Phylogeny, sequence conservation, and functional complementation of the SBDS protein family

G.R.B. Boocock a,b , M.R. Marita a, J.M. Rommens a,b,*

a Program in Genetics and Genomic Biology, The Hospital for Sick Children, 101 College Street, East Tower, Toronto, Canada ON M5G 1L7
b Department of Molecular and Medical Genetics, University of Toronto, 1 King’s College Circle, Toronto, Canada ON M5S 1A8

Received 26 September 2005; accepted 31 January 2006
Available online 10 March 2006

Abstract

The Shwachman–Bodian–Diamond syndrome (SBDS) protein family occurs widely in nature, although its function has not been determined. Comprehensive database searches revealed SBDS homologues from 159 species, including examples from all sequenced archaeal and eukaryotic genomes and all eukaryotic kingdoms. Sequence alignment with ClustalX and MUSCLE algorithms led to the identification of conserved residues that occurred predominantly in the amino-terminal FYSH domain where they appeared to contribute to protein folding or stability. Only SBDS residue Gly91 was invariant in all species. Four distantly related protists were found to have two divergent SBDS genes in their genomes. In each case, phylogenetic analyses and the identification of shared sequence features suggested that one gene was derived from lateral gene transfer. We also identified a shared C-terminal zinc finger domain fusion in flowering plants and chromalveolates that may shed light on the function of the protein family and the evolutionary histories of these kingdoms. To assess the extent of SBDS functional conservation, we carried out complementation studies of SBDS homologues and interspecies chimeras in Saccharomyces cerevisiae. We determined that the FYSH domain was widely interchangeable among eukaryotes, while domain 2 imparted species specificity to protein function. Domain 3 was largely dispensable for function in our yeast complementation assay. Overall, the phylogeny of SBDS was shared with a group of proteins that were markedly enriched for RNA metabolism and/or ribosome-associated functions. These findings link Shwachman–Diamond syndrome to other bone marrow failure syndromes with defects in nucleolus-associated processes, including Diamond–Blackfan anemia, cartilage–hair hypoplasia, and dyskeratosis congenita.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Shwachman–Diamond; SBDS; Phylogeny; Domain fusion; Lateral transfer; RNA metabolism; Ribosome; Complementation

The Shwachman–Diamond syndrome protein, SBDS, is a member of a highly conserved protein family that lacks both detectable signal peptides and primary sequence homology to any previously characterized functional domain. It is broadly expressed and occurs widely in nature, with homologues in Archaea and eukaryotes [1]. In humans, SBDS deficiency leads to Shwachman–Diamond syndrome (SDS; OMIM 260400). The disease has variable presentation that may include failure to thrive, frequent infections, anemia, and short stature [2]. Patients experience maldigestion due to low levels of pancreatic enzymes resulting from a depleted acinar cell population. Neutropenia is the most common hematological finding, although deficiencies in other myeloid lineages are common. Bone marrow is typically hypoplastic and patients are at increased risk of leukemic transformation. Short stature occurs independent of nutritional status, and 50% of patients are below the third percentile for height.

To date, no patients homozygous for a very common early truncation mutation (183TA → CT) have been identified and the Saccharomyces cerevisiae and mouse homologues are essential genes ([3] and unpublished data). Together with its broad species distribution, these observations indicate that the SBDS protein family contributes to a fundamental and
conserved cellular process. This is reflected in the genomic context of archaeal SBDS homologues, which tend to cluster in a conserved "superoperon" that includes genes involved in transcription, RNA processing, translation, protein folding, and protein degradation [4]. Statistical clustering of yeast transcriptional profiles has also demonstrated that the SBDS homologue is coregulated with a functionally related group of RNA processing and ribosome-associated genes [5]. The essential nature of the SBDS gene indicates that some SDS-associated mutations must be hypomorphic. However, as the precise molecular functions of SBDS are unknown, it is not immediately obvious how suboptimal performance of a highly conserved process leads to the tissue-specific manifestations of SDS, including the serious consequences of bone marrow failure and leukemic transformation.

The wide occurrence of SBDS homologues presents opportunities to glean insights into the molecular basis of disease. For example, the structure of the *Archaeoglobus fulgidus* homologue (AF0491) was recently determined by X-ray crystallography [6,7]. It indicated a three-domain arrangement, with a novel amino-terminal FSH (fungal, YHR087wp, Shwachman) domain, a central domain corresponding to a three-helix bundle, and C-terminal domain with a topology reminiscent of a ferredoxin fold. While informed by structural comparisons, the interpretation of anticipated functional data from model organisms would be aided by a comprehensive survey of the diversity that exists within this protein family. There is also a need to determine if the sequence homology observed between family members translates into functional conservation. Simultaneously, the study of a highly conserved protein family is likely to yield phylogenetic data with implications for evolutionary models.

In the present study, we sought to identify the complete set of SBDS homologues present in current sequence databases, assess their phylogeny, and catalogue their features. We also assessed the ability of a subset of homologues to function in a yeast complementation assay. This study was extended to include a more detailed examination of interspecies chimeras to assess the functional conservation of each structural domain. We present evidence that acquired domain fusions are present in some lineages, while lateral gene transfer (LGT) may have influenced the evolution of this gene family in others. Overall, SBDS phylogeny was consistent with a role in RNA metabolism and/or ribosome biogenesis or function.

**Results and discussion**

**Identification of SBDS homologues**

SBDS homologues were identified through the use of the basic local alignment search tool (BLAST) [8]. Overall, the first two of three SBDS domains were highly conserved, while sequences corresponding to the C-terminal domain were often below default BLAST detection limits. Reinteractive BLAST analysis with a position-specific score matrix did not yield additional sequences with significant homology [9]. Full-length and partial homologues were detected in 159 species, including representatives of all completely sequenced archaeal and eukaryotic genomes (summarized in Table 1 and listed with GenBank accession numbers in Table 1S). The 23 archaeal homologues had a species distribution that reflected the availability of sequenced genomes and included 18 Euryarchaeota, 4 Crenarchaeota, and the single Nanoarchaeota *Nanoarchaeum equitans*. Although 228 bacterial genomes have been sequenced [10], no bacterial homologues were detected.

