

# Protist Homologs of the Meiotic *Spo11* Gene and Topoisomerase VI reveal an Evolutionary History of Gene Duplication and Lineage-Specific Loss

Shehre-Banoo Malik,<sup>\*1</sup> Marilee A. Ramesh,<sup>†1</sup> Alissa M. Hulstrand,<sup>\*</sup> and John M. Logsdon Jr.<sup>\*</sup>

<sup>\*</sup>Department of Biological Sciences, Roy J. Carver Center for Comparative Genomics, University of Iowa; and <sup>†</sup>Department of Biology, Roanoke College

*Spo11* is a meiotic protein of fundamental importance as it is a conserved meiosis-specific transesterase required for meiotic recombination initiation in fungi, animals, and plants. *Spo11* is homologous to the archaeobacterial topoisomerase VIA (*Top6A*) gene, and its homologs are broadly distributed among eukaryotes, with some eukaryotes having more than one homolog. However, the evolutionary relationships among these genes are unclear, with some debate as to whether eukaryotic homologs originated by lateral gene transfer. We have identified and characterized protist *Spo11* homologs by degenerate polymerase chain reaction (PCR) and sequencing and by analyses of sequences from public databases. Our phylogenetic analyses show that *Spo11* homologs evolved by two ancient eukaryotic gene duplication events prior to the last common ancestor of extant eukaryotes, resulting in three eukaryotic paralogs: *Spo11-1*, *Spo11-2*, and *Spo11-3*. *Spo11-1* orthologs encode meiosis-specific proteins and are distributed broadly among eukaryotic lineages, though *Spo11-1* is absent from some protists. This absence coincides with the presence of *Spo11-2* orthologs, which are meiosis-specific in *Arabidopsis* and are found in plants, red algae, and some protists but absent in animals and fungi. *Spo11-3* encodes a Top6A subunit that interacts with topoisomerase VIB (Top6B) subunits, which together play a role in vegetative growth in *Arabidopsis*. We identified *Spo11-3* (*Top6A*) and *Top6B* homologs in plants, red algae, and a few protists, establishing a broader distribution of these genes among eukaryotes, indicating their likely vertical descent followed by lineage-specific loss.

## Introduction

The Spo11 protein creates double-strand DNA breaks (DSBs) only during the early stages of meiosis. These breaks are repaired by homologous recombination. Spo11-mediated DSBs are required for the initiation of meiotic recombination and usually for homologous chromosome pairing during meiosis (Lichten 2001). Meiosis is necessary for sexual reproduction in eukaryotes and is the process by which eukaryotic cells divide twice, reducing their genetic material precisely to half. This results in the production of haploid cells called gametes. Following replication, the first meiotic division separates homologous chromosomes, whereas the second meiotic division separates sister chromatids. Sexual reproduction occurs when 2 gametes fuse, restoring the full diploid complement of genetic material in the resulting progeny. During Prophase I of meiosis, homologous chromosomes recombine, exchanging regions of DNA, which is a characteristic unique to meiotic division. Homologous recombination is necessary for successful meiosis: the exchange of homologous chromosomes and consequent formation of chiasmata cause the association of homologs until the proper time for separation, thus preventing chromosomal nondisjunction and the formation of aneuploid inviable meiotic products (reviewed by Baudat and Keeney 2001; Lichten 2001). That is, recombination physically ensures the proper segregation of chromosomes during meiosis. Chromosomes with new combinations of alleles are produced as a consequence of recombination, thereby increasing phenotypic variation; such a variation is advantageous to populations experiencing changing environments or strong selection (Agrawal 2006; Otto and Gerstein 2006).

<sup>1</sup> Equal contributions to this work.

Key words: Spo11, meiosis, recombination, evolution, recombination initiation, phylogeny, eukaryotes.

E-mail: john-logsdon@uiowa.edu.

