Biodiversity in sulfur metabolism in hemiascomycetous yeasts

Agnès Hébert\textsuperscript{1,2}, Serge Casaregola\textsuperscript{1,2} & Jean-Marie Beckerich\textsuperscript{1,2}

\textsuperscript{1}INRA, UMR1319, Institut MICALIS, AgroParisTech, Thiverval-Grignon, France; and \textsuperscript{2}AgroParisTech, UMR1319, AgroParisTech, Thiverval-Grignon, France

Abstract

The evolution of the metabolism of sulfur compounds among yeast species was investigated. Differences between species were observed in the cysteine biosynthesis pathway. Most yeast species possess two pathways leading to cysteine production, the transsulfuration pathway and the O-acetyl-serine (OAS) pathway, with the exception of \textit{Saccharomyces cerevisiae} and \textit{Candida glabrata}, which only display the transsulfuration pathway, and \textit{Schizosaccharomyces pombe}, which only have the OAS pathway. An examination of the components of the regulatory network in the different species shows that it is conserved in all the species analyzed, as its central component Met4p was shown to keep its functional domains and its partners were present. The analysis of the presence of genes involved in the catabolic pathway shows that it is evolutionarily conserved in the sulfur metabolism and leads us to propose a role for two gene families which appeared to be highly conserved. This survey has provided ways to understand the diversity of sulfur metabolism products among yeast species through the reconstruction of these pathways. This diversity could account for the difference in metabolic potentialities of the species with a biotechnological interest.

Introduction

Sulfur metabolism is well known in the hemiascomycetous yeast model \textit{Saccharomyces cerevisiae}, from the assimilation of inorganic sulfate to the synthesis of sulfur amino acids by the transsulfuration pathway (cysteine) and the methyl cycle (methionine). Sulfur metabolism is involved in numerous metabolisms through S-adenosylmethionine (SAM), a source of methyl groups, methylene groups, amino groups, aminopropyl groups and 5'-deoxyadenosyl radicals (Fontecave \textit{et al.}, 2004). Sulfur metabolism plays an important role in the response to cadmium stress by the intermediary of glutathione (Fauchon \textit{et al.}, 2002). The archeanascomycetous yeast \textit{Schizosaccharomyces pombe} presents divergent sulfur metabolism pathways, in comparison with \textit{S. cerevisiae}. In fact, \textit{S. pombe} has an incomplete transsulfuration pathway that does not lead to cysteine production (Baudouin-Cornu & Labarre, 2006). The synthesis of cysteine is carried out by the O-acetyl-serine (OAS) pathway, which does not exist in \textit{S. cerevisiae}. The euscomycetes \textit{Neurospora crassa} and \textit{Embericella nidulans} possess both cysteine synthesis pathways.

Some studies have revealed that yeasts transform sulfur amino acids into characteristic volatile sulfur compounds (VSC) that have a considerable impact on food flavor, notably in wine (Perpète \textit{et al.}, 2006) and cheese (Bonaiti \textit{et al.}, 2005). Sulfur metabolism has not been extensively studied in the hemiascomycete phylum, except in \textit{S. cerevisiae}. As the species of this phylum, which have similar morphology and unicellular lifestyle, display more sequence divergences than in the whole chordates taxon (Dujon, 2006), a comparative genomic study was undertaken to provide an insight into the evolution of the sulfur pathways in hemiascomycetes.

Materials and methods

Sulfur metabolism data

Data on sulfur metabolism from several databases including the \textit{Saccharomyces} Genome Database (SGD: \url{http://www.yeastgenome.org/}) for \textit{S. cerevisiae} and the KEGG database (\url{http://www.genome.jp/kegg/}), which regroups enzymatic activities of various species, were combined.

Search for orthologues

Orthologues were searched for incomplete genome databases using the BLASTP algorithm with the \textit{S. cerevisiae} and...
the S. pombe protein sequences of interest as bait. Searches were completed using sequences from other genomes as bait. Further searches were performed on complete genome sequences using TBLASTN. The sources of complete sequences were SGD (http://www.yeastgenome.org/) for S. cerevisiae, Genolevures (http://www.genolevures.org/) for Yarrowia lipolytica, Kluyveromyces lactis, Zygosaccharomyces rouxii, Lachancea kluyveri, Kluyveromyces thermotolerans, Debaryomyces hansenii and Candida glabrata, NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi) and Swiss-prot (http://www.expasy.ch/tools/blast/) for Eremothecium gossypii, Pichia stipitis, S. cerevisiae, Candida albicans, S. pombe, N. crassa and E. nidulans. Pichia stipitis is now recorded as Scheffersomyces stipitis. Complementary information was also obtained from the Candida Genome Database (http://www.candidagenome.org/), from the Aspergillus Genome Database (http://www.aspergillusgenome.org/) and from the Saccharomyces Resequencing Project (Sanger Institute; http://www.sanger.ac.uk/Teams/Team118/sgrp/). The search for specific functional domains was performed using the Pfam database (http://pfam.sanger.ac.uk/). The TARGETP tool (http://www.cbs.dtu.dk/services/TargetP/) was used to identify proteins that have a mitochondrial location.

