Comparative Genomics in Hemiascomycete Yeasts: Evolution of Sex, Silencing, and Subtelomeres

Emmanuelle Fabre, Héloïse Muller, Pierre Therizols, Ingrid Lafontaine, Bernard Dujon, and Cécile Fairhead

Unité de Génétique Moléculaire des Levures, (URA2171 CNRS, UFR Université Pierre et Marie Curie), Département Structure et Dynamique des Génomes, Institut Pasteur, 75724 Cedex Paris, France

The recent release of sequences of several unexplored yeast species that cover an evolutionary range comparable to the entire phylum of chordates offers us a unique opportunity to investigate how genes involved in adaptation have been shaped by evolution. We have examined how three different sets of genes, all related to adaptative processes at the genomic level, have evolved in hemiascomycetes: (1) the mating-type genes that govern sexuality, (2) the silencing genes that are connected to regulation of mating-type cassettes and to telomere position effect, and (3) the gene families found repeated in subtelomeric regions.We report new combinations of mating-type genes and cassettes in hemiascomycetous species; we show that silencing proteins diverge rapidly. We have also found that in all species studied, subtelomeric gene families exist and are specific to each species.

Introduction

The question of whether the gene content of a genome can be correlated to adaptative properties can be addressed by comparing genome sequences from species with different lifestyles. In addition to partial genome sequences of closely related *Saccharomyces* species (Cliften et al. 2003; Kellis et al. 2003), the complete sequences of six unexplored yeast species (i.e., *Candida glabrata*, *Kluyveromyces waltii*, *Kluyveromyces lactis*, *Ashbya gossypii*, *Debaryomyces hansenii* and *Yarrowia lipolytica*) are now available (Dietrich et al. 2004; Dujon et al. 2004; Kellis, Birren, and Lander 2004). These species cover a vast range of lifestyles and exhibit a variety of life cycles and mating mechanisms (table 1).

We have examined, in the new hemiascomycete genomic sequences available, three different sets of genes, all related to adaptative processes at the genomic level. First, we analyzed genes involved in sexual reproduction because of the major role played by sex, or the loss of it, in a species' evolution. Second, we analyzed genes involved in silencing, because the silencing of sexual genes is an important feature in *S. cerevisiae*'s and other yeasts' sexual cycles and because silencing is involved in regulation of gene expression that is linked to an organism's direct adaptation to the environment. Third, we studied subtelomeric genes because of the plasticity of these regions and their capacity for harboring large gene families and because expression of some genes in subtelomeres is regulated through silencing mechanisms (Ai et al. 2002; De Las Penas et al. 2003; Halme et al. 2004).

Most ascomycetes have only two different mating types, as opposed to certain other fungi, in which many more mating types exist (Coppin et al. 1997). In all ascomycetes, the *MAT* locus encodes transcription factors that regulate mating-type–specific genes involved in pheromone production, pheromone sensing, and signal transduction (Fraser and Heitman 2004). Haploid-specific gene products are involved in repression of meiosis and mating-type switching, such as the *HO* endonuclease, at least in *S. cerevisiae* (Herskowitz and Oshima 1981; Haber 1998). In the case of *S. cerevisiae* and *Schizosaccharomyces pombe*, a very distant ascomycete species, homothallism is caused by gene conversion between the *MAT* locus and two *MAT*-like loci during cellular division of haploid cells (Herskowitz, Rine, and Strathern 1992; Klar 1992; Haber 1998).

The duplicated *MAT*-like cassettes must be transcriptionally repressed, and this is achieved through a process called silencing, which consists of the formation of a specialized compacted chromatin structure. In *S. cerevisiae*, the *MAT*-like cassettes, *HMR* and *HML*, are both surrounded by ''silencers,'' short specific sequences that are binding sites for DNA-binding proteins and are also involved in transcriptional activation and DNA replication (for recent reviews, see Gasser and Cockell [2001]. Grewal and Moazed [2003], and Rusche, Kirchmaier, and Rine [2003]). The other known proteins involved in this process are the SIR proteins, Sir1p through Sir4p, for which an enzymatic activity has only been described for the deacetylase Sir2p. Deacetylation of amino-terminal tails of histones (H3 and H4) can lead to nuclesome compaction correlated with transcriptional repression.

Silencing mechanisms have also been described in the regions close to telomeres and in telomeric repeats themselves. Such mechanisms exhibit the telomere position effect (TPE; for review, see Tham and Zakian [2002]). TPE is a repressor effect observed when reporter genes are inserted in a terminal position in a chromosome. It also occurs, to different extents, at about half of natural telomeres (Pryde and Louis 1999). In *S. cerevisiae*, chromosome ends consist of telomerase-dependent ($C_{1\text{-}3}A$) imperfect repeats, upstream of which several particular features are found. A retrotransposon-like Y′ element is present in two third of the chromosomes in strain S288C and a 500-bp core X element is found at all chromosome ends (Louis and Haber 1992; Louis 1995). In natural telomeres, silencing involves binding to the core X domain of almost the same proteins that also bind to *MAT*-like silent cassettes (Pryde and Louis 1999). Additionnal proteins that

Key words: mating type, gene families, silencing of chromatin, subtelomeres, hemiascomycetes.

E-mail: efabre@pasteur.fr; cfair@pasteur.fr

Mol. Biol. Evol. 22(4):856–873. 2005
doi:10.1093/molbev/msi070
Advance Access publication December 22, 2004

Table 1
Habitats and Life Cycles of Yeasts Presented Here

Species[a]	Habitat	Cycle	Mating
Saccharomyces cerevisiae	Grape must, other fruits	Haplo-diplontic[b]	Homothallic[c]
Candida glabrata	Human pathogen, animal feces[d]	Haploid cells only	Nonmater[e]
Kluyveromyces lactis	Dairy produce	Haplontic	Heterothallic, some homothallic strains[f]
Kluyveromyces waltii[g]	Probably insect-associated	Haplo-diplontic?	Homothallic
Ashbya gossypii[h]	Cotton-plant pathogen	Haploid cells only	Type strain is nonmater[i]
Debaryomyces hansenii	Salted foods[j]	Haplo-diplontic	Homothallic
Yarrowia lipolytica	foods (e.g., sausages, cheeses), oil fields[k]	Haplo-diplontic	Heterothallic

[a] Yeasts species are presented according to their phylogenetic distances. *Candida glabrata* is the closest yeast species to *Saccharomyces cerevisiae*, and *Yarrowia lipolytica* is the most distant one. Data were taken from the Genolevures Web site and Kurtzman (1994).
[b] It can be mitotically propagated in both diploid and haploid form in the laboratory.
[c] Genetically set to be a heterothallic (i.e. self-sterile) organism, *S. cerevisiae* is homothallic as a result of obligate mating-type switching at each haploid generation.
[d] do Carmo-Sousa and Barroso-Lopes (1969).
[e] No mating has yet been reported. see *Discussion* for details.
[f] Herman and Roman (1966).
[g] Habitat and cycle are not well characterized.
[h] Also named *Eremothecium gossypii*
[i] (P. Philipsen, personal communication).
[j] Cosentino et al. (2001).
[k] Barth and Gaillardin (1997).

bind telomeric repeats, such as the Ku complex, participate in TPE (Laroche et al. 1998). Subsequent to the binding of the Sir complex and hypoacetylation of amino-terminal tail of histone H3, nucleosome compaction spreads along the chromatin up to 5 to 8 kb from telomeres (Kimura, Umehara, and Horikoshi 2002; Suka, Luo, and Grunstein 2002). Histone acetyltransferase activities, as well as displacement of nucleosomes, are supposed to counteract such a spreading (Rusche, Kirchmaier, and Rine 2003).

We have further analyzed the gene content in larger subtelomeric regions than those subject to TPE. Recent characterization of subtelomeric regions from a variety of organisms from yeast to man has led to the realization that many chromosome ends are similar in structure (Mefford and Trask 2002). These subtelomeric regions correspond to about 30 kb of AT-rich, gene-poor sequences that contain no essential genes but contain some unique genes and genes present in multiple copies in other subtelomeric region. These subtelomeric regions are compacted in a heterochromatin domain extending from 10 to 25 kb from the telomere end (Robyr et al. 2002; Martin et al. 2004). Heterochromatin-like structure is then dependent on hypoacetylation and methylation of H3 but not on the Sir complex (Wyrick et al. 1999; Robyr et al. 2002; Martin et al. 2004). Industrial and ''wild'' yeasts are polymorphic for characters such as chromosome sizes and often heterozygous for sugar-utilization genes (Johnston, Baccari, and Mortimer 2000; Carro et al. 2003). These phenomena have been linked to plastic subtelomeric regions, which maintain gene families that are positively selected for in various brewing or baking strains of yeast.

By comparing genomic evolution of mating-type cassettes, we were able to describe further the evolutionary relationship between heterothallism and homothallism, corroborating the paradigm that heterothallism is the ancestral mode of sexual reproduction in ascomycetes (Coppin et al. 1997). We confirm that changes in modes of sexual reproduction are frequent when hemiascomycetous species adapt, indicating that sexual dysfunction is tolerated, possibly leading to complete loss of sexuality. We also show that characteristics of subtelomeric regions are probably shared widely among hemiascomycetes, although subtelomeres in each species present some specificities. As for silencing, we show that the evolutionary divergence of silencing mechanisms is far greater than the genes that are silenced themselves, thus, reinforcing the notion that transcription factors diverge faster than other proteins.

Materials and Methods

Genomes Analyzed in This Work

Six genomes of non- *Saccharomyces* hemiascomycetes were compared with *S. cerevisiae* in this work, four genomes from the Genolevures II effort (Dujon et al. 2004) (*Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii* and *Yarrowia lipolytica*) and two others (*Kluyveromyces waltii* [Kellis, Birren, and Lander 2004] and *Ashbya gossypii* [Dietrich et al. 2004]). For some analyses, partial sequences of six *Saccharomyces* genomes (Cliften et al. 2003; Kellis et al. 2003) (*S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S.castellii*, *S. kluyveri*), partial unpublished *K. thermotolerans* sequences (E. Talla et al., 2005), and partial *Z. rouxii* from the Genolevures I program (Souciet et al. 2000) were also examined. *C. albicans* data is from Magee and Magee (2000), Tzung et al. (2001), and our own Blast searches. The phylogenetic relationship between species is represented in figures and tables by the order of their names in rows and columns.

Comparative Gene Analysis

Analyses pertaining to genes encoding proteins involved in mating and in silencing and to subtelomeric gene families were performed by Blast searches using *S. cerevisiae* sequences as queries, and hits were usually found within annotated genes. To search for specific subtelomeric families, intraspecies Blast results for individual subtelomeric gene were examined. Genes in *MAT* loci and genes encoding pheromones, being smaller than average and sometimes containing introns, have sometimes been overlooked in annotation. Their identification is detailed below.

S. cerevisiae query sequences were retrieved from the Saccharomyces Genome Database (SGD, http://www.yeastgenome.org/). Most Blast searches, annotation and synteny data examination were done at SGD for *sensu stricto* and *sensu latto Saccharomyces* and for *A. gossypii*. Searches were done at the Genolevures site (http://cbi.labri.fr/Genolevures/) for *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, and *Yarrowia lipolytica* and for Genolevures I and *C. albicans* sequences. Were also used the *A. gossypii* Database (http://data.cgt.duke.edu/ashbya/Blast.html, http://agd.unibas.ch/), partial genomic sequences of *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii*, *S. kluyveri*at from the Genome Sequencing Center Database (http://genome.wustl.edu/projects/yeast/), and Blast on fungi at NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi). For *K. waltii*, contig and protein sequences were downloaded from (http://www.nature.com/nature/journal/v428/n6983/extref/nature02424-s1.htm) and analyzed locally. Depending on the site used and the type of sequences, BlastP, BlastN, tBlastN, or tBlastX was used, usually with default parameters (Gish, W. [1996–2004] http://Blast.wustl.edu). For some subtelomeric sequences, BlastP was used with the filters off, to take into account the low-complexity sequences of amino acid segments in certain proteins.

Homologs found by Blast searches were examined visually to check whether alignments covered at least roughly two thirds of the length of each protein. Homologs were classified in three categories as follows: ''highly similar'' (more than 80% similarity in amino acid sequence), ''similar'' (50% to 80% similarity), and ''weakly similar/some similarities with'' (less than 50% similarity and similarity spread throughout sequence/short, highly conserved segments).