*Mol. Biol. Evol.* 24(12):2827–2841. 2007

doi:10.1093/molbev/msm217

Advance Access publication October 5, 2007

Meiosis is one of several defining characteristics of eukaryotes. Current evidence supports a single origin for meiosis occurring early during the evolution of eukaryotes (Ramesh et al. 2005). Thus, meiosis likely played an important role in the success of the eukaryotic lineage. Genetic, molecular, and molecular phylogenetic analyses have shown that several key meiotic proteins are conserved, mainly in animals, fungi, and plants (Villeneuve and Hillers 2001; Ramesh et al. 2005). However, the range of organisms studied is relatively narrow, excluding many eukaryotic micro-organisms (protists). Any universal conclusions that may be made about meiosis are severely limited by the exclusion of protists, which represent the greatest phylogenetic diversity within the eukaryotic lineage (Dacks and Doolittle 2001). To fully understand such a fundamental eukaryotic process as meiosis, it is vital to analyze the level of conservation of its component machinery throughout all major eukaryotic lineages.

*Spo11* homologs are evolutionarily conserved and known to be present in animals, fungi, some plants, and a few protists (Villeneuve and Hillers 2001; Ramesh et al. 2005). These genes are orthologous to archaeobacterial topoisomerase VIA (*Top6A*), first identified in the crenarchaeote *Sulfolobus shibatae* and later in the euryarchaeote, *Methanocaldococcus jannaschii* (Atcheson et al. 1987; Bergerat et al. 1997; Nichols et al. 1999). Archaeobacterial topoisomerase VI is a type IIB topoisomerase that functions as a heterotetramer comprising of 2 molecules each of Top6A and topoisomerase VIB (Top6B), which act together to separate replicated chromosomes (Bergerat et al. 1997; Nichols et al. 1999; Corbett and Berger 2003b; Corbett and Berger 2005; Corbett et al. 2007).

Based on previous work characterizing the function of Spo11 in a variety of eukaryotic model organisms, we considered it among a set of genes that usually indicate the presence of meiosis (Ramesh et al. 2005). Species in which Spo11 was characterized include the fungi *Saccharomyces cerevisiae* (Atcheson et al. 1987; Keeney et al. 1997), *Schizosaccharomyces pombe* (Lin and Smith 1994), *Sordaria*

*macrospora* (Storlazzi et al. 2003), *Neurospora crassa* (Bowring et al. 2006), and *Coprinus cinereus* (Celerin et al. 2000; Merino et al. 2000); the animals *Caenorhabditis elegans* (Dernburg et al. 1998), *Drosophila melanogaster* (McKim et al. 1998; McKim and Hayashi-Hagihara 1998) and mouse (Keeney et al. 1999; Romanienko and Camerini-Otero 1999; Baudat et al. 2000); and the plant *Arabidopsis thaliana* (Hartung and Puchta 2000; Grelon et al. 2001; Stacey et al. 2006). In these systems, the meiosis-specific transesterase Spo11 is essential to the initiation of meiotic recombination and the production of viable meiotic products. These reports show that synapsis does not occur in Spo11 null mutants in mammals, fungi, and plants, although synaptonemal complex formation is not inhibited by Spo11 mutants in *D. melanogaster* and *C. elegans*. Mutational studies where Spo11 function was eliminated resulted in sterility or extremely low viability of gametes. Without Spo11, meiotic recombination is either absent or extremely infrequent in animals, fungi, and plants; as a result, meiosis fails (reviewed by Keeney 2001). Thus, the Spo11 protein serves an essential function in meiosis.

Spo11 is meiosis-specific; it is known to function only during meiosis in animals, fungi and plants and, where tested, creates DSBs that are required to initiate meiotic recombination among homologous chromosomes (Lin and Smith 1994; Keeney et al. 1997; Dernburg et al. 1998; McKim and Hayashi-Hagihara 1998; Romanienko and Camerini-Otero 1999; Baudat et al. 2000; Celerin et al. 2000; Merino et al. 2000; Grelon et al. 2001). Details of the initiation of meiotic recombination are best understood in *S. cerevisiae*, where 10 proteins are required to initiate meiotic recombination: Spo11, Ski8/Rec103, Rec102, Rec104, Rec114, Mei4, Mer2/Rec107, Xrs2, Mre11, and Rad50 (Malone et al. 1991; Arora et al. 2004; Keeney and Neale 2006; Maleki et al. 2007). However, this study focuses on Spo11 because sequences of the other 5 meiosis-specific proteins in this complex—Rec102, Rec104, Rec114, Mei4, and Mer2/Rec107—are not as well conserved. Generally less is known about their function compared with Spo11 and their homologs are not yet identifiable outside of fungi.

