Available online at www.sciencedirect.com
ScienceDirect
Genomics 88 (2006) 387–393
GENOMICS
www.elsevier.com/locate/ygeno

# Conflicting phylogenetic position of *Schizosaccharomyces pombe*

Eiko E. Kuramae [a,*], Vincent Robert [a], Berend Snel [b], Teun Boekhout [a,c]

[a] *Centraalbureau voor Schimmelcultures, Comparative Genomics and Bioinformatics, Uppsalalaan 8, 3584 Utrecht, The Netherlands*
[b] *Nijmegen Center for Molecular Life Sciences, University Medical Center St. Radboud, pa CMBI, Toernooiveld 1, 6526 ED Nijmegen, The Netherlands*
[c] *University Medical Center, Department of Medicine, Division of Acute Medicine and Infectious Diseases, Utrecht, The Netherlands*

Received 7 February 2006; accepted 6 July 2006
Available online 9 August 2006

**Abstract**

The phylogenetic position of the fission yeast *Schizosaccharomyces pombe* in the fungal Tree of Life is still controversial. Three alternative phylogenetic positions have been proposed in the literature, namely (1) a position basal to the Hemiascomycetes and Euascomycetes, (2) a position as a sister group to the Euascomycetes with the Hemiascomycetes as a basal branch, or (3) a sister group to the Hemiascomycetes with Euascomycetes as a basal branch. Here we compared 91 clusters of orthologous proteins containing a single orthologue that are shared by 19 eukaryote genomes. The major part of these 91 orthologues supports a phylogenetic position of *S. pombe* as a basal lineage among the Ascomycota, thus supporting the second proposition. Interestingly, part of the orthologous proteins supported a fourth, not yet described alternative, in which *S. pombe* is basal to both Basidiomycota and Ascomycota. Both topologies of phylogenetic trees are well supported. We believe that both reflect correctly the phylogenetic history of the species concerned. This apparent paradox may point to a heterogeneous nuclear genome of the fungi. Importantly, this needs to be taken in consideration for a correct understanding of the fungal Tree of Life.


*Keywords:* KOG database; Comparative genomics; Fungal phylogeny

*Schizosaccharomyces pombe* was the first fission yeast species to be discovered and it formed the basis of the "fission yeast" genus *Schizosaccharomyces* [1]. The species is used widely as a model organism in molecular and cellular biology. It is a unicellular fermentative eukaryote, with short cylindrical cells that maintain their shape by growing at the cell tips and divide by medial fission to produce two daughter cells of equal size [2]. Phylogenetically, *S. pombe*, together with *Taphrina*, *Protomyces*, *Saitoella*, and *Pneumocystis*, has been classified in the Archiascomycetes, which are believed to represent an ancestral assembly within the Ascomycota [3].

The genome of *S. pombe* has been sequenced and comprises three chromosomes [4]. The fission yeast is evolutionarily considerably diverged from the budding yeast *Saccharomyces cerevisiae*. For instance, the species has no large genome duplication of the type that occurred in budding yeasts [4].

The phylogenetic position of *S. pombe* is a point of ongoing discussion. Three possible phylogenetic relationships of *S. pombe* have been hypothesized: (1) Archiascomycetes (*S. pombe*) is a basal lineage to both the Euascomycetes and the Hemiascomycetes, (2) Euascomycetes and Archiascomycetes (*S. pombe*) are sister groups with the Hemiascomycetes as a basal lineage, and (3) Hemiascomycetes and Archiascomycetes (*S. pombe*) are sister groups with the Euascomycetes as a basal lineage (Fig. 1). Hypothesis 1 corresponds to *S. pombe* as a basal lineage within the Ascomycota and is supported by sequence analysis of the nuclear small subunit rRNA [3]. However, only limited statistical support was observed for this phylogenetic tree. Hypothesis 2 suggested by Prillinger et al. [5] and Diezmann et al. [6] is based on 18S rRNA sequences [5] and a combined analysis of 18S and 26S rDNA sequences [6]. Finally, hypothesis 3 was supported by concatenation of mitochondrial genes or proteins [7,8]. To test these three phylogenetic hypotheses, sequence alignments of 91 orthologous proteins shared among 19 complete eukaryote genomes were compared

* Corresponding author. Fax: +31 302512097.
*E-mail address:* kuramae@cbs.knaw.nl (E.E. Kuramae).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.07.001

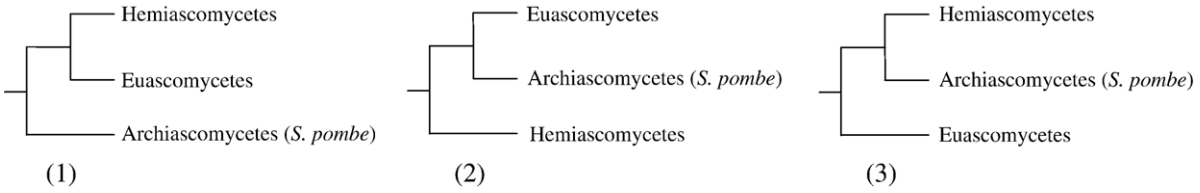


Fig. 1. Three published hypotheses on the phylogenetic position of *Schizosaccharomyces pombe*.

and analyzed to determine the most likely phylogenetic position of *S. pombe*.