**Phylogenetic analysis**

Primary protein sequence alignments were generated using the MUSCLE program (Edgar, 2004) implemented on the platform phylogeny.fr (http://www.phylogeny.fr; Dereeper et al., 2008), unless otherwise stated. Alignments were manually adjusted with GENEDOC (http://www.nrbsc.org/gfx/genedoc/). Phylogenetic trees were reconstructed with the PHYML program (Guindon & Gascuel, 2003) implemented on the platform phylogeny.fr, with 100 bootstrap replicates. Phylogenetic trees were visualized with NJ-Plot (Perrière & Gouy, 1996).

**Regulatory protein analysis**

The data obtained for the regulatory proteins analyzed were generated with the GAP software of GCG (gap weight = 2; length weight = 8).

**Results and discussion**

Using BLASTP and TBLASTN, the annotated genes involved in sulfur metabolism of S. cerevisiae and of the archaesomy-cete S. pombe were used as bait against 10 complete and carefully annotated genomes of hemiascomycetes (Fig. 1), which cover the whole hemiascomycete evolutionary tree (Dujon, 2006): Y. lipolytica, K. lactis, Z. rouxii, L. kluyveri, K. thermotolerans, D. hansenii, C. glabrata, E. gossypii, S. (Pichia) stipitis, and C. albicans. We also searched for genes involved in sulfur metabolism in two euascomycetes, N. crassa and E. nidulans (Fitzpatrick et al., 2006). The genes of these two filamentous fungi were used both as outgroups and as internal controls of the distribution of sulfur metabolism genes in ascomycetes, S. pombe being somewhat distantly related to the hemiascomycetes.

Unless otherwise stated, the names of genes and proteins refer to S. cerevisiae. For each gene studied, information on orthologues is available in Supporting Information, Appendices S1 and S2. It is necessary to take into account that the ancestor of S. cerevisiae and C. glabrata has undergone a whole-genome duplication (WGD), followed by massive differential gene loss. The genes duplicated during this event are called ohnologues (Byrne & Wolfe, 2005). As C. albicans is a diploid yeast, only one allele was utilized so as not to destabilize the phylogenetic trees obtained.

In this paper, we primarily deal with three topics: (1) sulfur metabolism on the basis of three axes (sulfate assimilation, sulfur amino acid biosynthesis, the methionine salvage pathway); (2) the sulfur amino acid catabolic pathway; and (3) regulatory proteins.

**Sulfate assimilation**

The first steps of sulfate assimilation

Sulfate cannot be directly used by S. cerevisiae (Thomas & Surdin-Kerjan, 1997). Instead, sulfate needs to be transformed into sulfide to be incorporated into a carbon chain. However, this reaction requires several steps. In fact, the electropotential of sulfate is too strong to be directly reduced.
by NADPH and therefore sulfate has to be activated in two steps by the ATP sulfurylase, and the adenylyl sulfate (APS) kinase (Fig. 2). The first enzyme (encoded by MET3) catalyzes the formation of APS from ATP and sulfate and the second (encoded by MET14) carries out the phosphorylation of APS to yield PAPS (3′-phospho-5′-adenyl sulfate). The latter compound has sufficient electropotential to be reduced by NADPH. Although the ATP sulfurylase sequences are tightly conserved among the species considered, *Y. lipolytica* ATP sulfurylase displays a divergent C-terminal end (from Gly426 to the end Asn572) from the other yeasts. The APS kinase-related genes are highly conserved.

The reduction of PAPS involves two enzymes that act successively. These enzymes (Fig. 2) are PAPS reductase (encoded by MET16), which produces sulfite, and sulfite reductase, which leads to the production of sulfide. The fungal PAPS reductase enzymes present a divergent N-terminal sequence compared with that of yeasts, including an insertion of about 30 amino acids. The sulfite reductase is composed of two subunits (encoded by MET5 and MET10) (Fig. 2). As PAPS is toxic when it is accumulated in the cell, its concentration is regulated by the reverse reaction (Fig. 2) catalyzed by the bisphosphate-3′-nucleotidase (encoded by MET22). The characteristics of the *D. hansenii* orthologue were extensively studied by Aggarwal et al. (2008), who linked this enzyme to the high halotolerance of *D. hansenii*. *Candida albicans* displays a duplication of the gene MET22. Besides these peculiarities, all these genes could be found in the studied species, underlining the importance of this system for yeasts.

**Fig. 2.** Sulfate assimilation in hemiascomycetous yeasts. The gene names indicated in this figure are those of *Saccharomyces cerevisiae*. The function of these genes is indicated in Appendix S2. This figure depicts the essential steps for the incorporation of sulfate into a carbon chain, leading to homocysteine production.