Some *MAT* genes had already been characterized (Kurischko et al. 1992; Kurischko et al. 1999; Srikantha, Lachke, and Soll 2003; Wong et al. 2003), but some new *MAT* loci were retrieved with genes described in other species than *S. cerevisiae*. Multiple hits indicated the presence of extra cassettes. Allocation of *MAT* and *HML*/*HMR* nomenclature is based on subtelomeric localization of extra cassettes and synteny analysis of neighboring genes when conserved. In *K. waltii*, only two loci are identifiable; they are not yet mapped and cannot be classified. *MAT* and *HML*/*HMR*-like loci were aligned so they could be subdivided into smaller boxes: the X and Z1 boxes are present at all three loci, the W and Z2 are present in two copies, and the Y box determine mating type. In *A. gossypii*, all three loci code for *MAT**a*** information and therefore, the X and Z1 boxes cannot be identified. The genes encoding mating factor **a** were not annotated in the new genomes, except in *A. gossypii*. We found them by Blast search and because of the presence of a conserved CAAX motif (Chen et al. 1997). The genes encoding alpha pheromones were annotated as such in databases and are characterized by the presence of an octapeptidic motif repeated several times in the sequence (Singh et al. 1983).

Synteny Analyses and Search for Gene Relics

Syntenic clusters as defined in Dujon et al. (2004) were used to confirm the identity of a hit found by Blast analyses. In this case, neighboring genes of a putative ortholog are also homologs to neighboring genes of the query in *S. cerevisiae*. When no hit could be found by Blast analyses, syntenic clusters were also used to confirm that the lack of detection of an ortholog was not caused by accumulation of mutations in the ancient open reading frame. The search of such a relic was performed according to Lafontaine et al. (2004). When all attempts at finding an ortholog failed, we cannot evidently exclude that functional homologs, with no easily detectable sequence similarity, are nonetheless present in the considered species (this is the case for *SIR4* in *K. lactis*, [see *Results*]).

Results

Mating-Related Genes

Genes involved in the primary steps of the sexual cycle—pheromone production, pheromone sensing, signal transduction, meiosis, and mating-type switching—are regulated by the *MAT* locus that encodes transcription factors. As seen on table 2, genes coding for the sexual pheromones and their receptors have homologs in all six genomes, except *D. hansenii* and *Y. lipolytica*, which lack a recognizable mating factor **a** gene (see *Materials and Methods*) but have homologs of the genes encoding maturation proteins of the missing pheromone. In *Y. lipolytica*, two potential homologs are found for the mating factor alpha gene. A putative duplicate pair is also found in *A. gossypii*, with one gene annotated as such, but the octapeptidic repeat is absent (see *Materials and Methods*). Even when the gene coding for one of the pheromones seems to be missing in some species, genes encoding modification factors of the pheromone and genes encoding some of the proteins from the specific signal transduction cascade are present. This absence could result in the pathway not being functional or, alternatively, the actual gene is too diverged to be recognized. The conservation of genes encoding transcription factors at *MAT*-like loci is shown in table 3.

Overall, we show that most of the genes presented here are conserved across species, with levels of divergence increasing with phylogenetic distance.

MAT-like and *HMR*/*HML*-like Loci

We examined the homologs of the *MAT* and *MAT*-like loci in all species (fig. 1). These were already described for *Y. lipolytica*, *K. lactis*, and *C. glabrata* (Kurischko et al. 1992; Astrom et al. 2000; Butler et al. 2004), and we now extend this description to the other sequenced species. As shown on table 3, all have at least one *MAT* locus homolog; however, the sizes of these loci are variable. In *S. cerevisiae*, the *MAT* locus is less than 3 kb long, but in *A. gossypii*, it is almost 5 kb long (fig. 1). Variation not only in gene composition but also in gene size explains this phenomenon. Most *MAT* loci have either ''**a**'' or ''alpha'' type information, but interestingly, *D. hansenii* has a *MAT*-like locus that seems to be a mosaic of *MAT**a*** and *MATalpha* genes. We also searched for the existence of silent cassettes. *Y. lipolytica* and *D. hansenii* have a single *MAT*-like locus, whereas *K. lactis* and *C. glabrata*, as well as *A. gossypii*,

Table 2
Putative Homologs of *S. cerevisiae* Genes Involved in Sexual Cycle in Hemiascomycetes

SACE Gene		Protein Function	CAGL	KLWA	KLLA	ASGO	DEHA	YALI
HO	YDL227C	Mating-type switching endonuclease	CAGL0G05423g	nf	KLLA0E: *ca* 1638300–1640350 see figure 2	nf	nf	nf
ASH1	YKL185W	Daughter-cell specific inhibitor of HO expression	*CAGL0D00462g*	*4069*	*KLLA0B02651g*	*AER088C*	*DEHA0F03586g*	*YALI0E16577g*
MFa1 *MFa2*	YDR461W YNL145W	Precursors of **a** factor (pheromone)[a]	CAGL0C: 202488–202387[b]	Ctg_181: 25737,25639[b]	KLLA0E: 1149629–1149531 KLLA0C: 1339261–13393561[b]	ABL196C	nf	nf
MFalpha1 *MFalpha2*	YPL187W YGL089C	Precursors of alpha factor (pheromone)[a]	CAGL0H03135g	11171	KLLA0E19173g	*AFL062W* *AAR163C*	*DEHA0F20900g*[d]	*YALI0E16533g* *YALI0E30415g*
STE2	YFL026W	Alpha factor receptor	CAGL0K12430g	18044	KLLA0F25102g	AFR522C	DEHA0A11110g	*YALI0F03905g*
STE3	YKL178C	**a** Factor receptor	CAGL0M08184g	*13772*	KLLA0A06534g	AEL131C	DEHA0D04708g	*YALI0F11913g*
RAM1	YDL090C	Involved in farnesylation of **a** factor	CAGL0L07106g	3644	KLLA0F07161g	ADR275W	DEHA0G18007g	YALI0D14762g
RAM2	YKL019W	Involved in farnesylation of **a** factor	CAGL0L09801g	896	KLLA0E18051g	ACR094C	DEHA0F12034g	YALI0B16126g
STE24	YJR117W	Involved in maturation/proteolysis of **a** factor	**CAGL0I08217g**	**19761**	**KLLA0D10846g**	ABR163W	DEHA0F06820g	YALI0F11033g
RCE1	YMR274C	Involved in maturation/proteolysis of **a** factor	CAGL0F03443g	10314	KLLA0D01705g	AAL186W	DEHA0F08041g	*YALI0F07359g*
STE14	YDR410C	Involved in methylation of **a** factor	CAGL0F02805g	18741	KLLA0A02167g	AAR122C	DEHA0G06017g DEHA0G06028g[e]	*YALI0E22913g*
STE23	YLR389C	Involved in maturation/proteolysis of **a** factor	**CAGL0H06457g**	**4237**	**KLLA0E05049g**	AER053C	DEHA0A05214g	YALI0E17831g
AXL1	YPR122W	Involved in maturation/proteolysis of **a** factor	CAGL0D04686g	20464	KLLA0D15631g	AGR251C	*DEHA0E03179g*	YALI0F25091g
STE6	YKL209C	**a** Factor export	CAGL0K00363g	15561	KLLA0B14256g	AFR432W	DEHA0F18667g	YALI0E05973g
KEX2	YNL238W	Involved in maturation/proteolysis of alpha factor	CAGL0J07546g	24800	KLLA0D19811g	ABL203W	DEHA0C11308g	YALI0F13189g
STE13	YOR219C	Involved in maturation/proteolysis of alpha factor	CAGL0L02651g	8646	KLLA0D06919g	ACR103C	DEHA0G24486g	nf
KEX1	YGL203C	Involved in maturation/proteolysis of alpha factor	CAGL0G01232g	18993	KLLA0F09999g	AFR549W	DEHA0F23760g+ DEHA0F23749g[f]	YALI0B05170g
CDC24	YAL041W	Cdc42p guanine exchange factor	CAGL0M11968g	6865	KLLA0C12969g	ADR388C	DEHA0E12452g	*YALI0C14828g*
STE5	YDR103W	Adaptator	*CAGL0L06336g*	*23543*	*KLLA0F12023g*	*AGR104W*	*DEHA0G13684g*	nf

Table 2
Continued

SACE Gene		Protein Function	CAGL	KLWA	KLLA	ASGO	DEHA	YALI
FUS3	YBL016W	MAPKinase	**CAGL0J04290g**	**2687**	**KLLA0E10527g**	**AFR019W**	DEHA0E21219g	YALI0E23496g
FAR1	YJL157C	Inhibitor of cdc28p kinase	*CAGL0I06138g*	*13886*	*KLLA0B07469g*	*AEL104W*	*DEHA0F14905g*	nf
STE12	YHR084W	Specific transcription factor	CAGL0M01254g *CAGL0H02145g*[c]	20171	KLLA0E17193g	ADR304W	*DEHA0F27445g*	*YALI0E16236g*

NOTE.—First and second columns contain gene name and systematic nomenclature from *S. cerevisiae* genes, respectively. Third column gives a brief description of gene product, according to SGD. Next columns give gene names of orthologs in the different species examined. For *K. waltii*, gene names consist of numbers only. Species names are SACE: *S. cerevisiae*; CAGL: *C. glabrata*; KLWA: *K. waltii*; KLLA: *K. lactis*; ASGO: *A. gossypii*; DEHA: *D. hansenii*; YALI: *Y. lipolytica*. Ortholog conservation: bold indicates highly similar; regular indicates similar; italics indicate weakly similar or some similarities with corresponding gene from *S. cerevisiae* (see *Materials and Methods*). When genes used as queries are part of paralogous families in *S. cerevisiae* and other species, only the most likely ortholog is indicated according to level of similarity and conservation of synteny.

[a] Two genes exist in *S. cerevisiae*. When two possible orthologs are found, both are indicated.

[b] Nonannotated gene; coordinates are given.

[c] Two possible orthologs found.

[d] No significant similarity with gene from *S. cerevisiae* (see *Materials and Methods*).

[e] Tandemly repeated paralogs.

[f] Two contiguous genes align with the query (possible intron or pseudogene).

like *S. cerevisiae,* each contain two additional loci with similar information. Only two copies are found in the sequence available for *K waltii*, but a third one may have not yet been assembled from the sequence (8X coverage). Both *C. glabrata* and *K. lactis* have two copies on the same chromosome, and the third is elsewhere. *A. gossypii* has each copy on a different chromosome, and *K. waltii* has two copies on the same contig. All sequenced strains contain opposite mating-type information in the duplicated cassettes, except *A. gossypii*, in which all three copies contain the *MAT**a*** type information. The examination of duplicated cassettes from all species containing them shows that *MAT*, *HML*, and *HMR*-like loci are composed of repeated boxes, denoted W, X, Y, Z1, and Z2 in reference to *S. cerevisiae* (see Material and Methods), but that the sizes of the different boxes, their combination between loci, and even their presence is not conserved across species (fig. 1).

The presence of duplicated cassettes suggests that corresponding species undergo mating-type switching, a phenomenon driven by the *HO* endonuclease in *S. cerevisiae*. As shown on table 2, we have not found any *HO* homolog in species that have only one *MAT*-like cassette, such as *Y. lipolytica* and *D. hansenii*, but we have also not found it in either *A. gossypii* or *K. waltii*, which contain duplicated cassettes. We have found the *HO* homolog in *C. glabrata*, as published (Butler et al. 2004), and discovered a potential relic of the *HO* gene in *K. lactis*. As shown in figure 2, this relic is highly degenerate, but the dodecapeptide motif of *HO*-type endonucleases is still recognizable (Gimble 2000). This family of endonucleases includes the intein *VDE*, which can be confused with *HO* homologs and is inserted in *VMA1* (or *TFP1*) in *S. cerevisiae*. Because the *VDE* intein gene is present in *K. lactis* in a *VMA1* homolog, we can hypothesize that the sequence we have found is a relic of an ancient *HO* gene. Finally, we have found a dodecapeptide motif endonuclease homolog in *K. thermotolerans* that is not located inside the vacuolar ATPase gene and which we propose to be an *HO* homolog (fig. 3).