## Results and discussion

### *Number of shared KOGs among 19 genomes*

The 19 eukaryotic genomes shared a total of 1250 KOGs (Supplementary Table 1) from a list of 4852 KOGs. Only 91 of these 1250 KOGs were represented by a single protein (Table 1) and were used for comparison and further phylogenetic analyses. Around 50% of these selected proteins belonged to the information storage and processing category (Supplementary Fig. 1). From this category, the majority is involved in RNA processing and modification (38%) and replication, recombination, and repair (27%) (Supplementary Fig. 1).

### *Comparison of shared proteins*

The comparison of 91 protein distance matrices by neighbor joining resulted in a KOG protein tree (Fig. 2). This KOG tree illustrates the relationship among the 91 proteins, and six main clusters (I–VI) could be discerned. The clusters I, II, III, IV, V, and VI are represented by 15, 8, 11, 9, 16, and 32 proteins, respectively. Each protein from each cluster was aligned among the 19 genomes, the gaps were eliminated, and subsequently, the proteins from the same cluster were concatenated for phylogeny reconstruction using maximum parsimony (MP) and quartet puzzling (QP) (see Material and methods). The concatenation of proteins from the various clusters resolved the position of *S. pombe* into two different branches of the phylogenetic tree (Fig. 3). Cluster I proteins placed *S. pombe* as a lineage basal to the phylum Basidiomycota and the classes Euascomycetes and Hemiascomycetes (Fig. 3a), while all other clusters (II–VI) supported the position of *S. pombe* as a lineage basal to Euascomycetes and Hemiascomycetes (Fig. 3b), with the Basidiomycota being more basal. Support for both trees was rather high.

### *Cluster I proteins*

The proteins of cluster I supported *S. pombe* as a lineage basal to the Basidiomycota, Euascomycetes, and Hemiascomycetes with 100% (MP) and 95% (QP) bootstrap support (Fig. 3a). In addition, the support of the basidiomycetous lineage being basal to the Hemiascomycetes and Euascomycetes was relatively high (70% by MP and 87% by QP) (Fig. 3a). Low bootstrap supported occurred for the clustering of the Hemiascomycetes and Euascomycetes with 42% (MP) and 75% (QP) support. Interestingly, this tree topology based on cluster I proteins does not fit to any of the three hypotheses illustrated in Fig. 1. Most of the cluster I proteins are nuclear proteins, except KOG2633 (i.e., tyrosyl-tRNA synthetase) and KOG1119 (involved in the assembly of mitochondrial and cytoplasmic iron–sulfur proteins, localized on mitochondrial matrix), which are mitochondrial proteins. Several of the nuclear cluster I proteins of *S. pombe* are more similar to those of animals than to those of fungi. This may explain the phylogenetic position of *S. pombe* as being closer to the animals than to the fungi.

Some proteins of *S. pombe* have previously been reported to more closely resemble those of animals (human, rat, *Xenopus laevis*, and *Caenorhabditis elegans*) and plants (*Medicago sativa*) than those of the budding yeast *S. cerevisiae* [9]. These include inosin-5-monophosphate dehydrogenase, guanine nucleotide-binding protein beta subunit-like protein, and folylpolyglutamate synthase (all closer to human), ATP citrate-lyase (similar to rat), developing GTP-binding protein (similar to *X. laevis*), hypothetical protein ZK370.3 (similar to *C. elegans*), and glutamate synthase (NADH) precursor (similar to *M. sativa*). None of these proteins found by Yoshioka et al. [9] is present in our cluster I, as they have more than one protein per KOG (glutamate synthase, folylpolyglutamate synthase) or are absent (guanidine nucleotide-binding protein, ATP citrate-lyase) or are not assigned to any KOG family (inosin-5-monophosphate dehydrogenase, developing GTP-binding protein, hypothetical protein ZK370.3) in one of the 19 genomes compared.

The genome of *S. pombe* also has genes similar to human disease-related genes and the largest portion of these genes is implicated in cancer [4]. They sum a total of 23 that are involved in DNA damage and repair, checkpoint controls, and the cell cycle, which are all processes involved in maintaining genomic stability. One gene (SPBC1703.04) from these 23 that encodes for DNA mismatch repair protein-MLH1 family (KOG1979) corresponds to the HNPCC; *MLH*1 human cancer gene is present in our cluster I proteins. A second gene (SPCC533.03) that encodes for AAA+-type ATPase (KOG0735) protein is similar to Zellweger syndrome; *PEX*1, a human neurological disease gene, is also present in our cluster I. These *S. pombe* proteins are more alike to human proteins than to fungal proteins. Another gene, named dolichol phosphate mannose synthase, reported by Collussi et al. [10] in *S. pombe*, is also more similar to the human counterpart