**The production of homocysteine**

Homoserine corresponds to a cross-roads between sulfur and threonine metabolism. Three enzymes lead to homoserine production from aspartate (Fig. 2): the aspartate kinase (encoded by HOM3), the aspartic semi-aldehyde dehydrogenase (encoded by HOM2) and the homoserine dehydrogenase (encoded by HOM6). The synthesis of homocysteine from homoserine requires the action of two enzymes. First, the L-homoserine-O-acetyltransferase (encoded by MET2) synthesizes O-acetyl-homoserine. Then, the O-acetyl homoserine sulphydrylase (encoded by MET17) catalyzes the reaction between O-acetyl-homoserine and sulfide, which leads to the production of homocysteine (Fig. 2). The good conservation of this pathway in all the organisms studied demonstrates the importance of homocysteine, which is the base of the biosynthesis of sulfur amino acids. Only *C. albicans* displays a duplication of the gene HOM3.

**Siroheme biosynthesis**

Met10/Met5 sulfite reductase needs siroheme as a cofactor in order to function (Hansen et al., 1997), as this heme is the only molecule known to react with sulfite. In fact, siroheme possesses the exceptional capacity to perform the six-electron reduction of sulfite to obtain sulfide (Murphy et al., 1974). The enzymes (Fig. 2) involved in siroheme biosynthesis act on uroporphyrinogen III to produce siroheme in three successive steps (Raux et al., 1999). First, MET1, which encodes the S-adenosyl-L-methionine uroporphyrinogen III
transmethylase activity, produces precorrin-2. Secondly, MET8, which encodes the dehydrogenase and chelatase activities, produces sirohydrochlorin. Finally, siroheme is produced. All the genes coding for these enzymes were found in our species of interest, MET1 being duplicated in *D. hansenii*.

**NADPH supply**

The glucose-6-phosphate dehydrogenase (encoded by *MET19*) (Fig. 2), which is the first enzyme of the pentose-phosphate pathway, was previously recorded as an enzyme involved in methionine metabolism because its inactivation leads to methionine auxotrophy (Masselot & De Robichon-Szulmajster, 1975). A *met19* mutant strain of *S. cerevisiae* requires an organic sulfur source such as methionine, SAM, cysteine, glutathione or homocysteine. The Met19 glucose-6-phosphate dehydrogenase produces reduced NADPH, which is essential to the function of Met16 and Met5-10 enzymes (Thomas *et al.*, 1991). This key enzyme of the pentose-phosphate pathway and sulfur activation is highly conserved in the species tested. It is noteworthy that *S. pombe* displays three glucose-6-phosphate dehydrogenases.

**Sulfur amino acid biosynthesis**

In *S. cerevisiae*, homocysteine is the central molecule in the biosynthesis of sulfur amino acids, as it is (1) the starting point of the synthesis of cysteine by the transsulfuration pathway and (2) the substrate of cobalamin-independent methionine synthase (Suliman *et al.*, 2005), which produces methionine (Fig. 3).

**The transsulfuration pathway**

Two major pathways lead to cysteine: the transsulfuration pathway and the OAS pathway. *Schizosaccharomyces pombe* has an incomplete transsulfuration pathway (Brzywczy *et al.*, 2002). Only the synthesis of homocysteine from cysteine takes place in this organism, not the reverse reaction. Therefore, the supply of cysteine in *S. pombe* relies on the OAS pathway. The OAS pathway (Fig. 3, in gray) is composed of two enzymes, the serine-O-acetyltransferase and the cysteine synthase (encoded by YGR012w in *S. cerevisiae*). There are two cysteine synthases in all of the organisms studied, but *D. hansenii*, *C. glabrata* and *S. cerevisiae* have only one. The cysteine synthases can be separated into two groups, the first, which is present in all the organisms studied, and the second, which is less conserved (Brzywczy *et al.*, 2007). Interestingly, the gene coding for the serine-O-acetyltransferase enzyme is present in all the organisms studied except *S. cerevisiae* and *C. glabrata*, the two species that underwent WGD. This could be the result of a concerted loss, as, for example, already observed for specific pathways in *C. glabrata* (Dujon *et al.*, 2004). This enzyme is very different from serine-

![Fig. 3. Methionine and cysteine synthesis in hemiascomycetous yeasts: similarities and divergences. The gene names indicated in this figure are those of *Saccharomyces cerevisiae*. The function of these genes is indicated in Appendix S2. This figure represents methionine biosynthesis via the methyl cycle and cysteine biosynthesis via the transsulfuration pathway. The pathway that leads to glutathione production is also indicated. The enzyme of the OAS pathway that is absent in *S. cerevisiae* and *Candida glabrata* is shown in gray.](https://academic.oup.com/femsyr/article-abstract/11/4/366/564359)
O-acetyltransferases with a homology of bacterial type and presents a strong sequence similarity with homoserine-O-acetyltransferases (Fig. 4) (Grynberg et al., 2000). This result indicates that the only functional pathway for the cysteine synthesis in *S. cerevisiae* and *C. glabrata* is the transulfuration pathway, in agreement with the genetic work of Cherest & Surdin-Kerjan (1992). To complete this study, homologues of the O-acetylserine transferase gene of *S. pombe* (SPBC106.17c) was searched for in the 37 strains of *S. cerevisiae* and the 37 strains of *Saccharomyces paradoxus*, which were sequenced by the *Saccharomyces* Resequencing Project at the Sanger Institute (Liti et al., 2009). We did not find orthologues in any of these strains. There is a great deal of literature concerning this hypothetical pathway in *S. cerevisiae*, with controversial results. A serine-O-acetyltransferase activity was described in *S. cerevisiae* (Ono et al., 1999) in crude cell extract and was, consequently, not associated with a gene. Cherest & Surdin-Kerjan (1992) demonstrated that the OAS pathway is not functional in *S. cerevisiae*. An *in vitro* O-acetylserine transferase activity could be due to a secondary activity of the O-acetylhomoserine transferase catalyzed by Met2 without physiological significance.