Studies in *Arabidopsis* reveal three *Spo11* homologs, *Spo11-1*, *Spo11-2*, and *Spo11-3* (the latter gene is also called *Bin5* or *root hairless 2/Rhl2*) (Hartung and Puchta 2000; Grelon et al. 2001; Hartung and Puchta 2001; Hartung et al. 2002; Sugimoto-Shirasu et al. 2002; Yin et al. 2002). *Arabidopsis Spo11-1* and *Spo11-2* genetically interact together; both are meiosis specific and are required for normal meiotic recombination (Grelon et al. 2001; Stacey et al. 2006). However, *Spo11-3* is functionally more similar to Top6A, so we will refer to it herein as “*Spo11-3* (Top6A)” (Hartung et al. 2002; Sugimoto-Shirasu et al. 2002; Yin et al. 2002). Consistent with the presence of *Spo11-3* (Top6A), *Arabidopsis* also possesses a homolog of archaeobacterial Top6B (also called *Bin3* or *hypocotyl 6/Hyp6*), which encodes the other component of a putative topoisomerase VI A<sub>2</sub>B<sub>2</sub> heterotetrameric protein as found in archaeobacteria (Hartung and Puchta 2000; Grelon et al. 2001; Hartung et al. 2002; Sugimoto-Shirasu et al. 2002; Yin et al. 2002). *Arabidopsis Spo11-3* (Top6A/Bin5) and

*Bin3* (Top6B) lack transit peptides for localization to the mitochondria or chloroplasts, and their mutant genes resemble a brassinosteroid-insensitive phenotype and have strong pleiotropic effects on cell growth and proliferation (Hartung and Puchta 2001; Hartung et al. 2002; Sugimoto-Shirasu et al. 2002; Yin et al. 2002). All three *Spo11* homologs and a *Bin3* (Top6B) homolog were also recently identified in *Oryza*; their expression profiles resemble those for *Arabidopsis* orthologs, and the *Spo11-3* (Top6A) and *Bin3* (Top6B) mutants were shown to reduce stress tolerance (Jain et al. 2006). Both *Spo11-3* (Top6A) and *Bin3* (Top6B) are targeted to the nucleus in *Oryza* and *Arabidopsis* (Sugimoto-Shirasu et al. 2002; Jain et al. 2006). Analyses of gene expression in somatic tissues and flowers of *Arabidopsis* and *Oryza* and mutant phenotypes in *Arabidopsis* indicate that *Spo11-3* (Top6A) and *Bin3* (Top6B) share similar phenotypes during vegetative growth and are expressed simultaneously (Hartung et al. 2002; Sugimoto-Shirasu et al. 2002; Yin et al. 2002; Jain et al. 2006). Yeast 2-hybrid analyses indicate interactions between *Bin3* (Top6B) proteins and both *Spo11-3* (Top6A), and *Spo11-2* in *Arabidopsis* and *Oryza* (Hartung and Puchta 2001; Jain et al. 2006). However, the interaction between *Spo11-2* and Top6B could be an artifact of the yeast 2-hybrid analyses because no genetic interaction was found *in vivo* between *Arabidopsis Spo11-2* and *Spo11-3*, and *Spo11-2* null mutants lack any somatic, nonmeiotic phenotype (Stacey et al. 2006). Furthermore, antisense mRNA sequence matching *Arabidopsis Spo11-2* was identified (Hartung and Puchta 2000). Although its functional role has not been demonstrated, the antisense transcript could hypothetically be involved in posttranscriptional regulation of *Spo11-2 in vivo* that would prevent the formation of *Spo11-2* protein in tissues when Top6B is coexpressed. The complex data from the variety of plant *Spo11* (Top6A) and Top6B homologs are unparalleled in any other group of eukaryotes, inviting interest in a more comprehensive survey of these genes across diverse eukaryotes in order to better understand their origin and evolutionary relationships (Gadelle et al. 2003; Jain et al. 2006).