The *HO* recognition sites on chromosome III of *S. cerevisiae* are located at the border between the Y and Z1 boxes in all three cassettes, but only the one in the *MAT* locus is cut when switching occurs. The sequence from *C. glabrata* exhibits the three potential sites with only minor variations, as shown on figure 4. In *K. thermotolerans* and in *K. lactis*, the site is not recognizable, which is consistent with the absence of a functional *HO* gene.

Genes Involved in Silencing of *MAT*-like Loci, Telomeres, and Subtelomeres

We have analyzed how a set of proteins involved in mating-type silencing and the telomere position effect in *S. cerevisiae* were conserved in hemiascomycetes. Table 4 shows that proteins required for mating-type silencing and directly bound to *HMR* and *HML* silencers (i.e., Orc1p, Rap1p and Abf1p) present two behaviors. Rap1p and Abf1p are not detected in *D. hansenii* or *Y. lipolytica*; neither could be gene relics. This suggests either that they are too diverged to be recognized or that they have appeared in species close to *S. cerevisiae*. On the contrary, Orc1p is

Table 3
Homologs of Genes in MAT-like Loci

Species	Locus	**a1**	**a2**	Alpha1	Alpha2	Alpha3
SACE	MAT			YCR040w	YCR039c	na
	HMR/HML	YCR097w	YCR096c	YCL066w	YCL067c	na
CAGL	MAT			CAGL0B01243g	CAGL0B01265g	na
	HMR/HML	CAGL0E00341g	CAGL0E00319g	CAGL0B00242g	CAGL0B00264g	na
KLWA	MAT/HMR/HML	ctg_123: 21115..20879 ctg_123: 28935..28664[a]	ctg_123: 21506..22153[b]	12992	12995	na
KLLA	MAT	KLLA0C03135g	KLLA0C03157g			
	HMR/HML	KLLA0B14553g	KLLA0B14575g	KLLA0C00352g	KLLA0C00374g	KLLA0C00396g
ASGO	MAT	AFR643c	AFR643w-A			na
	HMR/HML	ADL394c, AER456w	ADL393w, AER455c			na
DEHA	MAT	DEHA0E:1595477..1596254[c]	DEHA0E:1596549..1597382[d]	DEHA0E20174g		na
YALI	MAT			YALI0C07480g	YALI0C07458g	na

NOTE.—Systematic gene names are given when available for homologs of **a1**, **a2**, alpha1, alpha2, alpha3, defined as explained in *Materials and Methods* and figure 1. No representation of the level of similarity is given, because not all are found as BlastP hits with *S. cerevisiae* queries. Paralogs are indicated one above the other, with MAT or HML/HMR allocation. Only genes from MAT of the sequenced strain have a name. Species names as in table 2; na indicates not applicable.

[a] **a1** is present twice in the genome. Coordinates of both copies are given, with only one annotated corresponding protein file (12991).

[b] **a2** was not annotated; coordinates are given.

[c] **a1** gene name is not conventional; coordinates are given for clarity.

[d] Coordinates of the whole gene are given. The two exons (see figure 1) correspond, respectively, to DEHA0E20218g and DEHA0E20210.

found conserved across species, perhaps because it belongs to the ORC complex also required for initiation of DNA replication, a strong selective pressure operating on this complex. The DNA-binding domain of Rap1p, which recognizes telomeric sequences, is, however, found conserved in yeast species in which Rap1p is detected (fig. 5), suggesting a conserved three-dimensional structure. Accordingly, a 6-bp AC-rich motif resembling the sequence recognized by Rap1p in *S. cerevisiae*, is present at least in *C. glabrata* and *K. lactis* telomeres (McEachern and Blackburn 1994; Dujon et al. 2004). Table 4 also shows that a homolog of Sir3p is found in *C. glabrata*, despite its weak sequence similarity. The lack of detection of a Sir3p homolog in the other hemiascomycetes and the resemblance between Orc1p and Sir3p in *S. cerevisiae*, in which they are paralogs, suggest that a duplication event occurred in the ancestor common only to these two species and that these duplicated copies of *S. cerevisiae* and *C. glabrata* have evolved rapidly (Kellis, Birren, and Lander 2004; Fabre et al., unpublished data). As shown in table 4, two additional proteins, Sir1p and Sir4p, have peculiar evolutive characteristics. They could not be detected in most of the yeasts studied here and neither could gene relics (I.L. and B.D., unpublished data). Indeed, we have failed to detect any structural homolog of Sir1p in any of the yeast species considered, including *C. glabrata*. A putative Sir1p was found in *S. castellii*, in which the amino acids required for binding to Orc1p are conserved (Bose et al. 2004, and data not shown), but because the phylogenetic position of *S. castellii* is ambigous (Kurtzman and Robnett 2003; Dujon et al. 2004), we cannot determine whether the lack of *SIR1* in *C. glabrata* is caused by rapid divergence of this gene, the loss of this gene in this species, or the de novo creation of *SIR1* in *Saccharomyces* species. It has to be noted that synteny between *S. cerevisiae* and *C. glabrata* is disrupted at the *SIR1* locus. In the case of Sir4p, we found it weakly conserved at the primary amino acid sequence in *C. glabrata*, but it belongs to the same syntenic block as in *S. cerevisiae*. In *K. lactis*, *K. waltii*, and *A. gossypii*, we could not find any obvious homolog unless synteny was examined (*Materials and Methods*). In fact, the syntenic *KLLA0F13420g* corresponds to the functional *K. lactis* homolog isolated by transcomplementation of a sir4Δ *S. cerevisiae* mutant but has no similarity to the gene from *S. cerevisiae* (Astrom and Rine 1998). Similarly, we found two putative homologs of *SIR4*, tandemly repeated, in *A. gossypii*. This is not the case in *D. hansenii*, in which the two genes that surround *SIR4* in *S. cerevisiae* in the same orientation, contiguous and separated by 50 bp, leave no space for a putative *SIR4* gene. Because the carboxy-terminal coiled-coil of Sir4p is essential for silencing at telomeres and silent mating loci (Chang et al. 2003; Murphy et al. 2003), we examined whether the putative orthologs of Sir4p show a conserved coiled-coil structure by using the algorithm multicoil (Wolf, Kim, and Berger 1997). We found a conserved coiled-coil at the carboxy-terminal part of these proteins (fig. 6), suggesting a conserved functional role. In the case of *A. gossypii*, the two copies show nonoverlapping coiled-coils. X-ray crystal structure of Sir4p coiled-coil reveals two interfaces, one formed between homodimeric, parallel coiled-coils and

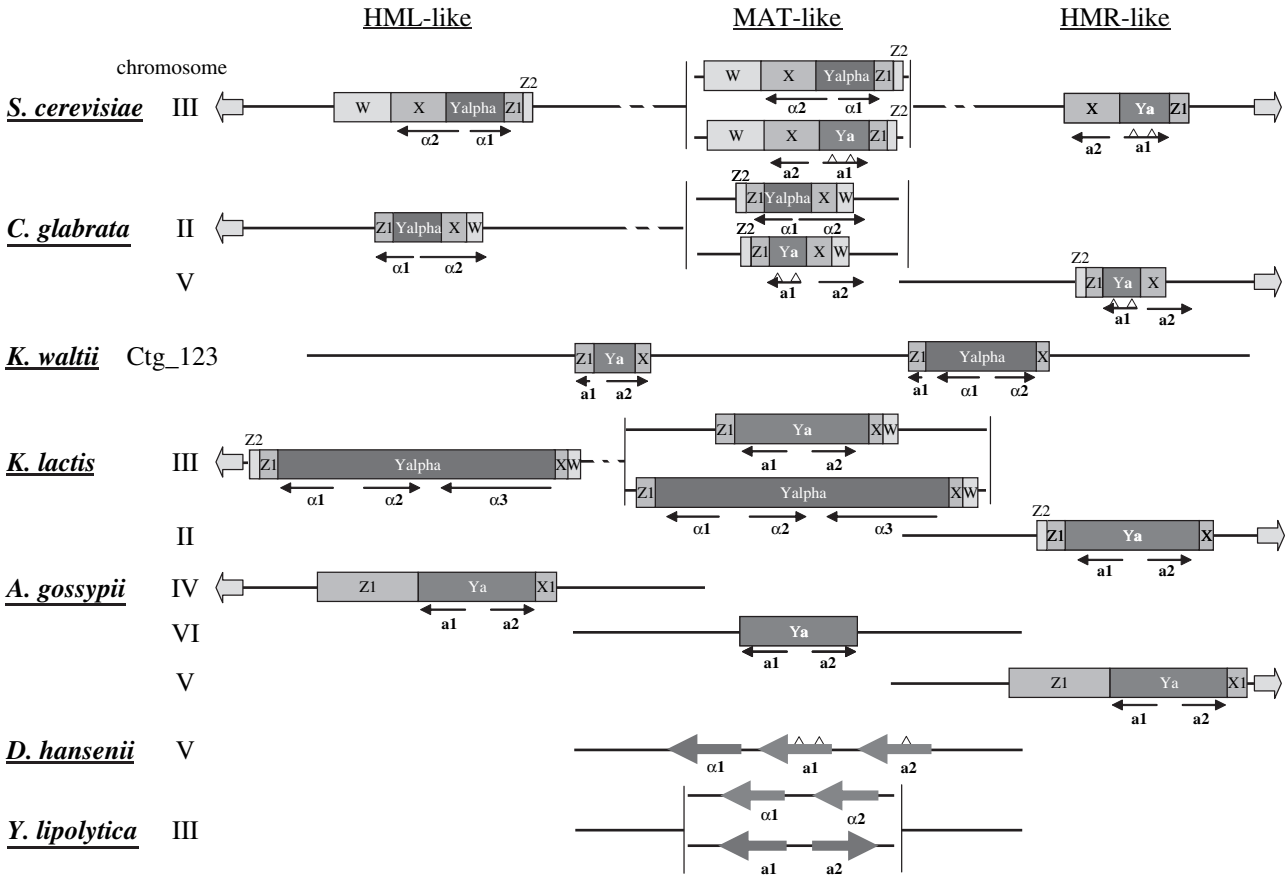

FIG. 1.—Representation of *MAT* and *MAT*-like loci in six yeast genomes in comparison to *S. cerevisiae*. Loci were identified from genome sequences, as described in *Materials and Methods*. On the left are shown the *HML* loci and homologs when present, in the center are the *MAT* loci, and on the right are the *HMR* loci and homologs. In *S. cerevisiae*, *HML*, *HMR*, and *MAT* itself are located on the same chromosome and contain identical sequence elements called W, X, Y, Z1, and Z2 "boxes." The Y box encodes either the "**a**" or the "alpha" information. We have noted all genes as **a** and alpha in reference to *S. cerevisiae*, even when genes had a different nomenclature, such as in *Y. lipolytica* (Barth and Gaillardin 1997). Black lines represent chromosomes, the number of which are indicated by roman numerals. Boxes represent identical elements between cassettes from the same species. Genes are represented by arrows beneath boxes or arrows on black line for genomes without duplicated cassettes and triangles represent putative introns. Telomeres are represented by arrows at ends of chromosomes. At the *MAT* locus, two alleles, or idiomorphs, exist in most species and are shown one above the other, where the top one is the one from the sequenced strain. Drawing is almost to scale. Sizes of *MAT**a*** loci are, in *S. cerevisiae*, in *C. glabrata*, in *K. waltii*, and in *K. lactis*, respectively, 2.4 kb, 1.4 kb, 1.3 kb, and 3 kb long. Sizes of *MATalpha* loci are, in the same order, 2.5 kb, 1.6 kb, 2.5 kb, and 5.9 kb long. In *A. gossypii*, the two identical cassettes are approximately 4 kb long. In *Y. lipolytica*, *MAT**a*** and *MATalpha* are approximately 3 kb long. In *D. hansenii*, the single locus is approximately 3.5 kb long. *Mata2* is probably not a true gene in *S. cerevisiae* (Johnson 1995), but homologs are found in *K. waltii* and in *C. glabrata*, where it has no start codon. The *Mata2* gene in all the other fungi has no obvious homology to *Mata2* from *S. cerevisiae*; all encoded proteins contain an HMG domain, and expression of this gene has been shown to prevent mating in *Y. lipolytica* (Kurischko et al. 1999). In *K. lactis*, an additional gene, *alpha3* is found; it is expressed from both *MAT* and *HML* (Astrom et al. 2000) and the double deletant is unable to mate.

the other between pairs of coiled-coils to form a large hydrophobic interface (Murphy et al. 2003). How each of these independent coiled-coils in *A. gossypii* participates in each of these interactions remains an open question.