A current view of eukaryotic phylogeny suggests six kingdoms [11]. The Opisthokonta are unified by molecular phylogenies and a single posterior flagellum on cells or gametes. Amoebozoa include most organisms that feed and move using finger-like pseudopodia [12]. Plantae includes the three major eukaryotic lineages with *Cyanobacterium*-derived primary chloroplasts: Glucophytes, Rhodophytes (red algae), and green plants, including Chlorophytes (green algae) and Embryophytes (land plants) [13]. Chromalveolata is a grouping of photosynthetic and nonphotosynthetic organisms including phytoplanktons and human pathogens such as malaria parasites. Some evolutionary models suggest that Chromalveolates are descended from a progenitor that acquired secondary plastids through endosymbiosis of a primary Rhodophyte [13,14]. Kingdom Excavata unites diverse unicellular protist groups, including human pathogens such as *Giardia, Leishmania, Trichomonas*, and *Trypanosoma* species [15]. Finally, Rhizaria is a grouping of diverse amoeboid protists that is largely dependent upon rRNA phylogenies [16].

The 136 eukaryotic SBDS homologues included representatives of all proposed eukaryotic kingdoms but were greatly

---

**Table 1**

<table>
<thead>
<tr>
<th>Superkingdom</th>
<th>Kingdom/ phylum</th>
<th>Subgroup</th>
<th>No. species</th>
<th>No. complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukaryota</td>
<td>Opisthokonta</td>
<td>Metazoans</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fungi</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Microsporidia</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Amoebozoa</td>
<td>Entamoeba</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mycetozoa</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pelobions</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Plantae</td>
<td>Embryophytes</td>
<td>38</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorophytes</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhodophytes</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glaucophytes</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Chromalveolata</td>
<td>Alveolates</td>
<td>10</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cryptophytes</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haptophytes</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stramenopiles</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Excavata</td>
<td>Euglenozoa</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jakobids</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malawimonas</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trimastix</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parabasalis</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diplomonads</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cercozoa</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Archaea</td>
<td>Euryarchaeota</td>
<td>18</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crenarchaeota</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nanoarchaeota</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>159</td>
<td>91</td>
</tr>
</tbody>
</table>
enriched for Opisthokont and Plantae species (their distribution is summarized in Table 1). Reflective of the current scope of sequencing projects, more than half of the data set comprised sequences from metazoaos \( (n = 43) \), flowering plants \( (n = 38) \), and fungi \( (n = 18) \). Four eukaryotes (Giardia lamblia, Trichomonas vaginalis, Entamoeba histolytica, and Cyanidioschyzon merolae) had two SBDS genes (discussed below).

**Sequence conservation across the SBDS protein family**

The SBDS gene encodes a polypeptide of 250 residues with a predicted molecular mass of 28.8 kDa. To assess the extent of sequence conservation across species, 87 full-length SBDS homologues were aligned using ClustalX and MUSCLE [17,18] (Figs. 1S and 2S). We excluded the four species with two SBDS genes from the initial analysis, presuming that the presence of two intact genes in a genome may have altered selective pressures. Full-length eukaryotic homologues ranged from 223 (Encephalitozoon cuniculi) to 638 residues (Plasmodium falciparum). Homologues longer than 343 residues had large C-terminal extensions beyond the region of domain 3 alignment (discussed below). Archaeal homologues ranged from 226 (N. equitans) to 242 (Haloarcula marismortui) residues and were shorter than their eukaryotic counterparts due to abbreviated N-terminal regions. Compared to the aligned eukaryote with the shortest N-terminus (E. cuniculi), Archaea had 4 (Methanopyrus kandleri) to 14 (N. equitans) fewer N-terminal residues.

We compiled a list of amino acid positions that were invariant or had only conservative substitutions in alignments generated by both algorithms (Table 2). Sequences corresponding to the amino-terminal FYSH domain were the most highly conserved, followed by domain 2 and domain 3, respectively. Of 19 residues conserved in eukaryotes and Archaea, 12 were in the amino-terminal FYSH domain, 4 were in the second domain, 1 was in the C-terminal domain, and 2 were situated between structural domains. Only 4 residues (corresponding to Glu28, Phe75, Gly91, and Pro214 of SBDS) were invariant. Stronger conservation was observed within eukaryotes, in which 35 residues were conserved: 25 in domain 1, 6 in domain 2, 1 in domain 3, and 3 in interdomain regions. This initial analysis was limited by the current sampling of species diversity. Fourteen partial sequences were available for protist groups that were otherwise poorly represented in the alignment. These sequences were added to the alignment and assessed with respect to conserved positions (sequences marked by asterisks in Figs. 1S and 2S). They included some of the most divergent eukaryotic homologues; only 12 of the 35 residues conserved in other eukaryotes were conserved in these additional species (divergences are italicized “Exceptions” in Table 2). Notably, *Cyanophora paradoxa* had a threonine at the otherwise-invariant Glu28 position, though we could not formally exclude the possibility of a sequencing error.

The 20 positions conserved in full-length eukaryotic and archaelal sequences were mapped onto the *A. fulgidus* crystal structure [6,7]. We found that they generally did not possess solvent-exposed side chains, suggesting that they are more likely to contribute to protein folding and stability than to conserved molecular interactions. For example, the near-invariant SBDS residue Glu28 corresponds to Glu20 in the *A. fulgidus* homologue (AF0491) and is situated in beta strand 2 \((\beta2)\) of the FYSH domain. The R group of Glu20 is hydrogen-bonded to Arg11 of AF0491 (situated in \(\beta1\)), while its backbone carbonyl and amide groups are hydrogen-bonded to Phe49 (situated in \(\beta4\)). As such, this residue is predicted to stabilize interactions of \(\beta2\) with \(\beta1\) and \(\beta4\). The interacting residues are also conserved (SBDS residues Arg19 and Phe57, Table 2), and we have noted that SBDS residue Arg19 is the site of a disease-associated missense mutation, R19Q (unpublished data). The invariant residue Gly91 corresponds to AF0491 Gly83 and is situated beyond alpha helix 4 (\(\alpha4\)) toward the C-terminus of the FYSH domain. \(\Phi\) and \(\Psi\) angles for this residue are \(-101^\circ\) and \(-147^\circ\), respectively. These values fall within the favored range for glycine, but would be disfavored for all other residues [19]. We conclude that this position is invariant due to structural constraints. Overall conserved features of the SBDS protein family appeared to contribute to protein folding and lacked characteristics of canonical motifs associated with enzyme active sites.