Although the other conserved meiosis-specific genes have evolved by gene duplication, the detailed evolutionary history of *Spo11* homologs has been unclear and reflects a paucity of information from protists (Hartung et al. 2002; Hartung et al. 2002; Yin et al. 2002; Ramesh et al. 2005; Jain et al. 2006). In animals and fungi, Spo11 represents the only homolog of Top6A, functioning specifically in meiosis and apparently without a subunit homologous to Top6B. Prior to this report, the only known eukaryotic homologs of *Bin3* (Top6B) or orthologs of *Spo11-3* (Top6A) were limited to plants (Hartung and Puchta 2001; Jain et al. 2006). However, recent studies revealed 2–3 paralogous *Spo11* genes in several eukaryotic lineages, where the resemblance of these paralogs to *Spo11-3* (Top6A) or to *Spo11-1* or *Spo11-2* was not clearly phylogenetically resolved (Hartung et al. 2002; Hartung et al. 2002; Ramesh et al. 2005). Although there is consensus in the literature that *Arabidopsis Spo11* homologs are not the products of recent gene duplications in *Arabidopsis*, recent literature is peppered with speculation that eukaryotes may have acquired *Spo11-2*, *Spo11-3*, and *Top6B*

homologs by lateral gene transfers from archaeobacteria to plants because they are apparently absent from all other eukaryotes (Hartung et al. 2002; Yin et al. 2002; Corbett and Berger 2003a; Jain et al. 2006). By comprehensively searching available sequence data and using an explicit phylogenetic comparative approach, we identify protist *Spo11* homologs as either as *bona fide* orthologs of meiosis-specific *Spo11-1* or *Spo11-2* genes or as orthologs of *Arabidopsis Spo11-3* (*Top6A*). Thus, a clearer understanding of the evolution of this important meiotic gene family and the distribution of meiosis-specific genes across eukaryotes can be reached. In a complementary approach for identifying putative protist *Spo11-3* (*Top6A*) homologs, we searched to find whether or not the organisms that bear more than one *Spo11* homolog also have a homolog of *Bin3* (*Top6B*). If so, it suggests that a topoisomerase function exists, given “guilt by association.”

Our recent study of meiotic genes (including *Spo11*) conserved in a broad range of eukaryotes indicates that meiosis arose once, early during eukaryotic evolution (Ramesh et al. 2005). A clear delineation of meiotic *Spo11* homologs distinct from nonmeiotic *Spo11-3* (*Top6A*) homologs is necessary to better understand the evolution of meiotic recombination and to elucidate whether or not the meiosis-specific transesterase evolved as a more recent addition to the rest of the meiotic recombination machinery. To address this question, we used degenerate polymerase chain reaction (PCR) to isolate homologs of *Spo11* genes from the genomes of diverse protists and inferred the phylogenies of Spo11, Top6A, and Top6B homologs to further elucidate their evolutionary histories.

## Materials and Methods

Spo11, Top6A, and Top6B homologs in prokaryotes and eukaryotes were identified from public databases using Blast searches. Homology of the genes was validated by multiple sequence alignments and phylogenetic analyses of their inferred protein sequences to find if they are related by speciation events (orthology) or by gene duplication (paralogy). Degenerate oligonucleotides were designed from alignments constructed in 2000, and these were used for PCR amplification of homologous genes from other eukaryotes; identical data from genome sequence projects made available in 2006 were included for phylogenetic analyses.

### Database Searches

Literature and keyword searches of the National Center for Biotechnology Information (NCBI) protein database revealed homologs of Spo11, Top6A, and Top6B from various organisms. These protein sequences were used as queries for BlastP and PSI-BlastP searches (Altschul et al. 1997) of the NCBI nonredundant database, and TblastN searches of the NCBI databases of expressed sequence tags (ESTs), high-throughput genome sequences, whole-genome shotgun, and genome survey sequences (GSSs) between August 2000 and December 2006. These searches included the genome sequences deposited at NCBI for several protists that are parasites or opportunistic pathogens, such as *Giardia*, *Entamoeba*,