The transulfuration pathway leads to the production of cysteine from homocysteine and the reverse reaction, via cystathionine synthesis (Fig. 3). The two enzymes involved in cysteine biosynthesis are the cystathionine β-synthase (encoded by *CYS4*) and the cystathionine γ-lyase (encoded by *CYS3*). These enzymes are very well conserved in all of the organisms studied, except *S. pombe*, as already shown.

The reverse reactions are successively due to a cystathionine γ-synthase (encoded by *STR2*) and a cystathionine β-lyase (encoded by *STR3*). These genes are all present in the studied species and interestingly, the *STR2*-like gene is

---

**Fig. 4.** Phylogram of homoserine-O-acetyltransferases and serine-O-acetyltransferases in hemiascomycetous yeasts. The gene tree was constructed as described in Materials and methods. All positions containing gaps and missing data were eliminated from the dataset. The *Saccharomyces cerevisiae* and *Candida glabrata* homoserine-O-acetyltransferase genes are boxed. The branch length linking the families is shortened. Scale bar = 0.5 substitutions per site.
duplicated in several species; there are two orthologues in *K. thermotolerans* and three in *N. crassa*. *Candida glabrata* contains two *STR2* ohnologues, whereas *S. cerevisiae* presents three orthologues of *STR2*, of which YJR130C and YML082W are ohnologues. Two *STR3* ohnologues can be found in *S. cerevisiae*.

This group of genes belongs to a protein family sharing a common motif linked to the use of pyridoxal-5′-phosphate as a cofactor which is widespread from bacteria to higher eukaryotes (see Appendix S2). In *S. cerevisiae*, this group is composed of four genes involved in sulfur metabolism (*MET17*, *CYS3*, *STR2* and *STR3*) and one ORF with no associated function *YHR112c* (see Appendix S1). The *YHR112c* gene family is well conserved, with a sequence similarity comprised between 43% and 73% over 98% of the sequence length, and it can therefore be inferred that it is subjected to a strong selection pressure. This gene is not essential in *S. cerevisiae* but it could be involved in sulfur metabolism under at least unidentified growth conditions. This gene is duplicated in *C. glabrata*, *Z. rouxii*, *D. hansenii* and *E. nidulans* further stressing its importance in the cell.

**Glutathione synthesis**

Sulfur metabolism leads to the production of a key molecule involved in the defense against oxidative stress, glutathione, from cysteine by way of γ-glutamylcysteine (Fig. 3). This pathway is composed of the γ-glutamylcysteine synthetase (encoded by *GSH1*) and the glutathione synthetase (encoded by *GSH2*), both present in the studied organisms.

**Methionine synthesis**

Methionine is synthesized by cobalamin-independent methionine synthase (encoded by *MET7*). To be functional, this enzyme needs the presence of polyglutamate on the methylytetrahydrofolate. The acquisition of polyglutamate on methylytetrahydrofolate involves four enzymes: the dihydrofolate reductase, two methylene tetrahydrofolate reductases and the folylpolyglutamate reductase encoded respectively by *FOL3*, *MET12*, *MET13* and *MET7* (Fig. 3) (Cherest et al., 2000).

Another important feature of sulfur metabolism is that methionine belongs to the methyl cycle (Fig. 3), which produces SAM, the major donor of methyl in *S. cerevisiae* (Katz et al., 2003). In *S. cerevisiae*, the methionine/SAM ratio is controlled (Thomas et al., 2000) by *MHT1* encoding the S-methylmethionine-homocysteine methyltransferase and *SAM4* encoding the S-adenosylmethionine-homocysteine methyltransferase, which have been described as ohnologues. During methyl transfer reactions, SAM is transformed into S-adenosylhomocysteine. The latter molecule is transformed into homocysteine by S-adenosyl-L-homocysteine hydrolase (encoded by *SAH1*), ending the methyl cycle. The ensemble of these genes can be found in one copy in the fungal species tested. However, *S. cerevisiae* carries two *FOL3* ohnologues. *Saccharomyces cerevisiae* and *C. glabrata* have two ohnologues encoding S-adenosylmethionine synthetases. *Candida glabrata* does not possess the regulators *MHT1* and *SAM4*.