**Four species have two SBDS genes**

Four protist species (G. lamblia, T. vaginalis, *Ent. histolytica*, and *Cyan. merolae*) each had two detected SBDS homologues. These proteins were divergent at several positions that were conserved in other eukaryotes (boldface “Exceptions” in Table 2). For example, both *G. lamblia* proteins were divergent at positions 19, 94, and 155. Both *T. vaginalis* homologues were divergent at positions 26, 32, 75, and 94, and both *Ent. histolytica* proteins were divergent at position 95. Specific divergences differed between proteins from the same species. In all four species, BLAST analysis indicated that one protein was more similar to eukaryotic SBDS homologues (primary homologues), while the other was highly divergent (secondary homologues). For example, primary homologues shared 37–41% amino acid sequence identity with SBDS, while secondary homologues were less than 25% identical to SBDS (Fig. 1). By comparison, homologue pairs from the same species were only 23–31% identical across domains 1 and 2 and had no obvious sequence homology across their third domains.

The four species with two SBDS genes are not closely related and represent three eukaryote kingdoms. The vast majority of eukaryotes, including species more closely related to these four, possess only one homologue. Ancient paralogy is therefore an improbable ancestral state. The second gene is more likely to be a derived trait that has arisen either by recent, independent paralogous gene duplications (PGD) or LGT events. A PGD model of evolution would be supported if primary and secondary homologues from each species were similar to one another. However, as discussed above,
primary homologues were more similar to SBDS than to their respective secondary homologues. Furthermore, secondary homologues shared abbreviated amino-termini (reminiscent of archaeal homologues) and possessed similar extreme C-termini enriched for lysine and glycine residues and including a common K-G-X-K-K/G motif of unknown
significance (Fig. 1). These features were not observed in the corresponding primary homologues. Secondary homologues also appeared to be more similar to one another at several positions across the sequence alignment (Fig. 1). These observations are not easily reconciled with independent PGD events.

Fig. 1. Sequence alignment of homologues from species with two genes. Amino acid positions are numbered with respect to SBDS (HSA). ClustalX alignment of homologues from species with two SBDS genes suggests that secondary homologues (CME2, EHI2, GLA2, and TVA2) are more closely related to one another than to primary homologues (CME1, EHI1, GLA1, and TVA1). Primary homologues are more similar to one another at seven positions across the alignment (boxed). Secondary homologues were more similar to one another at their extreme amino-termini, nine positions throughout the alignment, and their extreme C-terminal regions (shaded). The lysine/glycine-rich region at the C-terminus of secondary homologues appears in boldface, and the shared K-G-X-K-K/G motif is in a black box. Vertical lines denote nonconserved insertions (present in fewer than four of the nine sequences) that were removed from the alignment.

HSA, Homo sapiens; CME, Cyanidioschyzon merolae; EHI, Entamoeba histolytica; GLA, Giardia lamblia; TVA, Trichomonas vaginalis; 1, primary homologue; 2, secondary homologue.
Phylogenetic analyses support a LGT origin of secondary homologues

To ascertain whether the shared features of secondary homologues and their divergences from other eukaryotes were significant, phylogenetic analyses were performed with maximum likelihood (ML), maximum parsimony (MP), and distance matrix analysis (DM) methods [20,21]. The sequences of 95 full-length SBDS homologues were aligned using ClustalX and MUSCLE (subsets without asterisks in Figs. 1S and 2S) and the refined maximum aligned regions (corresponding to SBDS residues 18 to 237 in both cases) were used in subsequent analyses. Results indicated that secondary homologues constituted an outgroup with respect to other eukaryotic and archaeal homologues, while the 4 primary homologues were grouped among other eukaryotes (Figs. 2a and 2b).

To assess the impact of limited species sampling on these analyses, alignments that included partial sequences from several divergent and underrepresented protist species were generated (Figs. 1S and 2S, excluding regions in boldface). The maximum aligned region for both MUSCLE and ClustalX alignments corresponded to SBDS residues 18–179 and excluded domain 3. Phylogenetic analyses again separated secondary homologues from the vast majority of eukaryotic and archaeal homologues (Figs. 2c and 2d). However, examples from six unrelated protist species were also included in the outgroup, including Bigelowiella natans, Capsaspora owczarzaki, Cyanophora paradoxa, Mastigamoeba balamuthi, Malawimonas jakobiiformis, and Reclinomonas americana. Co-clustering of species from five eukaryotic kingdoms argues that we cannot infer a close relationship between secondary homologues and that long-branch attraction may explain their positions in these trees. Secondary genes could be recently derived and rapidly evolving paralogues, but this hypothesis fails to account for the similarities between these proteins (Fig. 1). It is also unlikely that congergent evolution could have generated sequence similarities that are distributed across the proteins. Collectively, these results suggest that secondary homologues may be derived from LGT and may be xenologues that have been dispersed from related source species among four unrelated protists. Broader species sampling may elucidate the origins of the xenologues and define the role that LGT has played in evolution of the SBDS gene family. It would be interesting to determine if the xenologues encode novel SBDS functions and what selective advantage, if any, they impart to their host genomes.

Fig. 2. Phylogenetic analyses of secondary homologues. Condensed radial trees are shown with the results of phylogenetic analyses of (a, c) MUSCLE and (b, d) ClustalX alignments using distance matrix analysis (DM), maximum parsimony (MP), and maximum likelihood (ML) methods. Bootstrap values for 100 replicates are shown for DM/MP/ML algorithms. (a) Analyses of the maximum MUSCLE-aligned region (SBDS residues 18 to 237) of 95 full-length SBDS homologues (nonasterisked subset of sequences in Fig. 2S) indicate that secondary homologues (2°) are an outgroup to other eukaryotic and archaeal examples. The node was supported by high bootstrap values in all analyses. Primary homologues (1°) were grouped among other eukaryotes. (b) Analyses of a ClustalX-generated alignment of full-length sequences (nonasterisked subset of Fig. 2S) yielded similar results, with high bootstrap values for DM and ML methods. (c) Partial protist sequences (asterisked sequences, Fig. 2S) were included in the MUSCLE alignment and the maximum-aligned region (SBDS residues 18–179) was used for phylogenetic analyses. These yielded a radial tree in which secondary homologues were grouped with six diverse protist species (Bigelowiella natans, Capsaspora owczarzaki, Cyanophora paradoxa, Mastigamoeba balamuthi, Malawimonas jakobiiformis, and Reclinomonas americana) with strong bootstrap support in MP- and ML-generated trees. (d) Analyses using the maximum ClustalX-aligned region of full-length and partial protist sequences (SBDS residues 18–179) produced similar results, although bootstrap values for these nodes were low and may reflect reduced alignment quality. Overall, phylogenetic analyses indicated that secondary SBDS homologues are highly divergent, but fall within the scope of total eukaryotic diversity.
Complementation testing indicates that domain 3 is dispensable