trypanosomes, and apicomplexans (see supplementary table S1, Supplementary Material online). Similarly, Spo11, Top6A, and Top6B homologs were retrieved (when present) by BlastP, TblastN, and keyword searches of genome sequence databases of several free-living protists including a ciliate (*Tetrahymena*), heterolobosean (*Naegleria*), choanoflagellate (*Monosiga*), stramenopiles, red and green algae, the pathogen *Trichomonas*, as well as plants and animals (see supplementary table S1, Supplementary Material online) at the Joint Genome Institute and The Institute for Genomic Research. The best hits from these searches were used as queries for BlastP or BlastX against the NCBI nonredundant database for validation as homologs of Spo11, Top6A, and Top6B. Once unannotated nucleotide sequences were retrieved from databases (see supplementary table S2, Supplementary Material online), the sequences were assembled and putative open reading frames annotated using Sequencher 4.6 (Genecodes, Ann Arbor, MI). Putative start and stop codons and exons were assigned with reference to pairwise comparisons (BlastX of GenBank) with homologous proteins and multiple amino acid sequence alignments that were made with ClustalX 1.83 (Chenna et al. 2003) and refined with MacClade 4.08 (Maddison WP and Maddison DR 2006). Similarly, because the GenBank entry for the *N. crassa* Spo11 predicted protein (GenInfo Identifier (GI) no. 7635881) differed from the published sequence (Bowring et al. 2006) in motif 3, we inferred a new mRNA-coding sequence and translation of the *N. crassa* *Spo11* gene GI no. 28922555:18032–19821) with reference to the predicted amino acid sequence of the *S. macrospora* Spo11 homolog (GI no. 33304618). The new *Neurospora* Spo11 protein annotation is given in supplementary figure S1, Supplementary Material online.

### Sources of DNA Templates for PCR

Miklós Muller (Rockefeller University) provided the genomic DNA from *Trichomonas vaginalis* strain NIH-C1. Jeff Cole American Type Culture Collection (ATCC, Manassas, VA) provided the genomic DNA of *Malawimonas jakobiformis* ATCC 50310. Michael Gray (Dalhousie University) provided the genomic DNA from *Acanthamoeba castellanii* strain Neff (ATCC 30010). Graham Clark (London School of Hygiene and Tropical Medicine) provided the genomic DNA of *Entamoeba histolytica* isolate HM-1:IMSS. Rick Tarleton provided the genomic DNA from *Trypanosoma cruzi* Brazil strain, Kojo Mensa-Wilmot provided the genomic DNA from *Trypanosoma brucei* and *Leishmania major* strain Friedlin, and Boris Striepen provided the genomic DNA of *Cryptosporidium parvum* Iowa, each from the University of Georgia.

### Isolation of Genes

Degenerate oligonucleotides were designed corresponding to conserved amino acid sequences (fig. 1 and supplementary fig. S2 and supplementary table S3, Supplementary Material online) and used in PCR to amplify partial sequences of eukaryotic *Spo11* homologs from *T. vaginalis*, *T. cruzi*, and *L. major*. Multiple (24) combinations