Table 1
Number and predict function of KOGs represented by a single protein shared among 19 eukaryote genomes

| KOG Number | Predict function |
|---|---|
| KOG1131 | RNA polymerase II transcription initiation/nucleotide excision repair factor TFIIH, 5′-3′ helicase subunit RAD3 |
| KOG1967 | DNA repair/transcription protein Mms19 |
| KOG2487 | RNA polymerase II transcription initiation/nucleotide excision repair factor TFIIH, subunit TFB4 |
| KOG3471 | RNA polymerase II transcription initiation/nucleotide excision repair factor TFIIH, subunit TFB2 |
| KOG0479 | DNA replication licensing factor, MCM3 component |
| KOG0970 | DNA polymerase alpha, catalytic subunit |
| KOG1942 | DNA helicase, TBP-interacting protein |
| KOG1969 | DNA replication checkpoint protein CHL12/CTF18 |
| KOG1979 | DNA mismatch repair protein-MLH1 family |
| KOG2267 | Eukaryotic-type DNA primase, large subunit |
| KOG2299 | Ribonuclease HI |
| KOG2671 | Putative RNA methylase |
| KOG2928 | Origin recognition complex, subunit 2 |
| KOG3303 | Predicted alpha-helical protein, potentially involved in replication/repair |
| KOG2038 | CAATT-binding transcription factor/60S ribosomal subunit biogenesis protein |
| KOG1063 | RNA polymerase II elongator complex, subunit ELP2, WD repeat superfamily |
| KOG3169 | RNA polymerase II transcriptional regulation mediator |
| KOG3438 | DNA-directed RNA polymerase, subunit L |
| KOG0050 | mRNA splicing protein CDC5 (Myb superfamily) |
| KOG0213 | Splicing factor 3b, subunit 1 |
| KOG0291 | WD40-repeat-containing subunit of the 18S rRNA processing complex |
| KOG0306 | WD40-repeat-containing subunit of the 18S rRNA processing complex |
| KOG0319 | WD40-repeat-containing subunit of the 18S rRNA processing complex |
| KOG1070 | rRNA processing protein Rrp5 |
| KOG1127 | TPR repeat-containing protein |
| KOG1135 | mRNA cleavage and polyadenylation factor II complex, subunit CFT2 (CPSF subunit) |
| KOG1272 | WD40-repeat-containing subunit of the 18S rRNA processing complex |
| KOG2051 | Nonsense-mediated mRNA decay 2 protein |
| KOG2771 | Subunit of tRNA-specific adenosine-34 deaminase |
| KOG2780 | Ribosome biogenesis protein RPF1, contains IMP4 domain |
| KOG2781 | U3 small nucleolar ribonucleoprotein (snoRNP) component |
| KOG2837 | Protein containing a U1-type Zn-finger and implicated in RNA splicing or processing |
| KOG2863 | RNA lariat debranching enzyme |
| KOG3013 | Exosomal 3′-5′ exoribonuclease complex, subunit Rrp4 |
| KOG3045 | Predicted RNA methylase involved in rRNA processing |
| KOG3068 | mRNA splicing factor |
| KOG3080 | Nucleolar protein-like/EBNA1-binding protein |
| KOG3117 | Protein involved in rRNA processing |
| KOG1069 | Exosomal 3′-5′ exoribonuclease complex, subunit Rrp46 |
| KOG1416 | tRNA(1-methyladenosine) methyltransferase, subunit GCD10 |
| KOG1612 | Exosomal 3′-5′ exoribonuclease complex, subunit Rrp42 |
| KOG2523 | Predicted RNA-binding protein with PUA domain |
| KOG2554 | Pseudouridylate synthase |
| KOG2623 | Tyrosyl-tRNA synthetase |
| KOG4089 | Predicted mitochondrial ribosomal protein L23 |
| KOG4548 | Mitochondrial ribosomal protein L17 |
| KOG0363 | Chaperonin complex component, TCP-1 beta subunit (CCT2) |
| KOG0396 | Uncharacterized conserved protein |
| KOG0687 | 26S proteasome regulatory complex, subunit RPN7/PSMD6 |
| KOG0735 | AAA+-type ATPase |
| KOG0938 | Adaptor complexes medium subunit family |
| KOG1119 | Mitochondrial Fe-S cluster biosynthesis protein ISA2 (contains a HesB-like domain) |
| KOG1173 | Anaphase-promoting complex (APC), Cdc16 subunit |
| KOG1299 | Vacuolar sorting protein VPS45/Stt10 (Sec1 family) |
| KOG1349 | Gpi-anchor transamidase |
| KOG1539 | WD repeat protein |
| KOG1763 | Uncharacterized conserved protein, contains CCCH-type Zn-finger |
| KOG1835 | Uncharacterized conserved protein |
| KOG1876 | Actin-related protein Arp2/3 complex, subunit ARPC4 |
| KOG2015 | NEDD8-activating complex, catalytic component UBA3 |
| KOG2055 | WD40 repeat protein |
| KOG2165 | Anaphase-promoting complex (APC), subunit 2 |
| KOG2268 | Serine/threonine protein kinase |
| KOG2340 | Uncharacterized conserved protein |
| KOG2463 | Predicted RNA-binding protein Nob1p involved in 26S proteasome assembly |
| KOG2490 | Predicted membrane protein |
| KOG2515 | Mannosyltransferase |
| KOG2564 | Predicted acetyltransferases and hydrolases with the alpha/beta hydrolase fold |
| KOG2635 | Medium subunit of clathrin adaptor complex |
| KOG2654 | Uncharacterized conserved protein |
| KOG2707 | Predicted metalloprotease with chaperone activity (RNAse H/HSP70 fold) |
| KOG2728 | Uncharacterized conserved protein with similarity to phosphopantothenoylcysteine synthetase/decarboxylase |
| KOG2750 | Uncharacterized conserved protein similar to ATP/GTP-binding protein |
| KOG2754 | Oligosaccharyltransferase, beta subunit |
| KOG2884 | 26S proteasome regulatory complex, subunit RPN10/PSMD4 |
| KOG2927 | Membrane component of ER protein translocation complex |
| KOG2973 | Uncharacterized conserved protein |
| KOG2978 | Dolichol-phosphate mannosyltransferase |
| KOG2986 | Uncharacterized conserved protein |
| KOG3048 | Molecular chaperone Prefoldin, subunit 5 |
| KOG3059 | N-acetylglucosaminyltransferase complex, subunit PIG-C/GPI2, required for phosphatidylinositol biosynthesis |
| KOG3159 | Lipoate-protein ligase A |
| KOG3228 | Uncharacterized conserved protein |
| KOG3237 | Uncharacterized conserved protein |
| KOG3239 | Density-regulated protein related to translation initiation factor 1 (eIF-1/SUI1) |
| KOG3244 | Protein involved in ubiquinone biosynthesis |
| KOG3273 | Predicted RNA-binding protein Pno1p interacting with Nob1p and involved in 26S proteasome assembly |
| KOG3318 | Predicted membrane protein |
| KOG3758 | Uncharacterized conserved protein |
| KOG3800 | Predicted E3 ubiquitin ligase containing RING finger, subunit of transcription/repair factor TFIIH and CDK-activating kinase assembly factor |
| KOG3954 | Electron transfer flavoprotein, alpha subunit |
| KOG4524 | Uncharacterized conserved protein |