Although the SBDS protein shows strong sequence conservation across species, it is not known if this translates into functional conservation among family members or their component domains. To address this issue, we exploited the essential nature of the *S. cerevisiae* homologue (YLR022c) to develop a complementation assay that was used to assess the function of a range of protein truncations, homologues, chimeras, and disease-associated mutations. Briefly, we generated a haploid yeast strain with a genomic deletion of YLR022c and carrying a single-copy, *Ura3*-bearing plasmid expressing YLR022cp with a carboxyl-terminal FLAG epitope under the control of a *Gal1* promoter (pRS416 [*Gal1*∷YLR022CFLAG]) [22]. Genes to be tested for complementation were cloned into a second plasmid with a *Leu2* marker (pRS415). All test genes were cloned with a C-terminal hemagglutinin (HA) epitope and were expressed with a regulatable *Met25* promoter [22]. Complementation testing was accomplished by assessing the ability of the second expression vector to sustain growth following active selection against the *Ura3*-bearing plasmid (and wild-type YLR022c) on medium containing 5-fluoroorotic acid. This plasmid shuffle was followed by a quantitative assessment of growth on medium promoting high (permissive condition) or low (restrictive condition) expression from the *Met25* promoter. This promoter yielded 5- (restrictive condition) to 20-fold (permissive condition) higher expression of YLR022cp compared to the endogenous promoter and overexpression of wild-type protein did not appear to impact growth rate (data not shown). Our complementation assay with high expression should increase sensitivity to residual protein activity. This was confirmed with disease-relevant mutations.
I87S and K118N. These were efficiently expressed and fully complemented under both restrictive and permissive conditions (Figs. 3a, Table 3). These mutations were previously reported to confer a slow growth phenotype when expressed with the endogenous promoter [6]. We found that the L71P mutant was expressed and conferred a subtle growth defect under the permissive condition and a markedly reduced growth rate under the restrictive condition and represents a conditional allele of YLR022cp (Figs. 3a and 3d and Table 3). L71P was previously reported to be a noncomplementing mutation when expressed from the endogenous promoter [6].

Domain 3 was dispensable for complementation; YLR022cpΔ174-250 complemented fully under both permissive and restrictive culturing conditions (Figs. 3c and 3d and Table 3). Shammas et al. [6] found that this deletion supported growth, but led to a quantitative growth defect when expressed with the endogenous promoter. We demonstrated that the FYSH domain was essential. YLR022cpΔ1-92 (Fig. 3c) was efficiently expressed (data not shown) but incapable of complementation (Table 3). Cumulatively, these results indicate the present assay is suitable for functional assessment of proteins from other species that may be susceptible to low expression, inefficient protein folding, or reduced function in yeast.

**SBDS-family proteins function in a species-specific manner**

To investigate the extent of SBDS functional conservation, 10 genes from nine diverse model organisms were tested for complementation in *S. cerevisiae*. This species included a fungus (*Schizosaccharomyces pombe*), three metazoa (*Homo sapiens, Drosophila melanogaster, and Caenorhabditis elegans*), a flowering plant (*Arabidopsis thaliana*), two Excavates (a Trypanosome, *Leishmania major*, and a Diplomonad, *G. lambia*), and two Archaea (*Giardia lamblia* and *Sulfolobus solfataricus*). The chimera was expressed (bottom, Fig. 3c) but led to a slow growth phenotype, indicating that it was low functioning with a rederived strain with a sequence-verified plasmid. The strain maintained a wild-type growth rate (Fig. 3d and Table 3) even in restrictive conditions, the *S. pombe* homologue led to a slow growth phenotype, indicating that it was low functioning in *S. cerevisiae* (Fig. 3d). Collectively, these results indicate that the SBDS protein family functions in a species-restricted manner.

**Domain 2 imparts species-specificity to SBDS function**

Species-restricted function may be a general feature of SBDS, or it may be determined by specific structural elements. To discriminate between these possibilities, human–yeast interspecies chimeras were cloned and tested for complementation (Fig. 3b). We determined that interchanging the yeast FYSH domain with cognate human sequences (HSD1C in Fig. 3b) had no detectable effect on protein function. The strain maintained a wild-type growth rate (Fig. 3d and Table 3) even though steady-state expression of the chimeric protein could not be detected (bottom, Fig. 3c). Identical results were obtained with a rederived strain with a sequence-verified plasmid. The yeast C-terminal domain could also be replaced by a homologous human sequence (HSD3C in Fig. 3b and Table 3). The chimera was expressed (bottom, Fig. 3c) but led to a significant growth defect under the restrictive culturing condition compared to the domain 3 Δ174-250 truncation (Fig. 3d and Table 3). This may reflect the difficulty encountered