All the other silencing proteins examined in table 4, such as the DNA-binding proteins yKu70p and yku80p, histone deacetylases, acetylases, or methyltransferases, were found in all hemiascomycetes, with a level of divergence that follows phylogenetic distances.

Subtelomeric Genes in Hemiascomycetes

We have first analyzed the conservation of subtelomeric genes of *S. cerevisiae* in other hemiascomycetes. The following set of genes were examined (table 5): gene families that encode proteins involved in carbohydrate metabolism, *SUC2*, *MAL*, *MEL*, *RTM*; the *FLO* family of adhesins; and families of uncharacterized proteins, *COS* and *PAU*, the largest subtelomeric family in S288C. We have extended our search to *Saccharomyces* species because previous partial data in these species have been central to the notion that these gene families are dynamic (see references in table 5).

The *FLO* family deserves special mention, as it has been the subject of many studies, some very recent (Halme et al. 2004). *FLO* genes encode serine/threonine-rich GPI-anchored cell wall proteins (Caro et al. 1997) that act as flocculins. The sequenced S288C strain has four copies of *FLO1* homologs and additional subtelomeric *FLO1* gene fragments and the nonsubtelomeric *FLO11* (Lo and Dranginis 1998). The *FLO* family has been directly implicated in molecular rearrangements of subtelomeric sequences (Carro et al. 2003). In *C. glabrata*, the *EPA* family of proteins, which has homology to the *FLO* family, has been

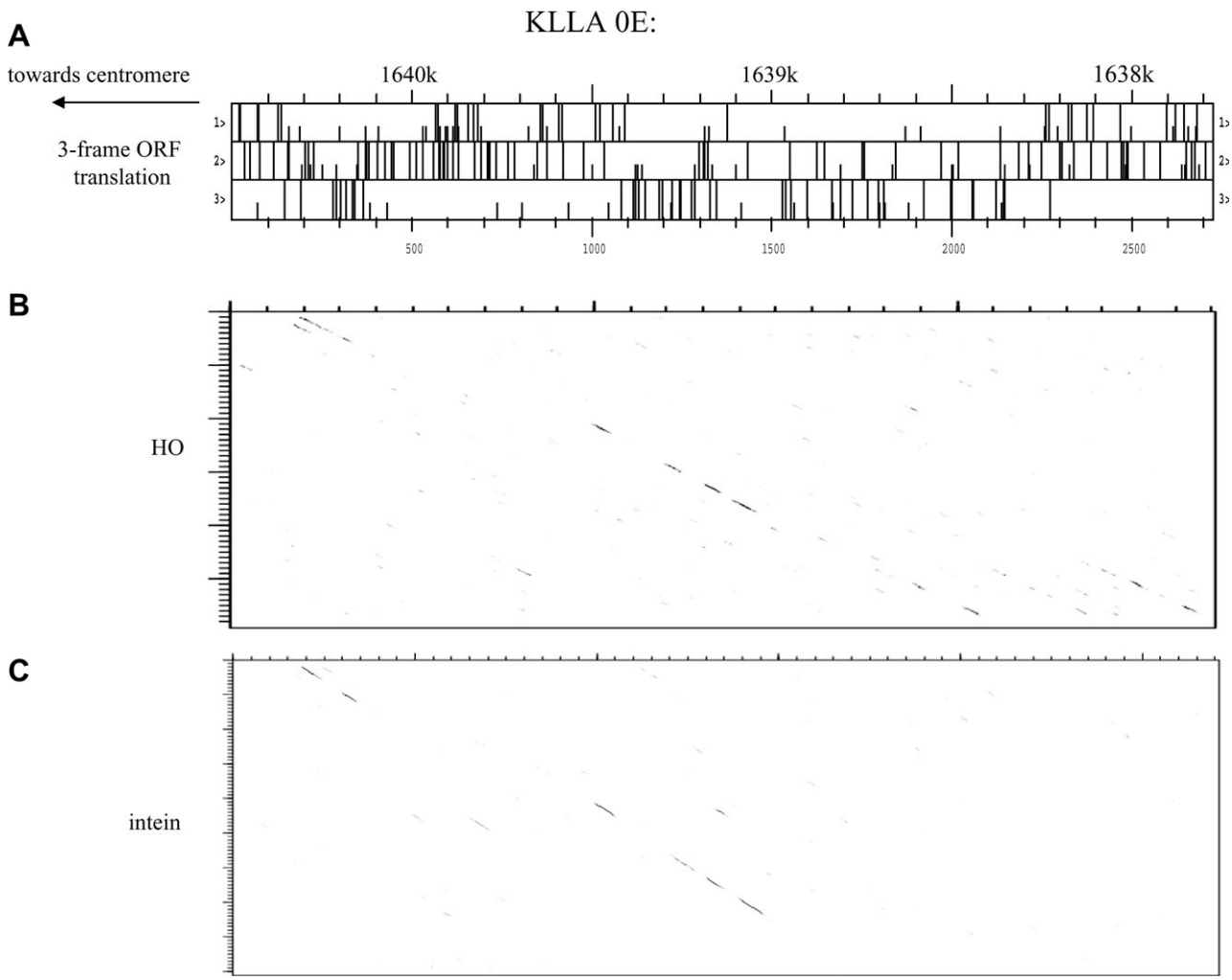

FIG. 2.—A putative *HO* gene relic in *K. lactis*. (*A*) Coordinates of the sequence segment of *K. lactis*, including the putative *HO* relic on chromosome 5 (KLLA0E). (*B*) and (*C*) Dotplot graphical outputs from DOTTER that compare nucleotide sequences translated into the three possible amino acid sequences (Sonnhammer and Durbin 1995) (sliding window of 16, threshold score of 35, BLOSUM62 matrix); *x*-axis same as in (*A*); *y*-axis in (*B*): *HO* (*YDL227c*); *y*-axis in C: intein from *TFP1* (*YDL185w*) (Chong et al. 1996).

shown to be involved in cellular adhesion (Cormack, Ghori, and Falkow 1999). Our *C. glabrata* data confirms that many entire copies and fragments of *EPA* genes exist, mainly in the last few kilobases close to the ends of the chromosomes, and, in fact, all genomes examined here contain multiple copies of genes and/or gene fragments with stretches of low-complexity serine/threonine amino acid sequences.

Overall, table 5 shows that many subtelomeric genes are conserved in *Saccharomyces* species. The situation is more complex outside of this group, as conservation or absence does not match evolutionary distance between the species. *D. hansenii* has many homologs of *S. cerevisiae* subtelomeric genes, whereas the phylogenetically closer *C. glabrata* has almost none, apart from the *EPA* family that has resemblance to the *FLO* family. This confirmed the high plasticity of these regions, and we, therefore, looked for the presence of specific families of genes that would be indicative of genes adapted to the species' environment. Indeed, this is what we found, by taking a direct look at the similarities between subtelomeric sequences of chromosomes in a given species, using data from the four Genolevures genomes (*Materials and Methods*). We have found several families in *Y. lipolytica* with one subtelomeric member; all others are internal. This situation exists in *S. cerevisiae*, for example, for the *HXT* genes (http://www.le.ac.uk/ge/ejl12/research/telostruc/ClustersLarge.html). In *K. lactis*, we have identified a sequence fragment of approximately 9 kb containing three genes that is repeated on seven different chromosome ends, in the same orientation relative to the telomere, two of which are represented on figure 7. This is a higher degree of redundancy at chromosome ends than what is observed in *S. cerevisiae* S288C. In all species examined so far, we found evidences of specific gene families at chromosome ends.

Discussion

In this work, we have examined three sets of genes directly linked to the divergence of species and their adaptation to an evolutionary niche: genes involved in the variation in the sexual transmission of genomes, genes that encode proteins involved in the variation of sexual mating

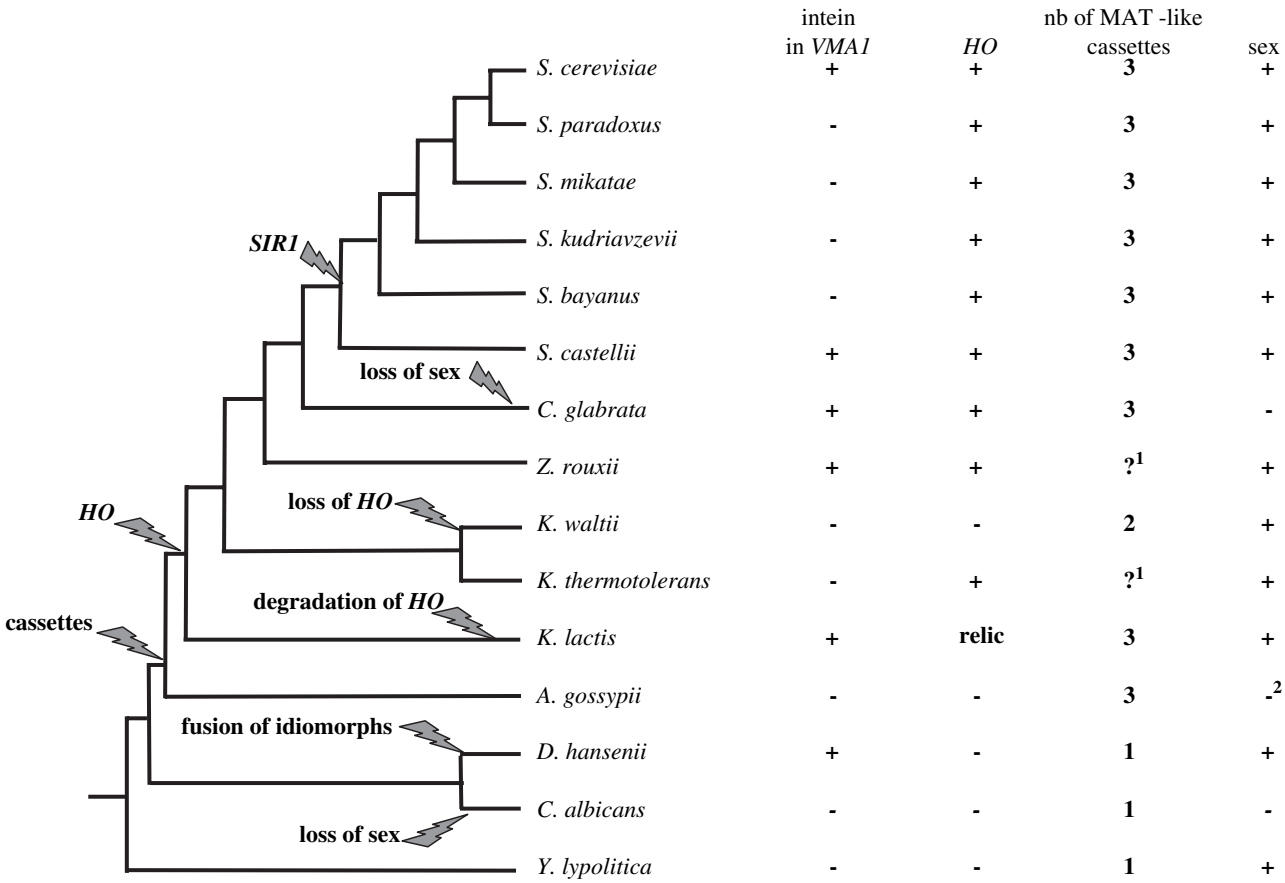

FIG. 3.—Intein gene, *HO* gene, and *MAT*-like cassettes in various yeast genomes. Approximate phylogenic relationships among the hemiascomycetes (adapted from Cai, Roberts, and Collins [1996], Kurtzman and Robnett [2003], and Dujon et al. [2004]) is represented on the left. The intein gene information is given for clarification of *HO* gene status (see text). A minus sign in the intein column means there is a vacuolar ATPase gene homolog in the genome, without the intein gene inserted in it. *S. kluyverii* was omitted because of negative search results in available sequences. The superscript 1 indicates negative search results in the incomplete sequences available. The superscript 2 means see table 1 and text for details. The appearance of the *HO* gene, MAT-like cassettes, and other features are indicated on the phylogenetic representation.

type and in the variation of expression of some gene families, and polymorphic gene families that lead to, and are created by, chromosomal rearrangements in genomes and that encode products needed in large quantities for adaptation to the environment. Many pathogens such as *Plasmodium*, *Trypanosoma*, and *Pneumocystis* generate variations of cell surface molecules by the presence of large subtelomeric families encoding surface glycoproteins, and human pathogens are prone to becoming asexual. Although several species here are animal or plant pathogens, these phenomena may represent a general situation in the hemiascomycetes' evolutionary branch.