than to that of budding yeast. According to Sipiczki [11], results from Yoshioka et al. [9] and Collussi et al. [10], and our findings, many *S. pombe* proteins are more similar to their mammalian homologs than to those of *S. cerevisiae*. This most likely explains the topology of the phylogenetic tree based on concatenation of cluster I proteins.

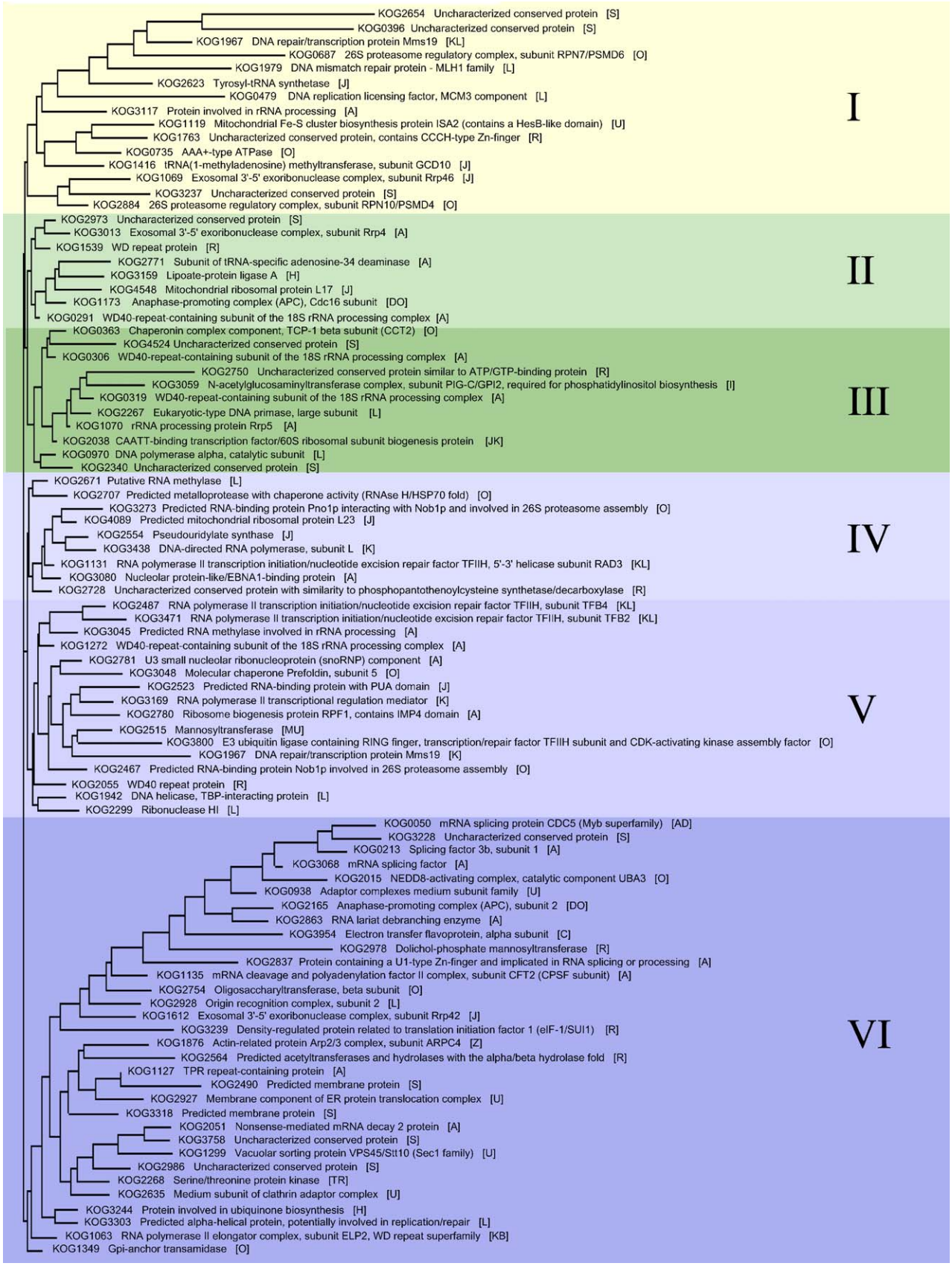


Fig. 2. KOG tree obtained by neighbor joining clustering. The proteins are grouped into six main clusters: I, II, III, IV, V, and VI. The legend of functional category of each KOG is represented in Supplementary Fig. 1.