### Table 3

<table>
<thead>
<tr>
<th>Gene or chimera</th>
<th>Expression</th>
<th>Comp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Homologues</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>YLR022cp</td>
<td>+++</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>SBDS</td>
<td>++</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>SPO</td>
<td>+++</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>CEL</td>
<td>++</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>DME</td>
<td>++</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>ATH</td>
<td>++</td>
</tr>
<tr>
<td>Giardia lamblia primary</td>
<td>GLA1</td>
<td>–</td>
</tr>
<tr>
<td>G. lamblia secondary A</td>
<td>GLA2a</td>
<td>++</td>
</tr>
<tr>
<td>G. lamblia secondary B</td>
<td>GLA2b</td>
<td>++</td>
</tr>
<tr>
<td>Sulfolobus solfataricus</td>
<td>SSO</td>
<td>–</td>
</tr>
<tr>
<td>Halobacterium salinarum</td>
<td>HNR</td>
<td>–</td>
</tr>
</tbody>
</table>

| YLR022cp mutants and truncations | | |
| L71P | ++ | + |
| I87S | +++ | +++ |
| K118N | +++ | +++ |
| YLR022c Δ1-92 | +++ | – |
| YLR022c Δ174-250 | ++ | +++ |

| Chimeras with YLR022cp | | |
| H. sapiens | HSD1C | – | +++ |
| H. sapiens | HSCb | ++ | – |
| H. sapiens | HSCc | ++ | – |
| H. sapiens | HSD3C | +++ | + |
| H. sapiens | HSCd | ++ | – |
| H. sapiens | HSc | ++ | – |
| H. sapiens | HSD2C | + | – |
| C. elegans | CELD1C | ++ | + |
| D. melanogaster | DMD1C | +++ | +++ |
| A. thaliana | ATHD1C | +++ | + |
| A. thaliana | ATHD3C | ++ | + |
| Leishmania major | LMAD1C | +++ | ++ |
| G. lamblia primary | GLA1D1C | +++ | – |
| G. lamblia secondary A | GLA2aD1C | ++ | – |
| G. lamblia secondary B | GLA2bD1C | ++ | – |
| S. solfataricus | SSO2D1C | ++ | – |
| H. salinarum | HNRD1C | ++ | – |

---

* Expression: ++++, normal expression; ++, reduced expression; +, very low expression; –, expression not detected.

* Complementation: ++++, full complementation; ++, growth defect under the restrictive condition; +, growth defect under both permissive and restrictive conditions; –, no complementation (no growth).

* D1C, domain 1 chimera; D2C, domain 2 chimera; D3C, domain 3 chimera. Refer to Fig. 3c for schematics of *H. sapiens* chimeras.
in defining a homologous boundary between domains two and three due to low human–yeast sequence similarity in this region. Other chimeras with humanized or partially humanized domain 2 sequences (HSCb, HSCc, HSCd, HSCe, and HSD2C in Fig. 3c) were expressed (bottom panel, Fig. 3c) but incapable of complementation (Table 3). Although human SBDS cannot functionally substitute for its S. cerevisiae counterpart, we have shown that domain 3 is dispensable and that yeast domain 1 may be replaced with a homologous human sequence. We conclude that domain 2 confers species specificity to YLR022cp function.

The function of the FYSH domain is widely conserved among eukaryotes

Complementation of the human FYSH domain in yeast led us to investigate the extent of functional conservation of this domain across a broader range of species. DNA sequences encoding the FYSH domains from the nine other cloned homologues were fused to YLR022cp domains 2 and 3. In contrast to the full-length homologues, steady-state expression of all such domain 1 chimeras could be detected (chimeras labeled with D1C in Fig. 3b and Table 3). Chimeras with the FYSH domain from C. elegans, D. melanogaster, A. thaliana, and L. major were functional in yeast (Table 3). Quantitative assessment of growth rates indicated that chimeras with Opisthokont-derived FYSH domains (i.e., H. sapiens and D. melanogaster) complemented more effectively than FYSH domains from other kingdoms (Fig. 3d). An exception was the C. elegans chimera (CELD1C), which complemented weakly (Table 3 and Fig. 3d). It is interesting to note that this homologue is one of four eukaryotes—and the only Unikont—that are divergent at conserved position Ala66 (Table 2). Although the L. major chimera was functional in yeast, neither the primary or the secondary G. lamblia homologue was capable of complementation. However, it is possible that the presence of two genes in the G. lamblia genome has altered the evolutionary constraints on both homologues, leading to sequence divergences that are not otherwise tolerated (Table 2 and discussed above). Positive complementation of Plantae- and Excavate-derived chimeras suggests that functional aspects of the FYSH domain are conserved across large evolutionary distances.

Plants and chromalveolates share a C-terminal extension with a C2H2 zinc finger

SBDS homologues from several species had extended C-terminal regions. For example, the Dicyostelium discoideum homologue has a C-terminal extension of about 22 residues including a low-complexity, serine/threonine-rich segment of 13 residues. Trypanosomatids, including two Leishmania and three Trypanosoma species, have C-terminal extensions of 199 (Trypanosoma vivax) to 269 residues (Leishmania species) that are rich in glycine and charged amino acids. A C-terminal extension with a putative C2H2 (Cys-Cys-His-His motif) zinc finger domain was previously identified in the SBDS homologues of flowering plants (Magnoliophyta) [1]. C2H2 zinc fingers occur in a broad range of proteins in which they mediate interactions with DNA and/or RNA [23,24]. In the present study, this feature was detected at the C-terminus of 17 flowering plant homologues (including all nine complete sequences) and 11 Chromalveolate homologues (including the seven with complete sequences), including 9 Alveolates (7 Apicomplexans, 1 Ciliate, and 1 Dinoflagellate) and 2 Stramenopiles. In Magnoliophyta, the first canonical cysteine of the zinc finger occurred 44 to 57 residues beyond the region of detectable human–Arabidopsis homology. In Chromalveolates, the intervening segment was more variable and ranged from 21 residues (Aphidnium carterae) to 71 residues (Toxoplasma gondii).

We could not unequivocally determine the status of a Glaucophyte (Cyano. paradoxa) or two Chlorophytes (Prototheca wickerhamii and Chlamydomonas reinhardtii) due to incomplete genomic and/or cDNA sequences. A third Chlorophyte, Ostreococcus tauri, had a C-terminal extension that lacked the canonical features of a C2H2 zinc finger domain. However, this should be interpreted cautiously because the gene was found in draft genomic sequence and has no EST support at present. As discussed previously, the only sequenced Rhodophyte genome (Cyanidi. merolae) has two SBDS homologues. Both lacked the C-terminal fusion, but this species may not be representative of other Rhodophytes because of its two SBDS homologues. We therefore sequenced cDNA clones corresponding to the homologue from the Rhodophyte, Porphyra yezoensis. This species also lacked the zinc finger fusions, suggesting that it does not occur in the Rhodophyte lineage.