The genomes presented here exhibit a great variety of situations pertaining to sexuality, mating-type switching, and homothallism, and evolution of sexuality among species is revealed by the genomic data (fig. 3). Two asexual species were analyzed, *A. gossypii* and *C. glabrata*. The sequenced type strain of *A. gossypii* contains three type ''**a**'' cassettes, which could explain the fact that it is described as a haploid nonmater. Such cassette configurations also exist in *C. glabrata* strains (Srikantha, Lachke, and Soll 2003) and in *S. cerevisiae*. Opposite mating types of strains of *A. gossypii* could however be searched for in nature, which would allow determination of whether asexuality is a peculiarity of the type strain possibly selected for loss of sexuality-induced clonality or whether it is related to the pathogenicity of this fungus in cotton plants. *C. glabrata* seems to have all the elements needed for mating and switching (Brockert et al. 2003; Wong et al. 2003; Butler 2004; H. Muller, and C. Fairhead, unpublished data), but only haploid cells have ever been isolated, and no mating has yet been observed. Nonetheless, two haploid types exist (Srikantha, Lachke, and Soll 2003; H. Muller, and C. Fairhead, unpublished data), and there are reports of mating-type switching (Brockert et al. 2003; Butler et al. 2004). Even though all three cassettes are not located on the same chromosome, switching is still theoretically possible; a slightly similar configuration has been shown in *S. cerevisiae* to allow switching in MAT**a** cells (Wu, Wu, and Haber 1997). Nonetheless, cells that have lost the capacity to switch mating types need not have lost the capacity to mate; these two characters are independent, as illustrated on figure3. In *C. glabrata*, we have observed that a synthetic alpha factor, deduced from the sequence from *C. glabrata*, has the property of arresting growth of MAT**a** cells of *S. cerevisiae*, but not of MAT**a** cells of *C. glabrata*, in a pheromone-mediated growth arrest assay (data not shown). The structure of the pheromone, although slightly diverged, is still recognized by the receptors in *S. cerevisiae*,

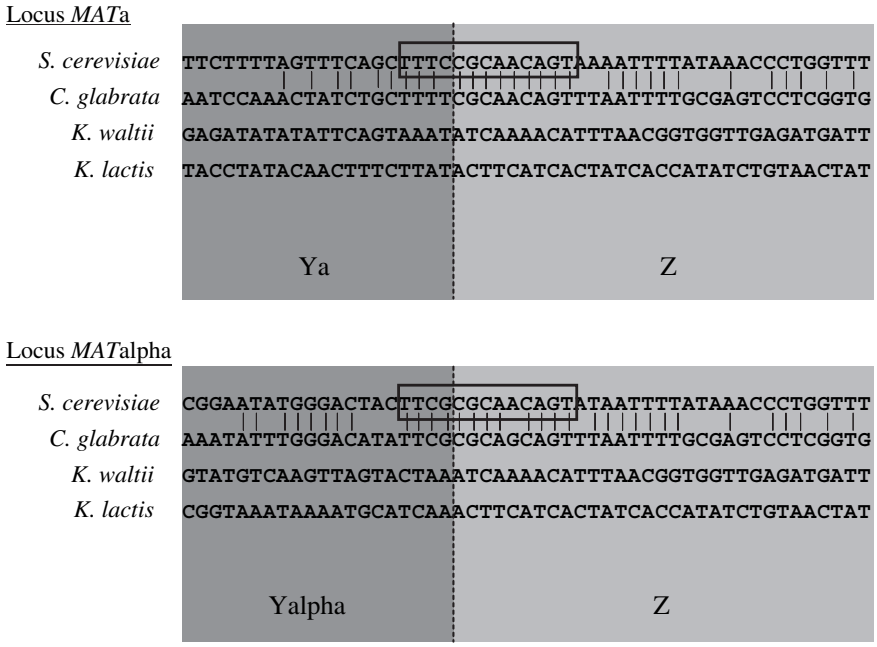

FIG. 4.—Y-Z1 junctions in yeast species with cassettes. Sequences from defined Y-Z1 limits are compared with the sequences of *S. cerevisiae* containing the *HO* recognition site. The sequences from *C. glabrata* are those that align best with *S. cerevisiae*, and it has been shown experimentally that the site present at the *MATalpha* locus from *C. glabrata* is cut by the *HO* endonuclease from *S. cerevisiae* (Nickoloff, Singer, and Heffron 1990).

raising the possibility of interspecies crosses. Because *C. glabrata* cells fail to respond, some component of the haploid pheromone-sensing pathway or of the mating response must be defective. It is possible that *C. glabrata*, like many other fungal pathogens, has lost its sexual mode of reproduction, a fact that may be correlated with the absence of transposable elements in this genome, as expected from theory (Hickey 1982; Wright and Finnegan 2001). Other asexual fungal pathogens include *C. albicans* (Magee and Magee 2000), which still contains mating-related genes but has lost meiosis-related genes (Tzung et al. 2001), *Aspergillus fumigatus* (Dyer, Paoletti, and Archer 2003), *Pneumocystis carnii* (Fraser and Heitman 2003), and *Cryptococcus neoformans*, in which pathogenicity is linked to the haploid state (Lengeler et al. 2000).

Most sexual species presented in figure 3 have *HML*/*HMR*-like cassettes, which appear after the split from *Y. lipolytica* and the *D. hansenii*/*C. albicans* branch. The *MAT* locus of *D. hansenii* was previously undescribed and seems to be a mosaic of type **a** and type alpha information, in accordance with the description of this species as homothallic. This is, to our knowledge, the first molecular description of a hemiascomycetous species that is homothallic without mating-type switching. The appearance of this new locus may result from the fusion of **a** and alpha loci, possibly in diploid cells, a mechanism speculated for the distant ascomycete *Cochliobolus* (Yun et al. 1999). No homolog of *MATalpha2* was found, and roles of transcriptional factors encoded by genes at the *MAT* locus may have diverged so as to control mating-related genes in a novel way. It must be noted that *D. hansenii* has genes for all maturation factors and receptors of both **a** and alpha pheromones, even though we have found no genes encoding **a** factor precursors.

In the branches of *Kluyveromyces* and *Saccharomyces*, the presence of the two duplicated *MAT*-like cassettes (fig.1), and their alternate transposition at the *MAT* locus, theoretically allows self-fertility in these species by cassette switching and *HO* recruitment. The *HO* gene, encoding the endonuclease that increases the frequency of switching, has been hypothetized to have appeared in the ancestor common to *S. cerevisiae* and *C. glabrata* but after the divergence from *K. lactis* (Butler et al. 2004). We have found a relic of the *HO* gene in the genome of *K. lactis* and a potentially functional homolog in *K. thermotolerans*; this implies that *HO* appeared before the separation with *Saccharomyces* species. *K. lactis* must have inherited this gene and lost it, and, therefore, become mostly heterothallic, as reported (table 1). *K. waltii* is described as homothallic and having a sexual cycle, and we identify no *HO* homolog and only two cassettes of opposite mating types. *K. waltii* may be a homothallic mater by a novel regulation of the two cassettes. Transcriptional analysis of these loci, in this and other species, would yield precious information on such processes.

A common characteristics of triplicated mating loci include the fact that they contain boxes of identical sequences and the fact that in all genomes in which the data was available, two cassettes are located in subtelomeric regions of chromosomes. This is compatible with the notion that the general organization of such loci was inherited from a common ancestor. However, the borders of the cassettes differ between species, and even the gene content of these loci varies, suggesting that they have evolved independently in each species, although in a concerted fashion between cassettes of the same genome. This may be the result of gene conversion mechanisms that are involved in mating-type switching. The cassettes could be homogenized

Table 4
Putative Homologs of *S. cerevisiae* Genes Involved in Chromatin Silencing in Hemiascomycetes

Protein Function	SACE		CAGL	KLLA	KLWA	ASGO	DEHA	YALI
DNA-binding proteins and chromatin silencing	*ORC1*[a]	YML065W	CAGL0L11264g	KLLA0B05016g	16977	*AER133C*	DEHA0D10439g	*YALI0D10104g*
	SIR3	YLR442C	*CAGL0M00770g*[cb]	nf	nf	nf	nf	nf
	RAP1	YNL216W	CAGL0K04917g[db]	KLLA0D19294g[eb]	4491	*ABL180W*	*DEHA0C13959g**	nf
	ABF1	YKL112W	*CAGL0J01177g*	*KLLA0F02970g*[fb]	*7699*	*ABL051W*	nf	nf
	SIR1	YKR101W	nf	nf	nf	nf	nf	nf
	SIR4	YDR227W	*CAGL0K11396g*	*KLLA0F13420g*[ghb]	*11611*[h]	*AGR188W/189W*[ho]	nf	nf
	ESC1[i]	YMR219W	*CAGL0F03971**	*KLLA0C07106g**	*9080**	*AER353W*[h]	nf	nf
	KU70[j]	YMR284W	CAGL0I02662g	KLLA0C06226g	10199	AFR443C/ABL117C	*DEHA0F11165g*	*YALI0C08701g*
	KU80[j]	YMR106C	CAGL0K03443g	*KLLA0B12672g*[kb]	*7624*	*ABL030W*	*DEHA0B01584g*	*YALI0E02068g*
Histone deacetylases	*SIR2*[l]	YDL042C	CAGL0K01463g/ CAGL0C05357g[n]	KLLA0F14663g[mb]	22853	AEL013C	DEHA0E14146g	*YALI0F11583g*
	HST1	YOL068C	CAGL0C05357g/ CAGL0K01463g[n]	KLLA0F14663g	22853	AEL013C	DEHA0E14146g	*YALI0F11583g*
	HST2	YPL015C	CAGL0L08668g	KLLA0F11033g	5070	AGL018C	DEHA0F08481g	*YALI0F11583g*
	HST3	YOR025w	CAGL0L08239g	KLLA0A07172g	18229	AEL229W	DEHA0E15763g	*YALI0D02145g*
	HST4	YDR191w	CAGL0F05621g	KLLA0D18535g	3264	AGL118W	*DEHA0E15763g*	*YALI0C23034g*
	RPD3	YNL330C	**CAGL0B01441g**	**KLLA0E01980g**	**22142**	**AGR395W**	**DEHA0F02706g**	**YALI0E22935g**
Histone acetylases	*SAS2*	YMR127C	CAGL0K03773g	KLLA0F17457g	17776	AFR221C	DEHA0G09504g	*YALI0D26598g*
	SAS3	YBL052C	CAGL0A04235g	*KLLA0E20009g*	*14800*	*ACR236W*	*DEHA0A07722g*	*YALI0A03861g*
	ESA1	YOR244W	**CAGL0J03696g**	**KLLA0F11209g**	**8779**	**ACR138W**	DEHA0C13585g	**YALI0E04675g**
	GCN5	YGR252W	CAGL0F08283g	*KLLA0A05115g*	*24776*	*AER297C*	*DEHA0G24794g*	*YALI0E02772g*
Histone methyltransferases	*SET1*	YHR119W	CAGL0L12980g	KLLA0F24134g	*14745*	*ABR136W*	*DEHA0F22176g*	nf
	DOT1	YDR440W	CAGL0J10516g	KLLA0B01287g	18501	AER326C	*DEHA0D02387g*	*YALI0E22715g*

NOTE.—Nomenclature is as in table 2. Asterisk (*) indicates similarity is weak but also found by synteny; nf indicates not found with our criteria.

[a] Orc1p and Sir3p are paralogs.

[b] Experimentally characterized function.

[c] De Las Penas et al. (2003).

[d] Haw, Yarragudi, and Uemura (2001).

[e] Krauskopf and Blackburn (1996).

[f] Goncalves et al. (1992).

[g] Astrom and Rine (1998).

[h] Found by synteny only; no significant sequence similarity.

[i] Esc1p has been recently shown to mediate silencing in a Sir4p-dependent, Ku80p- independent manner (Taddei et al. 2004). Evolution of Esc1p interestingly follows that of Sir4p.