**Methionine salvage pathway**

SAM is involved in polyamine biosynthesis (Fig. 5). It is decarboxylated by the S-adenosylmethionine decarboxylase (encoded by *SPE2*) to yield S-adenosylmethioninamine, which will be transformed into spermidine and spermine by the spermidine synthase and the spermine synthase encoded by *SPE3* and *SPE4*, respectively. Orthologues of *SPE3* are present in all the microorganisms studied, although *SPE4* was not found in *E. gossypii*, *N. crassa* or *E. nidulans*. In *S. cerevisiae*, spermidine is a polyamine essential for growth (Hamasaki-Katagiri et al., 1997). In addition to the production of spermidine and spermine, Spe3p and Spe4p produce methylthioadenosine, which is recycled into methionine by the methionine salvage pathway.

The enzymes of this pathway were recently studied in detail by Pirkov et al. (2008) in *S. cerevisiae* (Fig. 5). The methionine salvage pathway is composed of five enzymes (encoded by *MEU1*, *MR11*, *MDE1*, *UTR4* and *AD11*), leading to the production of keto-methyl thio butyrate (KMBA) from 5-methylthioadenosine. Each species has orthologues for each reaction of the cycle except *E. nidulans*, which lacks the enzyme encoded by *UTR4*.

The last step of the methionine salvage pathway, which produces methionine from KMBA, is carried out by a nonspecific transaminase. On the basis of the results of Pirkov et al. (2008), the branched-chain amino acid aminotransferases and aromatic amino acid aminotransferases are involved in this biosynthetic step as well as in the methionine catabolic pathway. As all the enzymes of the methionine salvage pathway have a cytoplasmic location, the cytoplasmic aminotransferases should be involved in this metabolism, but not the mitochondrial ones.

**Sulfur amino acid catabolism/VSC production**

The catabolism of sulfur amino acids has been less extensively studied; nevertheless, it produces VSC that are often important ‘character impact’ compounds in many fermented foods such as beer, wine and smear-ripened cheese. Previous studies were mainly focused on the identification and aromatic characterization of the sulfur compounds involved in the aroma of these products. However, their biosynthetic pathways are far from being understood (Fig. 5). Moreover, some steps seem to be purely
spontaneous. A comprehensive review of these aspects has been published recently (Landaud et al., 2008).

Despite our lack of knowledge, it is generally accepted that the catabolism of sulfur amino acids is initiated by a transamination step. Nevertheless, no specific aminotransferase for methionine has been identified in yeasts. This amino acid is therefore a substrate of enzymes with a fuzzy specificity. Consequently, the effects of various transaminases have previously been studied in our laboratory (Bondar et al., 2005; Kagkli et al., 2006; Cholet et al., 2008). It was demonstrated that methionine can be a substrate for branched-chain amino acid aminotransferases and aromatic amino acid aminotransferases, and that an overexpression of some of these genes leads to an increased production of VSC.

In *S. cerevisiae*, two branched-chain amino acid aminotransferases, the mitochondrial Bat1 and the cytoplasmic Bat2, were described. Phylogenetic analysis of all the Bat1 and Bat2 could not differentiate two clear groups (data not shown). We consequently took advantage of the presence/absence of the mitochondrial targeting sequence in this protein family to classify these proteins. Like *S. cerevisiae*, *C. glabrata*, *C. albicans* and *Y. lipolytica* display two enzymes with a different putative location. *Kluyveromyces thermotolerans*, *K. lactis*, *Z. rouxii*, *D. hansenii* and *L. kluyveri* only have a single enzyme with a mitochondrial targeting sequence. BAT2 of *L. kluyveri* was annotated as a pseudogene. *Eremothecium gossypii*, *D. hansenii* and *S. (Pichia) stipitis* only have the cytoplasmic enzyme Bat2p, the latter species carrying a duplication of this gene. The evolution of these genes in fungi is complex and needs to be extensively explored in relation to the functions they are associated with.

In *S. cerevisiae*, there are two cytoplasmic aromatic amino acid aminotransferases, Aro8 and Aro9. Phylogenetic analysis of aromatic aminotransferases differentiated three groups of proteins, one of them being associated to the *S. cerevisiae* ORF with no known function, YER152c. This group comprises potential aromatic amino acid aminotransferases. This gene is well conserved, with one copy in each genome and a duplication in *K. thermotolerans*. The strong conservation of this gene leads us to infer that this gene might play an important role under specific conditions. All these aromatic transaminases seem to be cytoplasmic except the Aro9 enzyme in *Y. lipolytica*, which presents a putative signal sequence for mitochondrial location. In the ARO8 group, a duplication event apparently occurred after the divergence of the group constituted of *E. gossypii*, *K. lactis*, *K. thermotolerans* and *L. kluyveri*, but just before the divergence of these species from each other, as there are two ARO8 orthologues in each of them (Fig. 6). At least one copy of ARO8 is in fact present in all the tested species. The presence of the gene ARO9 seems to follow different routes. In fact, this gene has been lost in two species, *E. gossypii* and *D. hansenii*, and has been duplicated in *S. stipitis*.