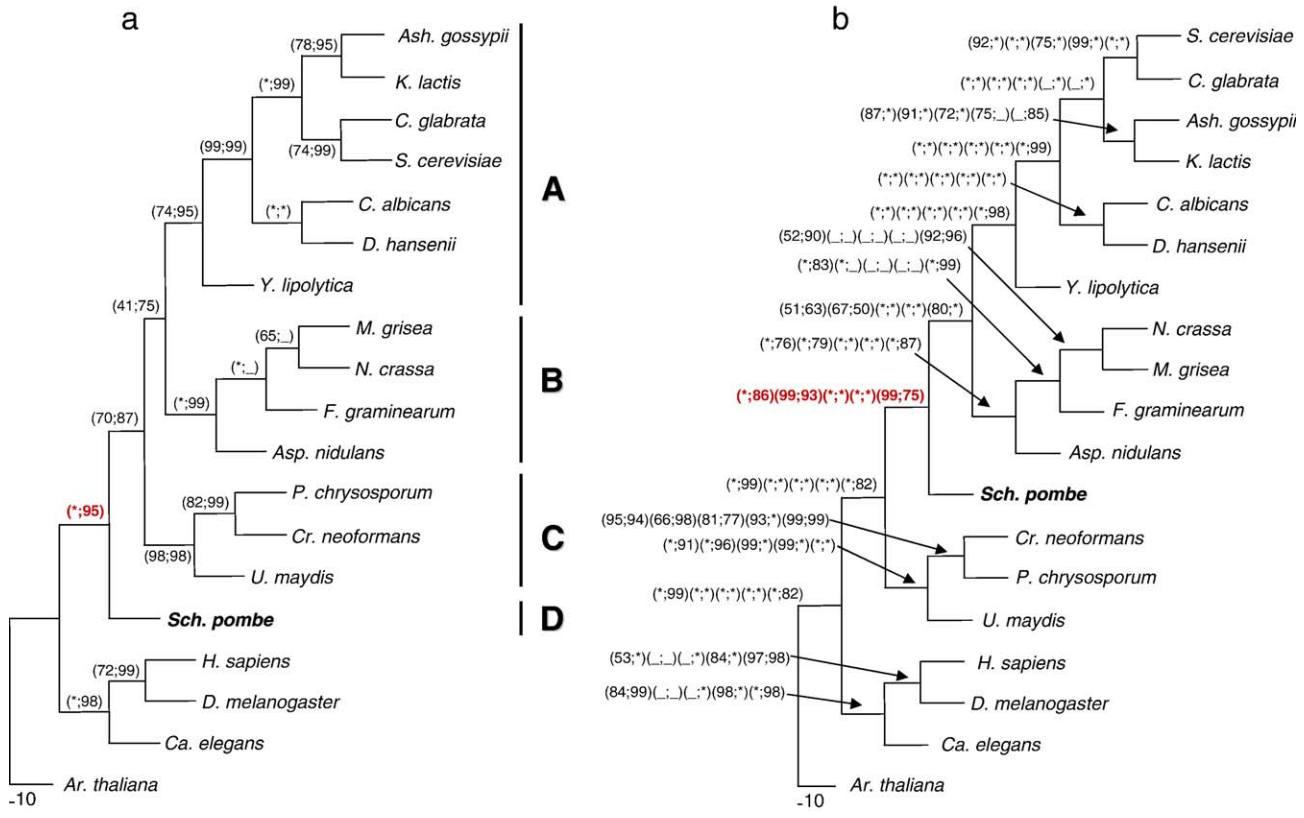


Fig. 3. Phylogenetic tree based on concatenation of proteins of (a) cluster I and (b) clusters II, III, IV, V, and VI. The numbers from right to left are branch support values (maximum parsimony bootstrap; maximum likelihood quartet puzzling support values). In (b) each pair of parentheses represents bootstrap values obtained for each cluster (II)(III)(IV)(V)(VI). A, Hemiascomycetes; B, Euascomycetes; C, Basidiomycota; D, Archiascomycetes. *Represents 100% support; _ between parentheses represents clade not present. *Ash.*, *Ashbya*; *Asp.*, *Aspergillus*; *Cr.*, *Cryptococcus*; *Sch., Schizosaccharomyces*; *Ca.*, *Caenorhabditis*; *Ar.*, *Arabdopsis*.

*Clusters II to VI proteins*

The phylogenetic analyses of the remaining protein clusters II, III, IV, V, and VI all supported hypothesis 1, in which *S. pombe* forms a lineage basal to the Euascomycetes and Hemiascomycetes (Fig. 1). This finding is in agreement with Kuramae et al. [12] after analyzing 531 common proteins among 25 eukaryote genomes. The clusters II and III supported this phylogenetic position of *S. pombe* with less than 70% bootstrap value (Fig. 3b) by both MP and QP phylogenetic methods, but the support by the clusters IV and V (both with 100% bootstrap values Fig. 3b) and cluster VI (100% MP and 80% QP, Fig. 3b) was very high. Together, these five clusters (II–VI) represent the majority (83.5%) of the commonly occurring single-protein KOGs among these 19 genomes. Among the fungi, *Neurospora crassa* (Euascomycetes) has been reported to have an rRNA gene organization similar to that of *S. pombe* [13], thus implying that *S. pombe* may be more closely related to *N. crassa* than to *S. cerevisiae* (Hemiascomycetes), at least based on rRNA data.

A second genome feature that supports hypothesis 1 is the intron size. Fungi display a wide diversity of average intron size [14]. Recent studies provided new insight into the patterns and mechanisms of intron evolution in *A. nidulans*, *F. graminearum*, *M. grisea*, and *N. crassa* genomes [15,16] and in *S. cerevisiae*, *S. pombe*, and *Cr. neoformans* [16]. The class Hemiascomycetes represented by *S. cerevisiae* [17], *C. glabrata, K. lactis*, *Y. lipolytica*, *D. hansenii* [18], *A. gossypii* [19], and *C. albicans* [20] has fewer introns (Table 2) than *S. pombe* [4], *N. crassa*, *A. nidulans*, and *Cr. neoformans* [16]. *S. pombe* has many short introns with an average length of 81 nucleotides [4]. The relatively large number of introns in *S. pombe*, *N. crassa*, *A. nidulans*, and *Cr. neoformans* provide opportunities for alternative splicing, as has been determined in *Cr. neoformans* [21]. These alternative splicings generate protein variants, which could have regulatory roles and increase the range of protein types present in the cell [22].