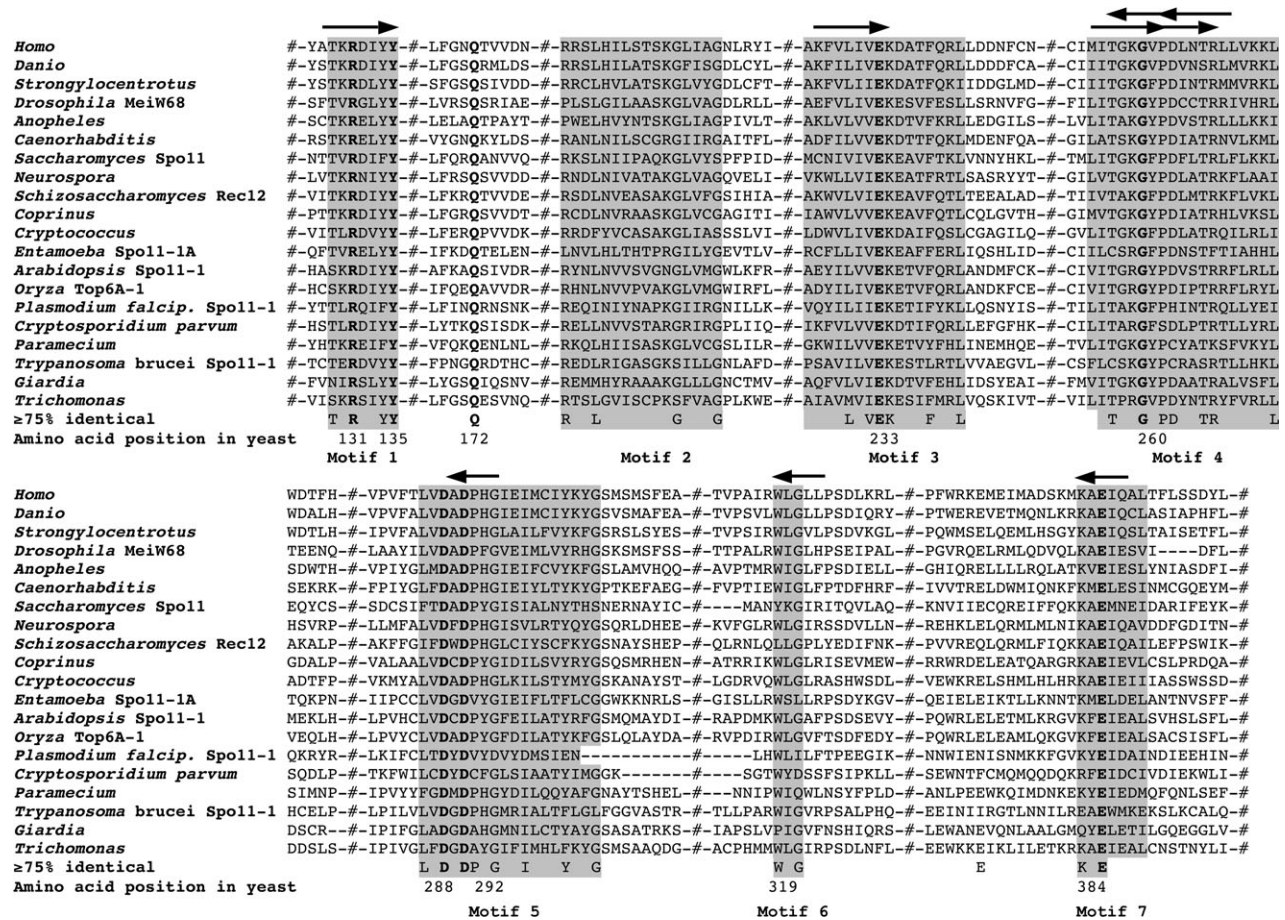


Fig. 1.—An alignment of Spo11-1 homologs reveals 7 conserved regions. Amino acid residues with 100% identity are bold. Conserved motifs are shaded in gray, including motifs 1–5 identified previously (Bergerat et al. 1997). Amino acid positions are numbered relative to the *Saccharomyces* ortholog (Diaz et al. 2002). Arrows designate conserved amino acid sequences from which degenerate PCR primers were designed, gaps in the alignment are represented by (-), and ambiguously aligned data by (#). For a comparison of Spo11-1 sequences with Spo11-2, Spo11-3, and Top6A, see supplementary figure S2, Supplementary Material online.

of forward and reverse degenerate primers were initially used. To verify and extend GSSs found in GenBank in 2001, exact match primers were designed from these data and used to amplify *Spo11* homologs from *E. histolytica* and *T. brucei*. Also, *Spo11* was amplified from *C. parvum* and *T. cruzi* with exact match primers designed in 2004 from genome sequences in GenBank, and *Acanthamoeba Spo11* and *Malawimonas Spo11-3 (Top6A)* were amplified with exact match primers designed from GSSs and ESTs found in GenBank in 2006 (supplementary table S4, Supplementary Material online).

Gene fragments were amplified from total DNA by PCR with Ex Taq PanVera (Invitrogen, Carlsbad, CA) and Eppendorf MasterTaq (Hamburg, Germany), as recommended by the manufacturers, with 10 ng DNA, 250 μM each dNTP (Stratagene, La Jolla, CA), 1.5 mM MgCl<sub>2</sub>, and 1 μM each primer (synthesized at Integrated DNA Technologies [IDT, Coralville, IA]) per reaction. Reaction conditions were 95 °C for 2 min followed by 40 cycles at 92 °C for 1.5 min, 45–60 °C for 1 min, 72 °C for 1.5 min + 6 s/cycle, and ending at 72 °C for 7 min. Secondary amplifications of degenerate PCR products using internally nested primers to preferentially amplify *Spo11* were performed using a 1:100 or 1:1,000 dilution of the initial degenerate PCR

product. Reaction reagents were the same concentrations as in the initial reaction. Reaction conditions were 95 °C for 2 min followed by 40 cycles at 92 °C for 1.5 min, 44 °C for 1.5 min, 72 °C for 1.5 min + 6 s/cycle, and ending at 72 °C for 7 min.