We identified three combinations of the five enzymes described above (Bat1, Bat2, Aro8, Aro9, Yer152c) in the hemiascomycetes. The first, which has all five of these enzymes, is found in *S. cerevisiae*, *C. glabrata*, *S. stipitis*, *C. albicans* and *Y. lipolytica*. The second, observed in *Z.
**rouxii**, *K. thermotolerans*, *L. kluyveri* and *K. lactis*, has all the enzymes except Bat2. The third lacks Bat1 and Aro9 and is found in *E. gossypii* and *D. hansenii*. Thus, two enzymes are always present in all these yeasts – Aro8 and Yer152c – stressing the importance of the aminotransferase Aro8 and of the protein of unknown function, Yer152c. We observed that each of the three technological yeasts, *K. lactis*, *D. hansenii* and *Y. lipolytica*, have a different aminotransferase profile. This could be indicative of different potencies for the biosynthesis of VSC in these three yeasts.

Methional is an important molecule for cheese flavor. This sulfur compound is a product of methionine degradation from KMBA, the first intermediary of methionine catabolism. In *S. cerevisiae*, the phenylpyruvate decarboxylase encoded by *ARO10* is thought to be responsible for this reaction (Vuralhan et al., 2003, 2005).

**Fig. 6.** Phylogram of the aminotransferases of the Aro8p and Aro9p families and of the related Yer152cp-like family in hemiascomycetous yeasts. The gene tree was constructed as described in Materials and methods. All positions containing gaps and missing data were eliminated from the dataset. The *Saccharomyces cerevisiae* proteins are boxed. Scale bar = 0.5 substitutions per site.
This enzyme belongs to a large family of decarboxylases including the pyruvate decarboxylases. The gene YAL10D 6930g clearly clusters with the fungal phenylpyruvate decarboxylase subfamily.

Yarrowia lipolytica possesses one ORF YAL10D10131g that produces a good alignment with the three genes of the pyruvate decarboxylases of S. cerevisiae (PDC1, PDC5 and PDC6). This enzyme has been described as an orthologue of PDC1 but seems to be closer to PDC6 according to this study (data not shown).

Regulatory proteins

In S. cerevisiae, the main regulator of transcription Met4 is regulated by the cysteine pool (Hansen & Johannesen, 2000). This regulator not only plays a central role in the regulation of the sulfur amino acid biosynthetic pathway, but in the cell cycle as well. It is assisted in its functions by a number of regulatory cofactors: Met28, Met31/Met32, Cbf1 and SCF^{Met30}. The latter is an ubiquitination complex composed of Skp1, Cdc53, Cdc34, Rbx1 and Met30 as specific components. Models of regulation of cell metabolism by Met4 and its partners have been proposed (Chandrakasan & Skowyra, 2008).

The majority of the components of the SCF^{Met30} complex are highly conserved (Table 1). It can be observed that there is no orthologue of RBX1 in the N. crassa genome. CDC53 is slightly less conserved than other components of the SCF^{Met30} complex, with sequence similarities in the same range as that of MET31/MET32. MET31 and MET32 are ohnologues and are also retrieved in C. glabrata. MET31/MET32 appears to be absent or too divergent in the S. pombe genome. CBF1 is conserved in all species studied, but the BLASTP analysis reveals a sequence alignment only in the C-terminal region.

In S. cerevisiae, Met4 is a 672 amino acid long polypeptide that can be split into functional domains (Kuras & Thomas, 1995). These domains are involved in transcriptional activation, ubiquitination, interaction with the negative regulator Met30, interaction with the Met31/Met32 dimer and interaction with Met28. The Met4 sequence also presents a degenerated leucine zipper (BZIP). This BZIP motif is not canonical, as it is disrupted by a large insertion (D504 to Q599).

Met4 orthologues display large differences among the species studied. The family of MET4-related genes can be separated into two subfamilies: a family of large polypeptides related to S. cerevisiae Met4 of about 670–710 amino acids, and a family of shorter polypeptides of 319–371 amino acids, including D. hansenii and Y. lipolytica. This second family is related to fungal regulators such as Cys3 and MetR from N. crassa and E. nidulans, respectively. We can raise the question whether or not these orthologues carry all the interaction domains such as Met4 from S. cerevisiae.

One argument for the conservation of these interactions is that the partners of Met4 are globally conserved except for Met28, which appeared to be absent in the clade of the short-length Met4 subfamily. However, the shorter Met4 sequences display a canonical BZIP motif with a basic region, allowing direct interaction with DNA, and a leucine zipper required for dimerization. It can be hypothesized that the shorter Met4 homologues with a complete BZIP domain can directly interact with DNA but do not depend on Met28 function. Therefore, Met28, which is not conserved in the studied species, appears to have an ancillary function.