Derived gene fusions have been used to infer ancestral relationships between kingdoms and suggest a root for the eukaryote tree [25,26]. The occurrence of a domain fusion in SBDS homologues from at least two major Chromalveolate lineages is supportive of the monophyly of this kingdom. Its co-occurrence in both Magnoliophyta and Chromalveolata is intriguing from an evolutionary perspective. It may reflect an ancestral trait signifying a monophyletic origin of these two kingdoms. Alternatively, LGT could also explain the existence of a shared trait in Plantae and Chromalveolates. Evidence suggests that Chromalveolate progenitors acquired plastids at least once in their common evolutionary history through secondary endosymbiosis of a primary Plantae-lineage species [14]. Replacement of a host (Chromalveolate) gene by a symbiont (Plantae) gene could explain the presence of the zinc finger domain fusion in these two kingdoms. The noted absence of the zinc finger domain in two modern Rhodophytes necessitates that the fusion was lost in this lineage in either of these two hypotheses. A third possible explanation of the phylogeny of the zinc finger fusion that does not require its loss in Rhodophytes centers on an expanded definition of Plantae [27,28]. The revised clade includes members of the Plantae and Chromalveolata kingdoms, with Rhodophytes in a basal position.

At present, the shared zinc finger domain fusion remains an intriguing trait, which will undoubtedly contribute to our understanding of the evolutionary history of these major eukaryote lineages. While the limited resolution of SBDS
phylogenetic trees currently obscures its origin, interpretation
would be aided by availability of full-length sequences for other
Rhodophyte, Chlorophyte, and Glaucophyte species.

The C-terminal zinc finger fusion may be consistent with core
SBDS function

Compared to the highly variable region that links it to the
third domain, the zinc finger is itself conserved across species.
A sequence alignment of representative zinc finger domains
from flowering plants and Chromalveolates indicated that a
canonical C2H2 motif was present in all sequences (C-X2-C-
X11-13-H-X2-H). Other residues outside the C2H2 motif were
also conserved and a more specific consensus could be
derived (C-X2-C-X5-10-F/Y-K-R-X2-N-X10-I/L/V; Fig. 4). We sought to identify other proteins with
similar zinc finger domains, anticipating that their ascribed
functions may shed light on the role of the fused zinc finger
domain in plants and Chromalveolates. BLAST analysis of
seven representative zinc finger domains against human
sequences identified two proteins with significant homology:
zinc finger-like protein 9 (ZPR9) and FLJ10415 protein
(GenBank Accession Nos. AAL02121 and AAH08948).

However, C2H2 zinc fingers are extremely abundant in the
human genome and the significance of these results was
initially unclear [29]. The S. cerevisiae genome has only 53
C2H2 zinc finger domain-containing proteins [30]. BLAST
analyses independently identified the yeast orthologues of the
human genes, suggesting that the previous result reflected
meaningful sequence conservation. The yeast protein Rei1p
and its parologue, Reh1p (GenBank Accession Nos.
NP_009825 and NP_013491), were homologous to human
ZPR9, while the uncharacterized protein YDR049wp (Gen-
Bank Accession No. NP_010334) was homologous to human
FLJ10415. Both homologous groups contained multiple
dispersed C2H2 zinc fingers. In each case, only the N-
terminal zinc finger had significant homology to the C-
terminal zinc fingers of Chromalveolate and Magnoliophyte
SBDS proteins. BLAST-identified human and yeast zinc
fingers are shown aligned to SBDS zinc fingers (Fig. 4). Most
highly conserved residues in the specific consensus (derived
above) were maintained in these proteins (two exceptions are
boxed in Fig. 4).

The function of the zinc fingers in the ZPR9 and FLJ10415
families is not clear at present. Rei1p is localized to the
cytoplasm and has been implicated in mitotic signaling and bud

---

Fig. 4. Sequence alignment of zinc domains from plants and Chromalveolates and homologous sequences detected by BLAST analysis. Sequences from representative Magnoliophyte (flowering plants) and Chromalveolate SBDS homologues corresponding to the zinc finger domain (with downstream conserved residues) were aligned using ClustalX. Residues comprising the canonical C2H2 motif appear in bold and are denoted with arrows. We found that five residues outside this motif were also highly conserved (denoted with asterisks). Amino acid residues conserved in >50% of sequences and at least three taxonomic groups are shaded. These sequences are shown aligned to the best scoring sequence matches (“BLAST Hits”) from the S. cerevisiae (Rei1p, Reh1p, and YDR049wp) and human (ZPR9 and FLJ10415) genomes. Rei1p and Reh1p are the yeast homologues of ZPR9, while YDR049wp is the homologue of FLJ10415. Zinc finger domains in these sequences were homologous to the SBDS zinc finger (exceptions at highly conserved positions are boxed). ACA, Aphidium carterae; ATH, Arabidopsis thaliana; CHO, Cryptosporidium hominus; CPA, Cryptosporidium parvum; OSA, Oryza sativa; HVU, Hordeum vulgare; GMA, Gossypium arboreum; GMA, Glycine max; LSA, Lycopersicon esculentum; LSC, Lactuca sativa; PFA, Plasmodium falciparum; PRA, Phytophthora ramorum; PYY, Plasmodium yoelii yoelii; TAE, Triticum aestivum; TAN, Theileriua annulata; TGO, Toxoplasma gondii; TTH, Thalassiosira pseudonana.
formation in yeast [31] but the other proteins are not well characterized. Parallel studies of these proteins will likely prove relevant to the elucidation of the function of the Plantaee–Chromalveolate zinc finger.

We cloned a S. cerevisiae–Arabidopsis domain 3 chimera (ATHD3C) and determined that it was expressed in yeast (top, Fig. 3c) and fully complemented in the plasmid shuffle assay, yielding no significant growth defect (Fig. 3d and Table 3). Therefore, the C-terminal extension is compatible with core YLR022c function. Zinc finger domains often mediate protein interactions with nucleic acids [24]. Since they typically span and contact from 3 to 4 bases, the lone zinc finger domain in some SBDS homologues is not expected to confer a great deal of specificity to any related molecular interaction. Structural studies have indicated that SBDS domains 2 and 3 have homology to proteins with nucleic acid binding functions [6,7]. Therefore, the zinc finger domain may function in conjunction with other domains to mediate more specific interactions.