PCR products were run on a 0.5:1.5% low-melting agarose:NuSieve GTG agarose (Fisher, Pittsburgh, PA) and BioWhittaker, Walkersville, MD) in 1× TAE buffer. DNA bands were excised from the gel and cloned directly into the pCR4.0-TOPO vector (Invitrogen) according to the manufacturer’s instructions. Putative clones were screened according to the size of their plasmid inserts by PCR with M13 forward versus M13 reverse primers, cycling at 94 °C for 2 min followed by 30 cycles at 94 °C for 1 min, 57 °C for 1 min and 72 °C for 90 s, and ending at 72 °C for 5 min (reagents from Invitrogen, Promega, [Madison, WI] and Stratagene) (Sandhu et al. 1989). Plasmid clones were isolated (Eppendorf FastPlasmid Kit [Eppendorf, Hamburg, Germany] and Qiagen [Valencia, CA]) and sequenced (ABI BigDye 3.1 [Applied Biosystems, Foster City, CA]; MJ Research BaseStation [Waltham, MA]) completely on both strands of at least 2 clones using M13 forward, M13 reverse, and T7 primers (Invitrogen), except for *Leishmania Spo11*, which was sequenced in only one

direction from a single clone and immediately found to exactly match a scaffold in the genome project.

### Phylogenetic Analysis

Phylogenetic analyses were used to infer the evolutionary relationships of the Spo11, Top6A, and Top6B homologs. The full-length protein sequences derived from the databases were used for all analyses because the sequences of our PCR products were all identical to those available from genome sequencing projects in GenBank. Multiple alignments of amino acid sequences were initially constructed using ClustalX 1.83 (Chenna et al. 2003) and inspected and adjusted manually using MacClade 4.08 (Maddison WP and Maddison DR 2006). Alignments are available upon request from J.M.L. Only unambiguously aligned amino acid sites were used for phylogenetic analyses: columns of the alignments containing ambiguously aligned regions and gaps introduced by insertions or deletions were deleted. Phylogenies are unrooted and also rooted when possible, using either paralogs or prokaryotic orthologs as outgroups.

MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) was used for analyses of each protein alignment. MrBayes was run for  $10^6$  generations, with 4 incrementally heated Markov chains, sampled every 1,000 generations with the temperature set to 0.5. Among-site, substitution rate heterogeneity was corrected using in-varying and 8 gamma-distributed substitution rate categories and the WAG model for amino acid substitutions (Whelan and Goldman 2001), abbreviated herein as WAG + I + 8G. The consensus tree topology, the arithmetic mean log-likelihood for this topology, and branch support were estimated from the set of sampled trees with the best posterior probabilities. The number of trees included in this set varied among analyses. Means and 95% confidence intervals for the gamma-distribution shape parameter ( $\alpha$ ) and the proportion of invariable sites (pI) were also estimated for each alignment that was analyzed. The opisthokonts (animals and fungi), when present, were constrained as a group for Bayesian analyses. Bootstrap analyses were performed with 100 replicates with the SEQBOOT, PROML, and CONSENSE programs of PHYLIP 3.61 (Felsenstein 2005), with single categories of in-varying and gamma-distributed amino acid substitution rates and the JTT model for amino acid substitutions (Jones et al. 1992), and the pI and coefficient of variation input using values estimated by MrBayes.

Numerous phylogenetic analyses were performed on complete data sets as well as on subsets of the data, in an effort to reduce any systematic bias that may be introduced to the analyses by including very diverged or distantly related sequences (Felsenstein 1978). Thus, phylogenetic analyses were performed on Spo11 homologs with and without their prokaryotic Top6A orthologs (data not shown), without Spo11-3 (Top6A) paralogs and prokaryotic orthologs, without Spo11-2 or Spo11-3 (Top6A) paralogs or prokaryotic orthologs, with Spo11-3 (Top6A) and prokaryotic orthologs alone, and with only Spo11-3 (Top6A) orthologs. Phylogenetic analyses of Top6B homologs were performed on all eukaryotic and prokaryotic homologs, with and without apicomplexans,

which only shared homology in the N-terminal region of the alignment.