Table 1. Summary of regulatory proteins studied

<table>
<thead>
<tr>
<th></th>
<th>MET4</th>
<th>MET28</th>
<th>MET31/MET32</th>
<th>CBF1</th>
<th>MET30</th>
<th>SKP1</th>
<th>RBX1</th>
<th>CDC53</th>
<th>CDC34</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. nidulans</td>
<td>25</td>
<td>–</td>
<td>27</td>
<td>25</td>
<td>46</td>
<td>61</td>
<td>53</td>
<td>30</td>
<td>52</td>
</tr>
<tr>
<td>N. crassa</td>
<td>26</td>
<td>–</td>
<td>31</td>
<td>30</td>
<td>43</td>
<td>62</td>
<td>–</td>
<td>29</td>
<td>48</td>
</tr>
<tr>
<td>S. pombe</td>
<td>23</td>
<td>–</td>
<td>–</td>
<td>30</td>
<td>45</td>
<td>61</td>
<td>55</td>
<td>30</td>
<td>47</td>
</tr>
<tr>
<td>Y. lipolytica</td>
<td>22</td>
<td>–</td>
<td>28</td>
<td>24</td>
<td>59</td>
<td>65</td>
<td>53</td>
<td>28</td>
<td>62</td>
</tr>
<tr>
<td>D. hansenii</td>
<td>28</td>
<td>–</td>
<td>34</td>
<td>34</td>
<td>55</td>
<td>70</td>
<td>57</td>
<td>28</td>
<td>64</td>
</tr>
<tr>
<td>C. albicans</td>
<td>28</td>
<td>–</td>
<td>37</td>
<td>39</td>
<td>55</td>
<td>67</td>
<td>59</td>
<td>27</td>
<td>64</td>
</tr>
<tr>
<td>P. stipitis</td>
<td>28</td>
<td>25</td>
<td>35</td>
<td>35</td>
<td>57</td>
<td>70</td>
<td>60</td>
<td>28</td>
<td>65</td>
</tr>
<tr>
<td>K. lactis</td>
<td>32</td>
<td>46</td>
<td>37</td>
<td>41</td>
<td>66</td>
<td>83</td>
<td>65</td>
<td>32</td>
<td>68</td>
</tr>
<tr>
<td>K. thermotolerans</td>
<td>36</td>
<td>50</td>
<td>52</td>
<td>52</td>
<td>66</td>
<td>85</td>
<td>82</td>
<td>37</td>
<td>77</td>
</tr>
<tr>
<td>E. gossypii</td>
<td>36</td>
<td>19</td>
<td>51</td>
<td>48</td>
<td>67</td>
<td>87</td>
<td>69</td>
<td>29</td>
<td>82</td>
</tr>
<tr>
<td>L. kluyveri</td>
<td>41</td>
<td>21</td>
<td>50</td>
<td>48</td>
<td>66</td>
<td>90</td>
<td>77</td>
<td>38</td>
<td>80</td>
</tr>
<tr>
<td>Z. rouxii</td>
<td>44</td>
<td>47</td>
<td>51</td>
<td>43</td>
<td>69</td>
<td>92</td>
<td>82</td>
<td>42</td>
<td>84</td>
</tr>
<tr>
<td>C. glabrata</td>
<td>40</td>
<td>50</td>
<td>32/40</td>
<td>30</td>
<td>73</td>
<td>89</td>
<td>85</td>
<td>40</td>
<td>82</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>100</td>
<td>100</td>
<td>100/45</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

This table represents the sequence similarity percentages for each regulatory protein studied, in comparison with Saccharomyces cerevisiae. The percentages indicated for the MET31/MET32 column were generated using the Met31p sequence. The analysis was carried out using the GAP software from Accelrys (Accelrys Inc.)
The activation motif of S. cerevisiae (D82 to F100), which is highly conserved in neighbor species, is retrieved in D. hansenii (D23 to L41) but not in Y. lipolytica. Conversely, the motif upstream from the ubiquitinated lysine (K163 in S. cerevisiae) appeared to be well conserved among yeast species by homology and by hydrophobic cluster analysis (data not shown). The Met31/Met32-binding domain of S. cerevisiae (L376 to H397) matches relatively well the L182 to Q203 region of D. hansenii. However, it is difficult to identify the sites of interaction with these regulators when distant Met4 orthologues are analyzed. For instance, in Y. lipolytica, it only can be hypothesized that the domain from Q193 to Q205 is related to the domain from Q191 to Q203 in D. hansenii.

In conclusion, it seems that the coregulation network depicted in S. cerevisiae is conserved in all the yeast species, although there is a very large variation in size and a wide divergence in the main component, Met4. It is clear that these hypotheses should be validated by mutagenesis experiments.

Conclusions

An overview of the sulfur amino acid pathways in hemiascomycetous yeasts was generated. It was important to investigate whether the wealth of data accumulated in the model yeast S. cerevisiae can be extended to species as distant from S. cerevisiae as Y. lipolytica, which diverged from their common ancestor more than 600 million years ago. By and large, the pathways appeared to be conserved even beyond the yeast world, in the whole phylum of fungi. This is probably foreseeable as this pathway plays a central role in the general metabolism and homeostasis of the cell.

Moreover, we propose that the regulation network in charge of the sulfur amino acid biosynthesis and its central component Met4 are conserved from S. cerevisiae to fungi. The Kluyveromyces clade and the Saccharomyces sensu lato have undergone a genetic event leading to the impairment of the BZIP motif of the regulator and need therefore the mediation of a BZIP partner of the Met28 family to bind the Met4 DNA targets. All the other partners appeared to be conserved along the evolutionary tree, especially Met31 and Met30, which play a central role in the specificity of the regulation.