[j] yku70p and yku80p are involved in TPE, but not in mating-type silencing. This complex is also required for DNA double-strand break repair through DNA-end joining of broken ends, a process conserved through evolution (Critchlow and Jackson 1998).

[k] Kooistra, Hooykaas, and Steensma (2004).

[l] Sir2p belongs to a family of four other deacetylases (Hst1p-Hst4p); the closest homolog is Hst1p.

[m] Chen and Clark-Walker (1994).

[n] Two genes exist in *S. cerevisiae* and two possible orthologs are indicated because of their high degree of similarity and because they belong to the same two syntenic blocks.

[o] Tandemly repeated paralogs.


YNL216W
CAGL0K04917g
Kwal_4491
ABL180W
KLLA0D19294g
DEHA0C13959g

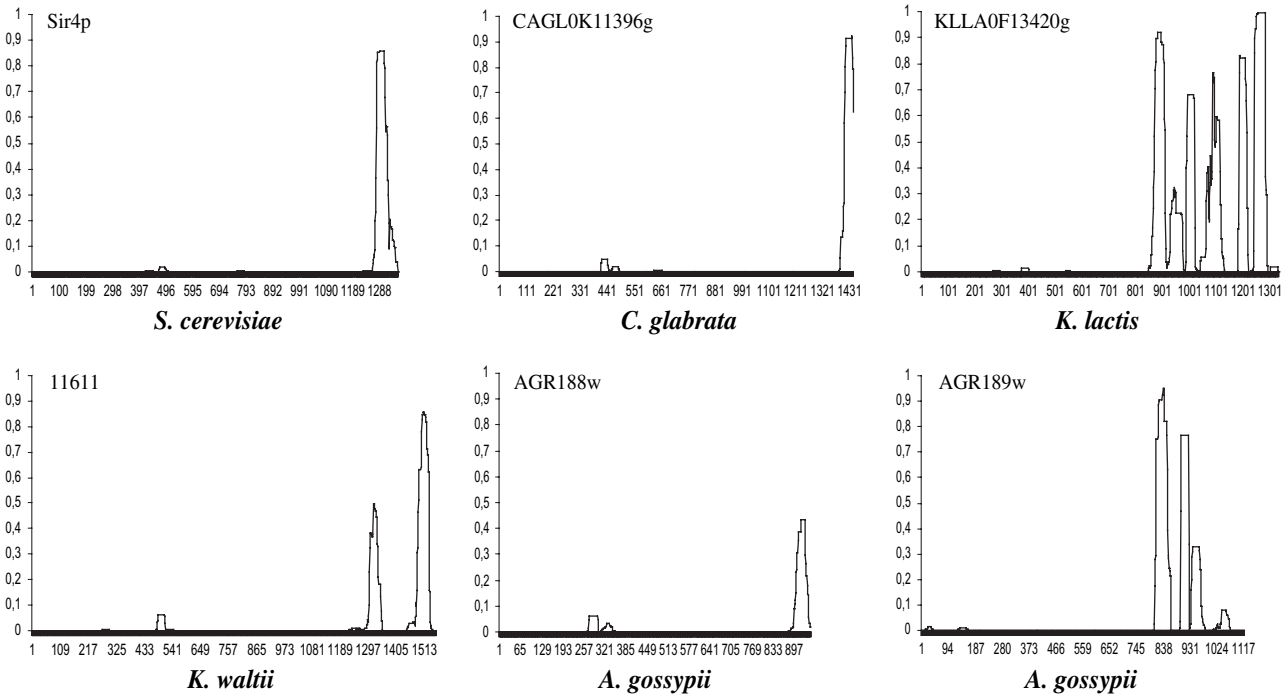

FIG. 6.—Coiled-coil structure prediction in putative Sir4p orthologs. The multicoil software (Wolf, Kim, and Berger 1997) was used to analyze the structure of putative *SIR4* orthologs as defined by conservation of synteny. A coiled-coil is predicted in the carboxy-terminal domain of each predicted protein from *S. cerevisiae* (Sir4p), *C. glabrata* (CAGL0K11396g), *K. lactis* (KLLA0F13420g), *K. waltii* (11611), and *A. gossypii (*AGR188w and AGR189w). The *x*-axes correspond to protein length in amino acid and the *y*-axes correspond to coiled-coil probability.

by the recombination process, the lowest percentage of mismatches between sequences itself allowing the most efficient recombination (Mezard, Pompon, and Nicolas 1992).

The second set of genes examined is involved in silencing. It is remarkable that along evolution, heterochromatin formation remains a conserved feature of eukaryotic genomes, allowing for genome stability and correct chromosome segregation (Perrod and Gasser 2003). Heterochromatin also plays a central role in the regulation of gene expression during development and cellular differentiation (Grewal and Moazed 2003). Heterochromatin-like structures are indeed involved in the stable inactivation of developmental regulators such as the homeotic gene clusters in *Drosophila* and mammals and, in *S. cerevisiae* and *S. pombe*, the mating-type genes.

In this work, we have analyzed how some central players of heterochromatin formation in *S. cerevisiae* have evolved in hemiascomycetes, and we have determined that a number of these silencing factors have appeared in branches close to *S. cerevisiae*. Because of this protein divergence, the question remains as to whether multiple mating-type cassettes are silenced in the species where they exist (fig.3). The first partners required in initial steps of establishment of silencing of mating-type cassettes in *S. cerevisiae* are DNA-binding proteins that include transcription factors. Studies in *K. lactis* have shown that mating-type silencing can occur through different transcription factors but that they belong to the same Myb domain transcription factor family (Astrom et al. 2000). Silencers have evolved quickly, and some transcription factors may have been substituted by others of the same family with more or less similar DNA-binding affinities. In subsequent steps of the formation of silenced chromatin, a linker protein containing a coiled-coil domain has a pivotal role in *S. cerevisiae*. We have shown that putative Sir4p homologs with such a conserved structural domain can be found in each species where more than one mating-type cassette is present, suggesting that silencing can be functional. Indeed, this has been proved for *K. lactis* (Astrom and Rine 1998). As for Sir3p, in most of the yeast species, we have no evidence for an ortholog, but it is possible that this function has been taken over by Orc1p or other proteins, although we cannot exclude that Sir3p function is not required for silencing in these yeasts. All *HML/HMR*-like loci may be silenced by a mechanism similar to that of *S. cerevisiae* and may corroborate the fact that these yeasts have a sexual cycle (except *A. gossypii* [see above]). The notable exception is *C. glabrata*, in which all the proteins required for mating-type silencing are present, except Sir1p (table 4).

←

FIG. 5.—RAP1and homolog sequences in different species. Multiple alignment between *S. cerevisiae RAP1* (*YNL216w*) and structural homologs are performed with ClustalW (Thompson, Higgins, and Gibson 1994). Sequences from *C. glabrata* (CAGL0K04917g), *K. waltii* (Kwal_4491), *A. gossypii* (ABL180W), *K. lactis* (KLLA0D19294g), and *D. hansenii* (DEHA0C13959g) are shown. On the top, DNA-binding domain (residues 361 to 596) of Rap1p, shown to crystallize with an 18-pb telomeric fragment and recognize DNA tandem repeats containing the ACACCA sequence (Konig et al. 1996), is surrounded by two black arrows. Black indicates identical residues; gray indicates similar residues.

Table 5
Subtelomeric Gene Families in *S. cerevisiae* and Their Putative Homologs

Syst. nom	Syn.	Function	S288C	Non S288C	*sensu stricto*	SACA	CAGL	KLWA	SAKL	KLLA	ASGO	DEHA	YALI
YIL162W	*SUC2*	β-Fructofuranosidase	unique, cen	6 subloci	**1** copy per genome	nf	nf	1	3	**1** cen	1	1sub	nf
YGR289C	*MAL11*	Symporter	2 sub+2[a]		**1–**3 copies per genome[b]	nf	nf	1	*1*	2[e] sub	nf	2sub+1cen[f]	1cen+1sub
YGR292W	*MAL12*	α-Glucosidase	2 sub+ 5[a]	5 subloci	**2–**5 copies per genome[b]	nf	nf	**1**	6	**2**[e] sub	nf	2sub[f]	nf
YGR288W	*MAL13*	Transcription factor	2 sub+2[a]		2–4 copies per genome	nf	nf	1	nf	nf	nf	1cen[f]	nf
na	*MEL1*	α-Galactosidase	Absent	10 subloci	**1** in SABA and SAMI, nf in SAKU and SAPA	nf *1*	nf 1cen	nf *1*	**1**[a] *4*	nf *1* sub	nf nf	1sub 1cen+1sub	nf nf
na	*RTM1*	Resistance to molasses	Absent[c]	7 subloci	1–4 copies per genome	nf	nf	nf	nf	nf	nf	nf	nf
YKL219W	*COS9*	Nuclear envelope[d]	12 sub	nd	1–10 copies per genome	1	*2* sub	**2**	nf	nf	nf	nf	nf
YCR104W	*PAU3*	Anaerobiosis induced	23 mostly sub	nd	**8–27** copies per genome	m w	m w	m w	m w	m w	m w[g]	m w	m w
YAR050W	*FLO1*	Flocculation lectin	4 sub+1 cen	sub	**3–**5 copies per genome[b]	m w	m w	m w	m w	m w	m w	m w	m w

NOTE.—Systematic nomenclature (when present in S288C) and synonymous gene names are indicated in the first two columns. Following columns contain information on function, status of family in the sequenced strain, and description of family in other strains. na indicates not applicable; nd indicates no data. Next columns show number of "hits" by Blast analysis, using *S. cerevisiae* genes as queries. Because most genes are part of families, and because many genomes are not fully annotated, we have indicated the number of potential copies we have found and the degree of identity but not their names. *sensu stricto*: SABA = *S. bayanus*, SAMI = *S. mikatae*, SAKU = *S. kudriavzevii*, SAPA= *S. paradoxus*; other species: SACA = *S. castellii*; SAKL = *S. kluyverii*. Other species' names as in table 2. Numbers of hits are bold when highly similar, regular plain when similar, and in italics when weakly similar or with some similarities. mw indicates multiple weak hits; nf indicates not found; sub indicates subtelomeric; cen indicates not subtelomeric. References: SUC (Naumov et al. 1992; Naumov et al. 1996); MAL (Naumov, Naumova, and Michels 1994), MEL (Turakainen, Aho, and Korhola 1993; Naumova 1996; Naumov, Naumova, and Louis 1995); RTM1 (Ness and Aigle 1995); COS (Spode et al. 2002); PAU (Rachidi et al. 2000); FLO (Teunissen and Steensma 1995; Lo and Dranginis 1998; Halme et al. 2004). Maps of subtelomeric regions in *S. cerevisiae* can be examined at the site of Professor E. J. Louis: http://www.le.ac.uk/ge/ejl12/research/telostruc/ClustersLarge.html.

[a] The complete *MAL* locus is present twice in the genome, but each gene also has dispersed homologs.

[b] Highly similar and similar paralogs coexist in each genome.

[c] There is a subtelomeric *RTM1* homolog in S288C: YER185w.

[d] The *COS* genes encode nuclear envelop proteins that might be involved in the cell's response to unfolded proteins.

[e] *MAL1* and *MAL2* homologs are contiguous and divergent, as in *S. cerevisiae*.

[f] One *MAL1* homolog is contiguous to one *MAL3* homolog, and the genes are convergent, as in *S. cerevisiae*, and another *MAL1* homolog is contiguous to a *MAL2* homolog in a convergent orientation.

[g] Although only two homologs of *FLO5* and one of *FLO8* are annotated, we have found multiple Blast hits in this genome.