SBDS phylogeny is suggestive of an RNA metabolism and/or ribosome-related function

Phylogenetic profiling can yield insight into protein function [32]. We hypothesized that the phylogeny of SBDS might be shared with a group of proteins of related function and therefore searched the NCBI Clusters of Orthologous Groups (COGs) database [33] for genes with an identical pattern of occurrence in nature. Fifty-five conserved genes shared the SBDS family’s phylogeny. Analysis of assigned functional categories indicated that these COGs were markedly enriched for RNA metabolism (n = 17) and/or translation-associated functions (n = 29). Notably, a COG representing the homologues of the ribosome-associated Diamond–Blackfan anemia protein (RPS19) occurred in this group. Therefore, the phylogeny of SBDS is characteristic of a functionally coherent group of proteins and highly suggestive of a role in translation and/or RNA metabolism.

A recent study of SBDS localization revealed that it is enriched in the nucleolus [34]. This nuclear subdomain is the site of tRNA transcription, modification, processing, and ribosome assembly. It is also the site of assembly and maturation for other ribonucleoprotein complexes, such as telomerase and the components of the signal recognition particle [35]. Cartilage–hair hypoplasia (CHH; OMIM 250250), dyskeratosis congenita (DKC; OMIM 127550 and 305000), and Diamond–Blackfan anemia (DBA; OMIM 105650) are inherited human disorders associated with defects in the maturation, structure, or function of ribonucleoprotein complexes that assemble in the nucleolus [36–39]. These syndromes share interesting characteristics with SDS. Most strikingly, all are associated with decreased proliferative capacity of at least one myeloid lineage. All four syndromes have associated increased cancer risks [40–43], while CHH, DBA, and SDS also share skeletal phenotypes that are variable in both severity and penetrance [43–45]. Furthermore DBA and DKC, like SDS, appear to arise due to reduced function (i.e., hypomorphic mutations or haploinsufficiency) as opposed to complete loss of function in the affected pathway [40,46]. A nucleolar function for SBDS can therefore be considered in the context of these other syndromes in which partial impairments of housekeeping functions produce variable pleiotropic effects, including developmental and tissue-specific manifestations. The function of the DKC genes in the telomerase complex provides an intriguing molecular link to genome instability and cancer risk in this syndrome. Predisposition of SDS patients to AML must also reflect a consequence of impaired SBDS function. Further investigation is required to determine whether clinical manifestations are the direct or indirect consequences of SBDS deficiency and whether they arise due to impaired ribosome function or disruption of other nucleolus-associated pathways.

Methods

Data sources

SBDS homologues were identified by BLAST analysis (blastp and/or tblastn with default parameters) of the H. sapiens or Ara. thaliana sequences against the GenBank nonredundant, expressed sequence tag (EST), high-throughput genome sequence, and whole genome shotgun databases [8,47]. BLAST analysis was similarly performed with the genome and/or EST databases generated by the Protist EST Program, the Sanger Center, The Institute for Genome Research, the Joint Genome Institute, the Kazusa DNA Research Institute, and The Chlamydomonas Center (refer to Web site/database list). EST sequences were manually assembled as necessary. Gene annotation in genomic sequence was confirmed where EST sequences were available for comparison. A list of identified homologues with their GenBank accession numbers and species abbreviations are available as supplementary material (Table 1S).

Generation of sequence alignments

Multiple protein sequence alignment of completely sequenced homologues was performed using both ClustalX and MUSCLE algorithms [17,18,48] to decrease sensitivity to the deficiencies of each algorithm. The alignments were manually refined. Insertions that were not present in more than 10% of species were removed. Conserved residues were initially identified in a subset of the alignments composed of full-length eukaryotic SBDS homologues and excluding those species with two SBDS homologues (in which evolutionary constraints may be relaxed). Conserved positions were noted only if they were supported by both ClustalX and MUSCLE alignments. The analysis was repeated three times with the inclusion of (1) archaean homologues, (2) species with two SBDS homologues, and (3) partial sequences from 14 diverse protist species, including 2 species of green algae (P. wickhami and Chl. reinhardii), a Glaucophyte (Cyano. paradoxa), 3 additional Excavates (Trinastis pyriformis, M. jakobiiformis, and R. americana), 5 additional Chromalveolates (Phytophthora infestans, Guillardia theta, Emiliana huxleyi, Aph. carterae, Tox. gondii), a second Amoebzoan (Mas. balamuthi), an Ichthyosporan fungus (Cap. owczarzaki), and a single representative of the kingdom Rhizaria (B. natans). Some of these sequences were produced by the Protist EST Program. ClustalX and MUSCLE alignments of all sequences appear in Figs. 1S and 2S, respectively (partial sequences denoted with asterisks).

Phylogenetic analyses

Only maximum-aligned regions (corresponding to SBDS residues 18 to 237) were used in phylogenetic analyses (Figs. 1S and 2S). These analyses were performed with a ML method using PhyML [20], a MP method using the PHYLIP 3.6a program prospars [21], and a DM analysis using the PHYLIP programs protdist and neighbour. Bootstrap support was estimated using 100 replicates for all three methods. MP and DM methods were used with default parameters. Maximum-likelihood calculations were based on the JTT
substitution matrix, one category of substitution rates, and an estimated proportion of invariant sites.

**Determination of the structural context of conserved residues**

We used the crystal structures of the *A. fulgidus* homologue (AF0491) to infer domain boundaries in other homologues [6,7]. Conserved residues were mapped onto homologous sites in AF0491 using the previously generated ClustalX alignments. Swiss-PDBViewer version 3.7 (http://www.expasy.org/spdbv/) was used with structural coordinates (IT95 and IP9Q in the Protein Data Bank) to determine the location of each conserved residue, calculate hydrogen bonds, identify interacting residues, determine Φ and Ψ angles, and estimate the surface accessibility of R groups [6,7].