The promoters of the MET regulon display in Saccharomyces, conserved binding sites for Cbf1 and Met31/Met32, which are therefore landmarks of membership to this regulon. It is tempting to investigate a conservation of these binding sites along the evolutionary tree. The Cbf1-binding hexamer CACGTG appeared to be widespread in the promoters of S. cerevisiae (in 592 promoters, i.e. about 10% of the genes, according to Kent et al., 2004). This binding site is shared with several transcription factors: Pho4, which belongs to the phosphate regulon (Gonze et al., 2005), Rtg3, a TOR-controlled transcription activator, Tye7, a transcription factor involved in the activation of glycolytic enzymes (Zhu et al., 2009). Moreover, Cormier et al. (2010) have recently studied the PDC6 promoter which belongs to a subclass of promoters responding to a starvation in sulfur amino acids induced by cadmium stress (Fauccon et al., 2002). This effect was shown to be fully dependent on the presence of the transcription factors of the Met regulon: Met4, Met32 and Cbf1. For Cbf1, no consensus CACGTG was found in the promoter area, only a related consensus TCACGTGT, whose deletion does not block the activation. The specificity of the interaction could be carried out through the cooperative binding of other transcription factors, for instance, Met31/Met32 for the Met regulon. The Met31/Met32-binding site was found to be AAACCTGTGGC (Blaiseau & Thomas, 1998). According to Cormier et al. (2010), the Met32 target is restricted to CACGTGGC in the PDC6 promoter. These data from S. cerevisiae clearly show that the rules of the syntax of the binding on promoter by the transcription factors are far from being understood. However, using the Meme (Bailey & Elkan, 1994) and RSAT (Thomas-Chollier et al., 2008) web services, we have found that in a set of genes belonging to the MET regulon in S. cerevisiae (MET1, MET2, MET3, MET6, MET14, MET17, MET28, CYS3, CYS4 and SAM2), such Cbf1 and Met31/Met32 sites can be found in close vicinity to each other. But we were unable to find the same combination of sites in a similar set of genes from K. lactis, D. hansenii or Y. lipolytica. Cbf1-like sites were easily found but they are very frequent in these genomes, as in S. cerevisiae. The Met31/Met32-binding site has probably also diverged and is therefore not easily identifiable. According to our hypothesis, Met4 homologues could bind directly to the DNA without Met28 help. In N. crassa, the DNA-binding specificity of the transcription factor Cys3 has been established to be ATGGGCCCAT (Paietta, 2008). We have therefore investigated the presence of such a binding site in the promoters of the MET regulon in Yarrowia. Hits could be obtained only if we used a more degenerated consensus WKRYRYMW. It is therefore difficult to give a firm answer on the specificity of the interaction between Met4 homologues and their specific promoters without experimental studies.

The majority of the studied species possess two pathways leading to cysteine production, the transsulfuration pathway and the OAS pathway, except S. cerevisiae and C. glabrata, which only display the transsulfuration pathway, and S. pombe, which only has the OAS pathway (Table 2).

The study of the gene families involved in this metabolism were conducted to identify well-conserved gene families without described function, although they are highly conserved and they diverge at the same speed as the essential genes of the pathway. This is the case for YHR112c gene.
family belonging to the family of the PLP-dependent enzymes such as the CY3, STR2, STR3 and MET17 gene families. This is also the case in the catabolic pathway for the YER152c gene family related to the aromatic amino acid aminotransferase family without identified function in S. cerevisiae. Their conservation during evolution suggests an important role for them in the metabolism of the sulfur compounds.

Concerning the catabolism of sulfur amino acids that leads to the production of VSC, the technological yeasts K. lactis, D. hansenii and Y. lipolytica have different sets of enzymes, which should account for the difference in aroma production observed between these species. The results of this work remain to be confirmed by genetic and molecular studies.

Acknowledgements
We would like to thank Sophie Landaud for helpful comments on the manuscript, and the PNRA (French National Research Agency: http://www.agence-nationale-recherche.fr) for the financing of the EcoMet project (ANR-06-PNRA-014).

Authors’ contributions
A.H. was responsible for the initial phylogenetic studies, tree analysis, manuscript and figure preparation. J.M.B. generated and analyzed the results on regulatory proteins, and participated in the writing of the manuscript. S.C. provided his expertise for the data analysis and the manuscript editing. All authors read and approved the manuscript.

References


Table 2. Summary of the components of the sulfur amino acids pathway highlighting the differences in cysteine synthesis according to the yeast clades

<table>
<thead>
<tr>
<th>Yeast Clade</th>
<th>Pathway</th>
<th>Transsulfuration reaction</th>
<th>OAS pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>WGD group (S. cerevisiae/C. glabrata)</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Other hemiascomycetes</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ascomycetes fungi</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

Appendix S1. List of genes involved in sulfur metabolism in each species studied.

Appendix S2. Table of proteins studied.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.