A third feature in the *S. pombe* genome that suggests a phylogenetic position closer to Basidiomycota and Euascomycetes than to the Hemiascomycetes (hypothesis 1) is the transposon elements. The majority of the transposons present in the genome of *S. pombe* occur clustered in single blocks on each chromosome [4] similar to their organization in *Cr. neoformans* [21] and *N. crassa* [23].

Interestingly, none of the 91 analyzed proteins support hypotheses 2 and 3 (Fig. 1). According to our analysis there are no groups of proteins in any cluster of the KOG tree that support hypothesis 3, in which the Hemiascomycetes and

Table 2
Eukaryote genome origin, source, and features used in this study

| Genome | Strain | Data source | Number of KOG | Genome size (Mb) | Number of genes | Total number of introns | GC% content |
|---|---|---|---|---|---|---|---|
| *Arabidopsis thaliana* | – | GeneBank | 3286 | 125 | 125,498 | – | – |
| *Ashbya gossypii* | ATCC 10895 | Stanford | 2592 | 9.2 | 4718 | 221 | 52 |
| *Aspergillus nidulans* | FGSC A4 | Whitehead | 2982 | 31 | 9396 | 11,520 | 50 |
| *Caenorhabditis elegans* | – | Sanger | 4235 | 97 | 19,000 | – | – |
| *Candida albicans* | SC5314 | Stanford | 2636 | 15 | 7677 | 2749 | 33.5 |
| *Candida glabrata* | CBS138 | Genolevures | 2505 | 13 | 5283 | 84 | 38.8 |
| *Cryptococcus neoformans* | JEC21 | TIGR | 2856 | 24 | 6572 | 35,025 | 48.6 |
| *Debaryomyces hansenii* | CBS767 | Genolevures | 2760 | 12–13 | 6906 | 356 | 36.3 |
| *Drosophila melanogaster* | – | GeneBank | 4352 | 137 | 14,300 | – | – |
| *Fusarium graminearum* | PH-1 (NRRL 31084) | Whitehead | 3063 | 36 | – | – | – |
| *Homo sapiens* | – | GeneBank | 4597 | 3200 | ~30,000 | – | – |
| *Kluyveromyces lactis* | CLIB210 | Genolevures | 2596 | 11.4 | 5329 | 130 | 38.7 |
| *Magnaporte grisea* | 70–15 | Whitehead | 2917 | 40 | – | – | – |
| *Neurospora crassa* | N-150 | Whitehead | 2962 | 40 | 10,082 | 17,118 | 49.9 |
| *Phanerochaete chrysosporium* | RP78 | JGI | 2945 | 29.9 | 11,777 | 30,309 | 57 |
| *Saccharomyces cerevisiae* | S288C | GeneBank | 2668 | 13 | 5807 | 301 | 38.3 |
| *Schizosaccharomyces pombe* | Urs Leupold 972 $h^-$ | Sanger | 2762 | 14 | 4940 | 4714 | 36 |
| *Ustilago maydis* | 521 | Whitehead | 2850 | 20 | – | – | 39.1 |
| *Yarrowia lipolytica* | CLIB99 | Genolevures | 2699 | 20–21 | 6703 | 740 | 49 |