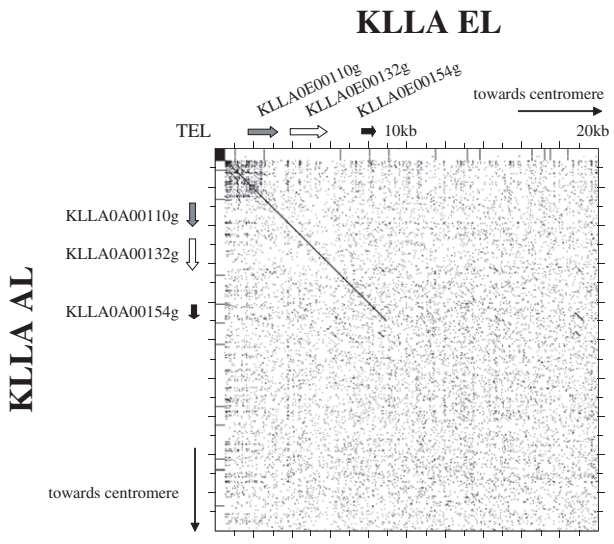

FIG. 7.—Subtelomeric families in *K. lactis*. Dot-matrix representation of nucleotide identity (15 identities per 23 nucleotides) between subtelomeric sequences from the left end of chromosome I (AL) and the left end of chromosome V (EL). Telomeric repeat alignments form the black square at top left corner (TEL), diagonal indicates the presence of an approximately 9-kb duplication between subtelomeres, encompassing three annotated genes on each sequence, schematized as gray, white, and black arrows, with their names indicated.

It could be hypothesized that the triplicate *MAT*-like loci of these yeast are, therefore, not silenced and would explain the failure to mate *C. glabrata* cells of opposite mating types, as this yeast would behave as a diploid.

As for subtelomeric silencing, it has been experimentally demonstrated that expression of genes involved in surface adhesion is regulated by an *SIR3*/*RAP1*-dependent mechanism in *C. glabrata* (De Las Penas et al. 2003), suggesting that subtelomeric silencing in this yeast species follows molecular mechanisms close to those of *S. cerevisiae*. Subtelomeric silencing in *K. lactis* is also dependent on Rap1p fixation on telomeric repeats (Gurevich et al. 2003). In the distant species *D. hansenii* and *Y. lipolytica*, the question as to how and whether subtelomeric silencing is achieved remains open. As we have observed that silencing proteins diverge rapidly, it is possible that subtelomeric regions in novel species are not subjected to silencing; this observation is consistent with the ability of these regions to harbor highly expressed gene families (see below).

Understanding how transcriptional inactive chromatin is established and maintained remains a challenge, but histones and their posttranslational modifications play a pivotal role in the assembly of heterochromatin-like structures (Grewal and Moazed 2003). Accordingly, histones and trans-modifying activities of amino-terminal histone H3 and H4 tails, such as the prominent histone deacetylases, acetylases, and methyltransferases, are also conserved in the hemiascomycetous analyzed here (table 4). However, the way heterochromatin complexes are targeted to specific domains are different and were probably invented several times, because in hemiascomycetous yeasts close to *S. cerevisiae*, silencers and DNA-binding proteins are the first key markers, whereas in metazoans, establishment of inactive chromatin implicate repetitive DNA and noncoding RNAs. Study of establishment of inactive chromatin in distant yeast species such as *D. hansenii* and *Y. lipolytica*, in which both DNA-binding proteins and Sir proteins and components of the RNAi pathway are not conserved (not shown), should, thus, provide precious information.

Finally, we have seen that subtelomeric gene families are species specific: even though sometimes orthologs do exist across species (i.e., the protein activity is conserved), they do not exhibit the same characteristics in terms of copy number and subtelomeric localization. This was noticeable in partial data on genes encoding sugar-utilizing enzymes in *Saccharomyces* species, because 69% of species-specific genes were reported to be subtelomeric (Kellis et al. 2003). Our preliminary analysis shows that the presence of subtelomeric gene families in all species examined is a general rule, even if there are specific characteristics that need further investigation.

Even though subtelomeres are regions where transcriptionally silenced genes have been described, one of the roles of subtelomeric DNA, at least in *S. cerevisiae*, is to allow amplification of gene families, the expression of which gives an evolutionary advantage to the cell (Turakainen, Aho, and Korhola 1993). We have seen that all species contain homologs of *FLO*-like proteins with low-complexity amino acid sequence segments, and *FLO*-like GPI- anchored cell wall proteins (GPI-CWP) have previously been characterized in the very distant hemiascomycete *Y. lipolytica* (Jaafar and Zueco 2004). The *FLO*-like families may have evolved as a reservoir of surface protein genes that can be expressed alternatively and genetically shuffled. The evolutionary mechanism may be of the same sort as described for interspecies divergence by Kellis, Birren, and Lander (2004) and Wolfe (2004), in which pairs of genes surviving from the whole-genome duplication have diverged so much as to not be recognizeable by BlastP analysis but are nonetheless related.

Apart from gene family amplification, another characteristic of subtelomeric sequences is the plasticity they confer onto chromosome ends. In *S. cerevisiae*, chromosome length polymorphism is mainly caused by illegitimate recombination between subtelomeric sequences and between Ty's (Rachidi et al. 1999). This sometimes results in the emergence of new, mosaic gene sequences (Kobayashi et al. 1998). Other properties of these repeated sequences in *S. cerevisiae* include allowing survival of a small number of cells when the telomerase complex is defective (Huang et al. 2001), where survival depends on Y′ sequence amplification, recombination between chromosome ends, and chromosome end fusion. Subtelomeric repeats also allow more cells to survive to a double-strand break than when the break is central, (Ricchetti, Dujon, and Fairhead 2003) because of break-induced replication and gene-conversion events occurring between *COS* and *FLO* genes. Even though Y′-type sequences are absent from the genomes of species outside the *Saccharomyces* group, other common features suggest that subtelomeric properties may be shared among ascomycetes and need experimental testing such as telomerase mutant analysis.

Mating-type cassette transposition and chromosomal rearrangements between subtelomeric sequences both rely

on the presence of the mitotic recombination machinery, which is shown to be conserved throughout hemiascomycetes (Richard et al., in press), thereby demonstrating the essential character of this machinery and emphasizing the generality of its involvement in numerous mitotic gene-conversion and genomic-rearrangement pathways.

Supplementary Material

Web Sites

Genolevures home page: http://cbi.labri.fr/Genolevures/

Blast on *Ashbya gossypii* at Duke: http://data.cgt.duke.edu/ashbya/blast.html

A*shbya gossypii* sequence home page: http://agd.unibas.ch/

Partial genomic sequences of *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii*, *S. kluyveri*at Genome Sequencing Center: http://genome.wustl.edu/projects/yeast/

Blast on fungi at NCBI: http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi

Blast and sequences from *K. waltii* in Supplementary Information of Kellis, Birren, and Lander (2004): http://www.nature.com/nature/journal/v428/n6983/extref/nature02424-s1.htm

Washington University Blast home page: http://blast.wustl.edu

Maps of Subtelomeres of *S. cerevisiae*: http://www.le.ac.uk/ge/ejl12/research/telostruc/ClustersLarge.html

*Saccharomyces*genome database: http://www.yeastgenome.org/

Acknowledgments

We thank Emmanuel Talla for access to *K. thermotolerans* sequences and Christophe Hennequin for help with mating-type gene analysis. We thank members of the Unité de Génétique Moléculaire des Levures and of the Génolevures consortium for general discussions and for annotation of the novel yeast sequences and the LABRI team for initiation and maintenance of the Génolevures database. This work was supported by the Consortium National de Recherche en Génomique (to Génoscope and to Institut Pasteur Génopole), the CNRS (GDR2354, Génolevures sequencing consortium), the Ministère de la Jeunesse, de l'Education et de la Recherche (ACI IMPBio n°IMPB114 ''Génolevures en ligne'') and the 'Conseil Régional d'Aquitaine' (''Génotypage et Génomique Comparée''). H.M. and P.T. are recipients of a doctoral fellowship of the Ministère de la Recherche through the Université Paris 6. B.D. is member of the Institut Universitaire de France.

Literature Cited

Ai, W., P. G. Bertram, C. K. Tsang, T. F. Chan, and X. F. Zheng. 2002. Regulation of subtelomeric silencing during stress response. Mol. Cell. **10**:1295–1305.

Astrom, S. U., A. Kegel, J. O. Sjostrand, and J. Rine. 2000. *Kluyveromyces lactis*Sir2p regulates cation sensitivity and maintains a specialized chromatin structure at the cryptic alpha-locus. Genetics **156**:81–91.

Astrom, S. U., and J. Rine. 1998. Theme and variation among silencing proteins in *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. Genetics **148**:1021–1029.

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

Bose, M. E., K. H. McConnell, K. A. Gardner-Aukema, U. Muller, M. Weinreich, J. L. Keck, and C. A. Fox. 2004. The origin recognition complex and Sir4 protein recruit Sir1p to yeast silent chromatin through independent interactions requiring a common Sir1p domain. Mol. Cell. Biol. **24**:774–786.

Brockert, P. J., S. A. Lachke, T. Srikantha, C. Pujol, R. Galask, and D. R. Soll. 2003. Phenotypic switching and mating type switching of *Candida glabrata* at sites of colonization. Infect. Immun. **71**:7109–7118.

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

Cai, J., I. N. Roberts, and M. D. Collins. 1996. Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces*, *Dekkera*, *and Kluyveromyces* deduced by small-subunit rRNA gene sequences. Int. J. Syst. Bacteriol. **46**:542–549.

Caro, L. H., H. Tettelin, J. H. Vossen, A. F. Ram, H. van den Ende, and F. M. Klis. 1997. In silicio identification of glycosylphosphatidylinositol-anchored plasma-membrane and cell wall proteins of *Saccharomyces cerevisiae*. Yeast **13**:1477–1489.

Carro, D., J. Garcia-Martinez, J. E. Perez-Ortin, and B. Pina. 2003. Structural characterization of chromosome I size variants from a natural yeast strain. Yeast **20**:171–183.

Chang, J. F., B. E. Hall, J. C. Tanny, D. Moazed, D. Filman, and T. Ellenberger. 2003. Structure of the coiled-coil dimerization motif of Sir4 and its interaction with Sir3. Structure (Camb.) **11**:637–649.

Chen, P., S. K. Sapperstein, J. D. Choi, and S. Michaelis. 1997. Biogenesis of the *Saccharomyces cerevisiae* mating pheromone a-factor. J. Cell Biol. **136**:251–269.

Chen, X. J., and G. D. Clark-Walker. 1994. sir2 mutants of *Kluyveromyces lactis* are hypersensitive to DNA-targeting drugs. Mol. Cell. Biol. **14**:4501–4508.

Chong, S., Y. Shao, H. Paulus, J. Benner, F. B. Perler, and M. Q. Xu. 1996. Protein splicing involving the *Saccharomyces cerevisiae* VMA intein: the steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. J. Biol. Chem. **271**:22159–22168.

Cliften, P., P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B. A. Cohen, and M. Johnston. 2003. Finding functional features in *Saccharomyces*genomes by phylogenetic footprinting. Science **301**:71–76.

Coppin, E., R. Debuchy, S. Arnaise, and M. Picard. 1997. Mating types and sexual development in filamentous ascomycetes. Microbiol. Mol. Biol. Rev. **61**:411–428.

Cormack, B. P., N. Ghori, and S. Falkow. 1999. An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. Science **285**:578–582.

Cosentino, S., M. E. Fadda, M. Deplano, A. F. Mulargia, and F. Palmas. 2001. Yeasts associated with Sardinian ewe's dairy products. Int. J. Food Microbiol. **69**:53–58.

Critchlow, S. E., and S. P. Jackson. 1998. DNA end-joining: from yeast to man. Trends Biochem. Sci. **23**:394–398.

De Las Penas, A., S. J. Pan, I. Castano, J. Alder, R. Cregg, and B. P. Cormack. 2003. Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to RAP1- and SIR-dependent transcriptional silencing. Genes Dev. **17**:2245–2258.

Dietrich, F. S., S. Voegeli, S. Brachat et al. (14 co-authors). 2004. The *Ashbyagossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae*genome. Science **304**:304–307.

do Carmo-Sousa, L., and C. Barroso-Lopes. 1969. A comparative study of the extracellular and cell wall polysaccharides of some *Candida* species. Antonie Van Leeuwenhoek **35**(Suppl.):A7-8.

Dujon, B., D. Sherman, G. Fischer et al. (67 co-authors). 2004. Genome evolution in yeasts. Nature **430**:35–44.

Dyer, P. S., M. Paoletti, and D. B. Archer. 2003. Genomics reveals sexual secrets of *Aspergillus*. Microbiology **149**:2301–2303.

Fraser, J. A., and J. Heitman. 2003. Fungal mating-type loci. Curr. Biol. **13**:R792–795.

———. 2004. Evolution of fungal sex chromosomes. Mol. Microbiol. **51**:299–306.

Gasser, S. M., and M. M. Cockell. 2001. The molecular biology of the SIR proteins. Gene **279**:1–16.