**Cloning of SBDS homologues**

All homologues were identified though BLAST analysis of respective genomic databases, with manual confirmation of gene annotation. The *S. cerevisiae*, *G. lamblia*, *L. major*, *Sul. solfataricus*, and *Hal. salinarum* sp. NRC-1 homologues are intronless genes and were isolated via two-step polymerase chain reaction (PCR) amplification [49] (sequences of all oligonucleotide primers used for PCR are available in Table 2S). In the first round of PCR, each open reading frame (ORF) was amplified from genomic DNA. In a subsequent PCR, each ORF was linked to an appropriate 5’ restriction site, 5’ Kozak consensus sequence [50], C-terminal HA epitope tag, stop codon, and unique 3’ restriction site through the use of tailed forward and reverse primers. The *G. lamblia* secondary homologue (GLA2a) had a 879-bp open reading frame, including two possible translation start sites upstream of the region of homology to other SBDS homologues. Expression plasmids were therefore prepared with the minimum-length (GLA2a) and maximum-length (GLA2b) N-termini. The *Sch. pombe* orthologue is composed of two coding exons. These were amplified in an initial round of PCR and linked together in a second round of PCR using a bridging oligonucleotide primer with homology to both exons. A final round of PCR was used to link 5’ and 3’ sequences for cloning and expression, as described above. *C. elegans* and *D. melanogaster* homologues were obtained as cDNA clones from Open Biosystems (clones OCE1182 and EDM1133, respectively). The *D. melanogaster* homologues carry a 163T → C mutation leading to a missense change (H55Y) with respect to reference sequences. This was cloned homologues using the forward primer (with a 5’ restriction site, Kozak consensus sequence, and 18 bp homologous to the ORF) employed previously. Reverse primers had 18 bp of sequences complementary to the coding strand of the YLR022c domain 2 and 18 bp complementary to the C-terminus of the test FYSH domain. Following amplification and purification of each FYSH domain, the PCR products were used as forward primers in a second PCR which linked them to YLR022c sequences encoding domains 2 and 3 via the homologous tail. In this PCR, a YLR022c reverse primer that included an HA epitope and appropriate restriction site for cloning was used. The FYSH domain forward primer (with a 5’ restriction site and Kozak sequence, as described above) was also added to the reaction to increase the yield of the full-length chimeric PCR fragment. These amplified products were purified, subjected to restriction enzyme digestion, and cloned into pRS415 as described above. Other chimeras were generated in a similar manner, using oligonucleotide primers spanning appropriate homologous junction points.

**Generation of the starting strain for plasmid shuffle**

The diploid yeast gene deletion strain 20519D (W303; Mat a/o; ura3-1/ura3-1; his3-11/his3-11; leu2-3_112/leu2-3_112; trp1Δ1/trp1Δ1; ade2-1/ade2-1; can1-100/can1-100; and YLR022C (1, 753): XanMX4/YLR022C) was purchased from Invitrogen. The starting strain for plasmid shuffle was generated as described by Shammam et al. [6], except that the expression vector pRS416 [Gal]: YLR022CFLAG was used to mask the genomic deletion.

**Plasmid shuffle**

The starting strain for plasmid shuffle was transformed with expression vectors for each homologue; YLR022c mutant, truncation, or interspecies chimera after a high efficiency transformation protocol [52]. Transformation reactions were carried out in duplicate. Each reaction was split and plated onto two types of synthetic dextrose (SD) medium to select for transformants through omission of leucine (SD –Leu). One type of medium promoted high expression from the Met25 promoter through exclusion of methionine (SD –Leu –Met). The other type promoted low expression through supplementation with a high level (800 μM) of methionine (SD –Leu +Met) as described [22]. The inclusion of uracil in both media relaxed selection for pRS416Gal:: YLR022CFLAG and allowed this plasmid to be lost if the expression cassette on the transformed plasmid was capable of functional complementation. We then actively selected against this plasmid by replica plating colonies onto either SD –Leu –Met or SD –Leu +Met, both supplemented with 0.1% 5-fluoroorotic acid (5-FOA; Toronto Research Chemicals). Positive complementation was indicated by growth of all replica-plated colonies following 5-FOA selection.

**Measurement of strain doubling times**

Following plasmid shuffle, three independent transformants for each strain were cultured overnight at 30°C to stationary phase (OD600 1.0) at the permissive culturing condition (SD –Leu –Met). Cultures were then diluted to OD600 0.05 in both SD –Leu –Met and SD –Leu +Met and grown for 5 to 6 h at 30°C to permit entry into the log phase. The OD600 of each culture was then measured at regular time intervals, and dilutions were made as required to maintain log phase (OD600 0.2 – 0.7). Average doubling times were calculated for each strain, and the statistical significance of any apparent difference from the wild-type growth rate was assessed using a Student t test.

**Detection of protein expression**

To assess protein expression, strains were grown to OD600 1.0 in SD –Leu –Met. Cells were collected by centrifugation and harvested by physical disruption with glass beads as described [52]. Protein was detected by immunoblotting as described [49] using a mouse anti-HA monoclonal antibody (Covance). Protein loading was assessed by subsequent immunoblotting with a rat anti-β-tubulin polyclonal antibody (Santa Cruz). We could not detect expression of the *Hal. salinarum* sp. NRC-1 homologue (data not shown), likely due to its high GC content (67%) compared to YLR022c (37%). We did not attempt to express the full-length *L. major* homologue, which had a similar
GC content (64%). We were also unable to detect steady-state expression of either the Sulfolobus solfataricus homologue or GLA1 (middle, Fig. 3c) possibly due to species differences in codon usage.

Acknowledgments

The authors thank Jennifer Haynes, Priya Kapoor, Richelle Sopko, and Brenda Andrews for advice regarding the yeast experiments; Andrew J. Roger, John Parkinson, and Patricia Romans for assistance with phylogenetic analysis; Alan J. Warren for advice on structural interpretations; members of the Protist EST Program (Patrick Keeling, B. Franz Lang, Robert W. Lee, Wolfgang Loeffelhardt, Ron Pearlman, and Andrew J. Roger) for access to their database; W. Ford Doolittle for

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2006.01.010.

References


J. He, et al., Targeted disruption of Dkc1, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice, Oncogene 21 (2002) 7740–7744.


URLs for internet databases

Chlamydomonas Center: http://www.chlamy.org/.
Joint Genome Institute, Department of Energy P. ramorum Project: http://genome.jgi-psf.org/ramorum1/ramorum1.home.html.
Kazusa DNA Research Institute Databases: http://www.kazusa.or.jp/eng/database/.
The Protein Data Bank: http://www.rcsb.org/pdb/.
The Sanger Institute Sequencing Genomics Projects: http://www.sanger.ac.uk/Projects/.
The Institute for Genome Research Genome Projects: http://www.tigr.org/tdb/.