Archiascomycetes (*S. pombe*) are sister groups with the Euascomycetes occurring as basal lineage. Hypothesis 3 is the result of phylogenetic analysis of mitochondrial proteins and genes [7,8] and this may explain the clustering of the yeast life style in a single branch. The mitochondrial genome of *S. pombe* was found to be more closely related to those of budding yeasts (*Kluyveromyces*, *Saccharomyces*, and *Torulopsis*) than to those of filamentous Ascomycota (*Aspergillus*, *Neurospora*, and *Podospora*) [24]. It appears quite likely that the mitochondrial genes have evolved differently from nuclear genes, therefore resulting in different phylogenetic trees. Thus, the phylogenetic position of *S. pombe* based on mitochondrial genomes, proteins, or genes may be different from that based on nuclear proteins because of different rates of organelle and genome evolution.

Here we showed that the phylogenetic position of *S. pombe* in the Tree of Life can be in two different branches, depending on the set of proteins considered. Probably, this reflects a heterogeneous background of genes, not only when one compares the nuclear proteins with the mitochondrial proteins but also among the nuclear proteins.

## Material and methods

### *Genomes and KOG assignment*

Nineteen eukaryote genomes used in this study are listed in Table 2. The group orthology framework presented in the KOG database [25] was the basis of our analyses. KOGs of *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe* were obtained from the KOG database ftp://www.ftp.ncbi.nih.gov/pub/COG/KOG/. Thirteen proteomes from *Ashbya gossypii*, *Aspergillus nidulans*, *Candida albicans*, *Candida glabrata*, *Cryptococcus neoformans*, *Debaryomyces hansenii*, *Fusarium graminearum*, *Kluyveromyces lactis*, *Magnaporthe grisea*, *Neurospora crassa*, *Phanerochaete chrysosporium*, *Ustilago maydis*, and *Yarrowia lipolytica* were assigned for orthologies using the STRING program [12,26]. Subsequently, we selected only the KOGs represented by a single protein to avoid problems in orthology assessment due to paralogy in the KOGs represented by a higher number of protein families.

### *Alignment, Gblocks, and distance matrix*

Each KOG protein shared among the 19 genomes and represented by a single protein was aligned by Clustal X [27] and the phylogeneticly informative blocks were selected by GBlocks program [28]. The distance matrix (percentage divergence) of each KOG protein alignment was determined by calculating the distances between all pairs of sequences from a multiple alignment. Then, we computed the Pearson's correlation between all 91 single-protein distance matrices. The clustering of the KOGs was calculated by using the neighbor joining method to build the KOG tree.

### *Concatenation of similar groups of KOGs and phylogenetic analysis*

The proteins comprising each single cluster represented in the KOG tree (Fig. 3) were concatenated, aligned, and analyzed by Gblocks; then a distance matrix was built. The phylogenetic analyses were carried out using maximum parsimony and quartet puzzling using maximum likelihood. MP analysis was done using PROTPARS (heuristic search with characters equally weighted) from the Phylip package [29]. Nonparametric bootstrap support for MP was calculated from 100 resampling rounds. QP trees were constructed by using the TREE-PUZZLE program [30] using the Whelan and Goldman [31] model of amino acid substitution. To root our phylogenetic trees, we have selected eight KOGs shared by eight eukaryotes, one archaea, and one bacterium genome [12]. Based on that result [12] we rooted the phylogenetic trees with *Arabidopsis thaliana.*

## Acknowledgment

This work was supported by the Renewal Fund of the Royal Netherlands Academy of Arts and Sciences (RNAAS–KNAW).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.07.001.

## References


[1] P. Lindner, *Schizosaccharomyces pombe* n. sp. neuer Gärungserreger, Wochenschr f Brauerei 10 (1893) 1298–1300.

[2] C.P. Kurztman, Discussion of teleomorphic and anamorphic ascomycetous yeasts and a key to genera, in: C.P. Kurtzman, J.W. Fell (Eds.), The Yeast A Taxonomy Study, Elsevier Science BV, 1988, pp. 111–121.