Gimble, F. S. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiol Lett. **185**: 99–107.

Goncalves, P. M., K. Maurer, W. H. Mager, and R. J. Planta. 1992. *Kluyveromyces* contains a functional ABF1-homologue. Nucleic Acids Res. **20**:2211–2215.

Grewal, S. I., and D. Moazed. 2003. Heterochromatin and epigenetic control of gene expression. Science **301**:798–802.

Gurevich, R., S. Smolikov, H. Maddar, and A. Krauskopf. 2003. Mutant telomeres inhibit transcriptional silencing at native telomeres of the yeast *Kluyveromyces lactis*. Mol. Genet. Genomics **268**:729–738.

Haber, J. E. 1998. Mating-type gene switching in *Saccharomyces cerevisiae*. Annu. Rev. Genet. **32**:561–599.

Halme, A., S. Bumgarner, C. Styles, and G. R. Fink. 2004. Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. Cell **116**:405–415.

Haw, R., A. D. Yarragudi, and H. Uemura. 2001. Isolation of a *Candida glabrata* homologue of RAP1, a regulator of transcription and telomere function in *Saccharomyces cerevisiae*. Yeast **18**:1277–1284.

Herman, A., and H. Roman. 1966. Allele specific determinants of homothallism in *Saccharomyces lactis*. Genetics **53**:727–740.

Herskowitz, I., and Y. Oshima. 1981. Control of cell type in *Saccharomyces cerevisiae*: mating type and Mating-type interconversion. Pp. 181–209 *in* The molecular biology of the yeast *Saccharomyces*: Life cycle and inheritance. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Herskowitz, I., J. Rine, and J. N. Strathern. 1992. Mating-type determination and mating-type interconversion in *S. cerevisiae*. Pp. 583–656 *in* J. R. Broach, J. R. Pringle, and E. W. Jones, eds. The molecular biology of the yeast *Saccharomyces*: Gene expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hickey, D. A. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics **101**:519–531.

Huang, P., F. E. Pryde, D. Lester, R. L. Maddison, R. H. Borts, I. D. Hickson, and E. J. Louis. 2001. SGS1 is required for telomere elongation in the absence of telomerase. Curr. Biol. **11**:125–129.

Jaafar, L., and J. Zueco. 2004. Characterization of a glycosylphosphatidylinositol-bound cell-wall protein (GPI-CWP) in *Yarrowia lipolytica*. Microbiology **150**:53–60.

Johnson, A. D. 1995. Molecular mechanisms of cell-type determination in budding yeast. Curr. Opin. Genet. Dev. **5**:552–558.

Johnston, J. R., C. Baccari, and R. K. Mortimer. 2000. Genotypic characterization of strains of commercial wine yeasts by tetrad analysis. Res. Microbiol. **151**:583–590.

Kellis, M., B. W. Birren, and E. S. Lander. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature **428**:617–624.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423**:241–254.

Kimura, A., T. Umehara, and M. Horikoshi. 2002. Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. Nat. Genet. **32**:370–377.

Klar, A. J. S. 1992. Molecular genetics of fission yeast cell type: mating type and mating-type interconversion. Pp. 745–777 *in* J. R. Broach, J. R. Pringle, and E. W. Jones, eds. The molecular biology of the yeast *Saccharomyces*: Gene expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kobayashi, O., N. Hayashi, R. Kuroki, and H. Sone. 1998. Region of FLO1 proteins responsible for sugar recognition. J. Bacteriol. **180**:6503–6510.

Konig, P., R. Giraldo, L. Chapman, and D. Rhodes. 1996. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell **85**:125–136.

Kooistra, R., P. J. Hooykaas, and H. Y. Steensma. 2004. Efficient gene targeting in *Kluyveromyces lactis*. Yeast **21**:781–792.

Krauskopf, A., and E. H. Blackburn. 1996. Control of telomere growth by interactions of RAP1 with the most distal telomeric repeats. Nature **383**:354–357.

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

Kurischko, C., M. B. Schilhabel, I. Kunze, and E. Franzl. 1999. The MATA locus of the dimorphic yeast *Yarrowia lipolytica* consists of two divergently oriented genes. Mol. Gen. Genet. **262**:180–188.

Kurtzman, C. P. 1994. Molecular taxonomy of the yeasts. Yeast **10**:1727–1740.

Kurtzman, C. P., and C. J. Robnett. 2003. Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res. **3**:417–432.

Lafontaine, I., G. Fischer, E. Talla, and B. Dujon. 2004. Gene relics in the genome of the yeast *Saccharomyces cerevisiae*. Gene **335**:1–17.

Laroche, T., S. G. Martin, M. Gotta, H. C. Gorham, F. E. Pryde, E. J. Louis, and S. M. Gasser. 1998. Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. Curr. Biol. **8**:653–656.

Lengeler, K. B., P. Wang, G. M. Cox, J. R. Perfect, and J. Heitman. 2000. Identification of the MATa mating-type locus of *Cryptococcus neoformans* reveals a serotype A MATa strain thought to have been extinct. Proc. Natl. Acad. Sci. USA **97**:14455–14460.

Lo, W. S., and A. M. Dranginis. 1998. The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by *Saccharomyces cerevisiae*. Mol. Biol. Cell. **9**:161–171.

Louis, E. J. 1995. The chromosome ends of *Saccharomyces cerevisiae*. Yeast **11**:1553–1573.

Louis, E. J., and J. E. Haber. 1992. The structure and evolution of subtelomeric Y′ repeats in *Saccharomyces cerevisiae*. Genetics **131**:559–574.

Magee, B. B., and P. T. Magee. 2000. Induction of mating in *Candida albicans* by construction of MTLa and MTLalpha strains. Science **289**:310–313.

Martin, A. M., D. J. Pouchnik, J. L. Walker, and J. J. Wyrick. 2004. Redundant roles for histone H3 N-terminal lysine residues in subtelomeric gene repression in *Saccharomyces cerevisiae*. Genetics **167**:1123–1132.

McEachern, M. J., and E. H. Blackburn. 1994. A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. Proc. Natl. Acad. Sci. USA **91**:3453–3457.

Mefford, H. C., and B. J. Trask. 2002. The complex structure and dynamic evolution of human subtelomeres. Nat. Rev. Genet. **3**:91–102.

Mezard, C., D. Pompon, and A. Nicolas. 1992. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell **70**:659–670.

Murphy, G. A., E. J. Spedale, S. T. Powell, L. Pillus, S. C. Schultz, and L. Chen. 2003. The Sir4 C-terminal coiled coil is required for telomeric and mating type silencing in *Saccharomyces cerevisiae*. J. Mol. Biol. **334**:769–780.

Naumov, G. I., E. S. Naumova, R. A. Lantto, E. J. Louis, and M. Korhola. 1992. Genetic homology between *Saccharomyces cerevisiae* and its sibling species *S. paradoxus* and *S. bayanus*: electrophoretic karyotypes. Yeast **8**:599–612.

Naumov, G. I., E. S. Naumova, and E. J. Louis. 1995. Genetic mapping of the alpha-galactosidase MEL gene family on right and left telomeres of *Saccharomyces cerevisiae*. Yeast **11**: 481–483.

Naumov, G. I., E. S. Naumova, and C. A. Michels. 1994. Genetic variation of the repeated MAL loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. Genetics **136**:803–812.

Naumov, G. I., E. S. Naumova, E. D. Sancho, and M. P. Korhola. 1996. Polymeric SUC genes in natural populations of *Saccharomyces cerevisiae*. FEMS Microbiol. Lett. **135**:31–35.

Ness, F., and M. Aigle. 1995. RTM1: a member of a new family of telomeric repeated genes in yeast. Genetics **140**:945–956.

Nickoloff, J. A., J. D. Singer, and F. Heffron. 1990. In vivo analysis of the *Saccharomyces cerevisiae* HO nuclease recognition site by site-directed mutagenesis. Mol. Cell. Biol. **10**:1174–1179.

Perrod, S., and S. M. Gasser. 2003. Long-range silencing and position effects at telomeres and centromeres: parallels and differences. Cell Mol. Life Sci. **60**:2303–2318.

Pryde, F. E., and E. J. Louis. 1999. Limitations of silencing at native yeast telomeres. EMBO J. **18**:2538–2550.

Rachidi, N., M. J. Martinez, P. Barre, and B. Blondin. 2000. *Saccharomyces cerevisiae* PAU genes are induced by anaerobiosis. Mol. Microbiol. **35**:1421–1430.

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

Ricchetti, M., B. Dujon, and C. Fairhead. 2003. Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast. J. Mol. Biol. **328**:847–862.

Richard, G. E., A. Kerrest, L. Lafontaine, and B. Dujon, 2005. comparative genomics of hemiascomycete yeasts: genes involved in DNA replication, repair, and recombination. Mol Biol Evol, in press.

Robyr, D., Y. Suka, I. Xenarios, S. K. Kurdistani, A. Wang, N. Suka, and M. Grunstein. 2002. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. Cell **109**:437–446.

Rusche, L. N., A. L. Kirchmaier, and J. Rine. 2003. The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. Annu. Rev. Biochem **72**:481–516.

Singh, A., E. Y. Chen, J. M. Lugovoy, C. N. Chang, R. A. Hitzeman, and P. H. Seeburg. 1983. *Saccharomyces cerevisiae* contains two discrete genes coding for the alpha-factor pheromone. Nucleic Acids Res. **11**:4049–4063.

Sonnhammer, E. L., and R. Durbin. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167**:GC1–10.

Souciet, J., M. Aigle, F. Artiguenave et al. (24 co-authors). 2000. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett. **487**:3–12.

Spode, I., D. Maiwald, C. P. Hollenberg, and M. Suckow. 2002. ATF/CREB sites present in sub-telomeric regions of *Saccharomyces cerevisiae* chromosomes are part of promoters and act as UAS/URS of highly conserved COS genes. J. Mol. Biol. **319**:407–420.

Srikantha, T., S. A. Lachke, and D. R. Soll. 2003. Three mating type-like loci in *Candida glabrata*. Eukaryot. Cell. **2**:328–340.

Suka, N., K. Luo, and M. Grunstein. 2002. Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. Nat. Genet. **32**:378–383.

Taddei, A., F. Hediger, F. R. Neumann, C. Bauer, and S. M. Gasser. 2004. Separation of silencing from perinuclear anchoring functions in yeast Ku80, Sir4 and Esc1 proteins. EMBO J. **23**:1301–1312.

Teunissen, A. W., and H. Y. Steensma. 1995. Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. Yeast **11**:1001–1013.

Tham, W. H., and V. A. Zakian. 2002. Transcriptional silencing at *Saccharomyces* telomeres: implications for other organisms. Oncogene **21**:512–521.

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.

Turakainen, H., S. Aho, and M. Korhola. 1993. MEL gene polymorphism in the genus *Saccharomyces*. Appl. Environ. Microbiol. **59**:2622–2630.

Tzung, K. W., R. M. Williams, S. Scherer et al. (13 co-authors). 2001. Genomic evidence for a complete sexual cycle in *Candida albicans*. Proc. Natl. Acad. Sci. USA **98**:3249–3253.

Wolf, E., P. S. Kim, and B. Berger. 1997. MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci. **6**:1179–1189.

Wolfe, K. 2004. Evolutionary genomics: yeasts accelerate beyond BLAST. Curr. Biol. **14**:R392–394.

Wong, S., M. A. Fares, W. Zimmermann, G. Butler, and K. H. Wolfe. 2003. Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast *Candida glabrata*. Genome Biol. **4**:R10.

Wright, S., and D. Finnegan. 2001. Genome evolution: sex and the transposable element. Curr. Biol. **11**:R296–299.

Wu, X., C. Wu, and J. E. Haber. 1997. Rules of donor preference in *Saccharomyces* mating-type gene switching revealed by a competition assay involving two types of recombination. Genetics **147**:399–407.

Wyrick, J. J., F. C. Holstege, E. G. Jennings, H. C. Causton, D. Shore, M. Grunstein, E. S. Lander, and R. A. Young. 1999. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature **402**:418–421.

Yun, S. H., M. L. Berbee, O. C. Yoder, and B. G. Turgeon. 1999. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. Proc. Natl. Acad. Sci. USA **96**:5592–5597.

Franz Lang, Associate Editor

Accepted December 14, 2004