[3] H. Nishida, J. Sugiyama, Phylogenetic relationships among Taphrina, Saitoella, and other higher fungi, Mol. Biol. Evol. 10 (1993) 431–436.

[4] V. Wood, R. Gwilliam, M.A. Rajandream, et al., The genome sequence of *Schizosaccharomyces pombe*, Nature 415 (2001) 845–848 (133 co-authors).

[5] H. Prillinger, K. Lopandic, W. Schweigkofler, R. Deak, H.J. Aarts, R. Bauer, K. Sterflinger, G.F. Kraus, A. Maraz, Phylogeny and systematics of the fungi with special reference to the Ascomycota and Basidiomycota, Chem. Immunol. 81 (2002) 207–295.

[6] S. Diezmann, C.J. Cox, G. Schonian, R.J. Vilgalys, T.G. Mitchell, Phylogeny and evolution of medical species of *Candida* and related taxa: a multigenic analysis, J. Clin. Microbiol. 42 (2004) 5624–5635.

[7] P.V Pramateftaki, V.N. Kouvelis, P. Lanaridis, M.A. Typas, The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts, FEMS Yeast Res. 6 (2006) 77–90.

[8] C.E. Bullerwell, J. Leigh, L. Forget, B.F. Lang, A comparison of three fission yeast mitochondrial genomes, Nucleic Acids Res. 31 (2003) 759–768.

[9] S. Yoshioka, K. Kato, K. Nakai, H. Okayama, H. Nojima, Identification of open reading frames in *Schizosaccharomyces pombe* cDNAs, DNA Res. 4 (1997) 363–369.

[10] P.A Collussi, C.A. Taron, J.C. Mack, P. Orlean, Human and *Saccharomyces cerevisiae* dolichol phosphate mannose synthases represent two classes of enzyme, but both function in *Schizosaccharomyces pombe*, Proc. Natl. Acad. Sci. USA 94 (1997) 7873–7878.

[11] M. Sipiczki, Phylogenesis of fission yeasts. Contradictions surrounding the origin of centure old genus, Antonie van Leeuwenhoek 68 (1995) 119–149.

[12] E.E. Kuramae, V. Robert, B. Snel, M. Weiß, T. Boekhout, Phylogenomics reveal a robust fungal tree of life, FEMS Yeast Res. (2006) doi: 10.1111/j.1567-1364.2006.00119.x.

[13] S.J. Free, P.W. Rice, R.L. Metzenberg, Arrangement of the genes coding for ribosomal ribonucleic acids in *Neurospora crassa*, J. Bacteriol. 137 (1979) 1219–1226.

[14] J.E. Galagan, M.R. Henn, L.J. Ma, C.A. Cuomo, B. Birren, Genomics of the fungal kingdom: insights into eukaryotic biology, Genome Res. 15 (2005) 1620–1631.

[15] C.B. Nielsen, B. Friedman, C.B. Birren, J.E. Galagan, Patterns of intron gain and loss in fungi, PloS Biol. 2 (2004) E422.

[16] D.M. Kupfer, S.D. Drabenstot, K.L. Buchanan, H. Lai, H. Zhu, D.W. Dyer, B.A. Roe, J.W. Murphy, Introns and splicing elements of five diverse fungi, Eukaryot. Cell 3 (2004) 1088–1100.

[17] A.B. Goffeau, G. Barrell, H. Bussey, et al., Life with 6000 genes, Science 274 (1996) 546–567 (16 co-authors).

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

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

[20] T. Jones, N.A. Federspiel, H. Chibana, et al., The diploid genome sequence of *Candida albicans*, Proc. Natl. Acad. Sci. USA 101 (2004) 7329–7334 (12 co-authors).

[21] B.J. Loftus, E. Fung, P. Roncaglia, et al., The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*, Science 307 (2005) 1321–1324 (54 co-authors).

[22] B.R. Graveley, Alternative splicing: increasing diversity in the proteomic world, Trends Genet. 17 (2001) 100–107.

[23] J.E. Galagan, S.E. Calvo, K.A. Borkovich, et al., The genome sequence of the filamentous fungus *Neurospora crassa*, Nature 422 (2003) 859–868 (77 co-authors).

[24] D. Sankoff, G. Leduc, N. Antoine, B. Paguin, B.F. Lang, R. Cedergren, Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome, Proc. Natl. Acad. Sci. USA 89 (1992) 6575–6579.

[25] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, et al., The COG database: an updated version includes eukaryotes, BMC Bioinformatics 4 (2003) 41 (17 co-authors).

[26] B. Snel, G. Lehmann, P. Bork, M.A. Huynen, STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene, Nucleic Acids Res. 28 (2000) 3442–3444.

[27] J.D Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.

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

[29] J. Felsenstein, Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods, Methods Enzymol. 266 (1996) 418–427.

[30] H.A. Schmidt, K. Strimmer, M. Vingron, A. von Haeseler, TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing, Bioinformatics 18 (2002) 502–504.

[31] S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach, Mol. Biol. Evol. 18 (2001) 691–699.