### Tree Topology Tests

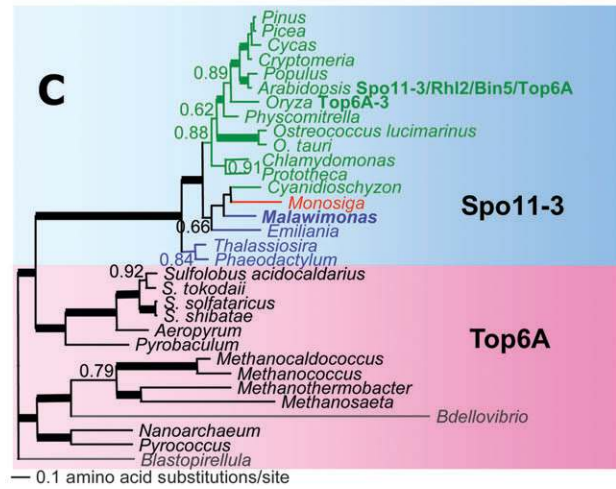
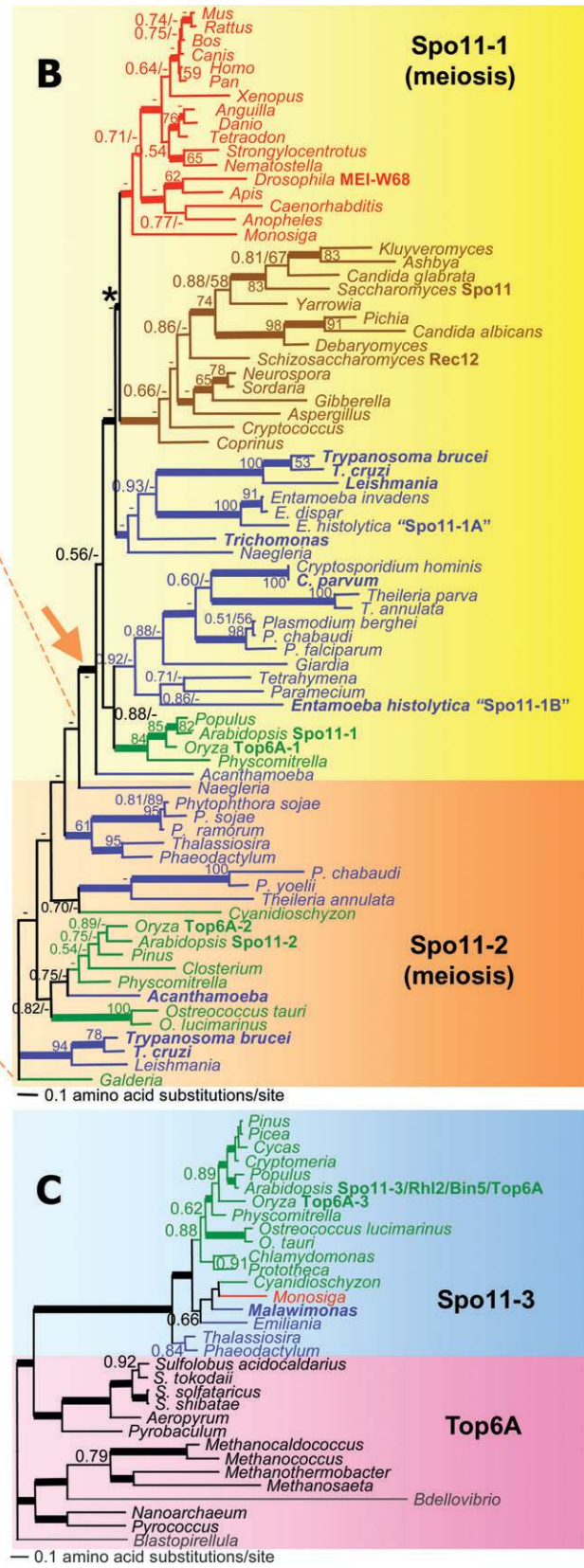
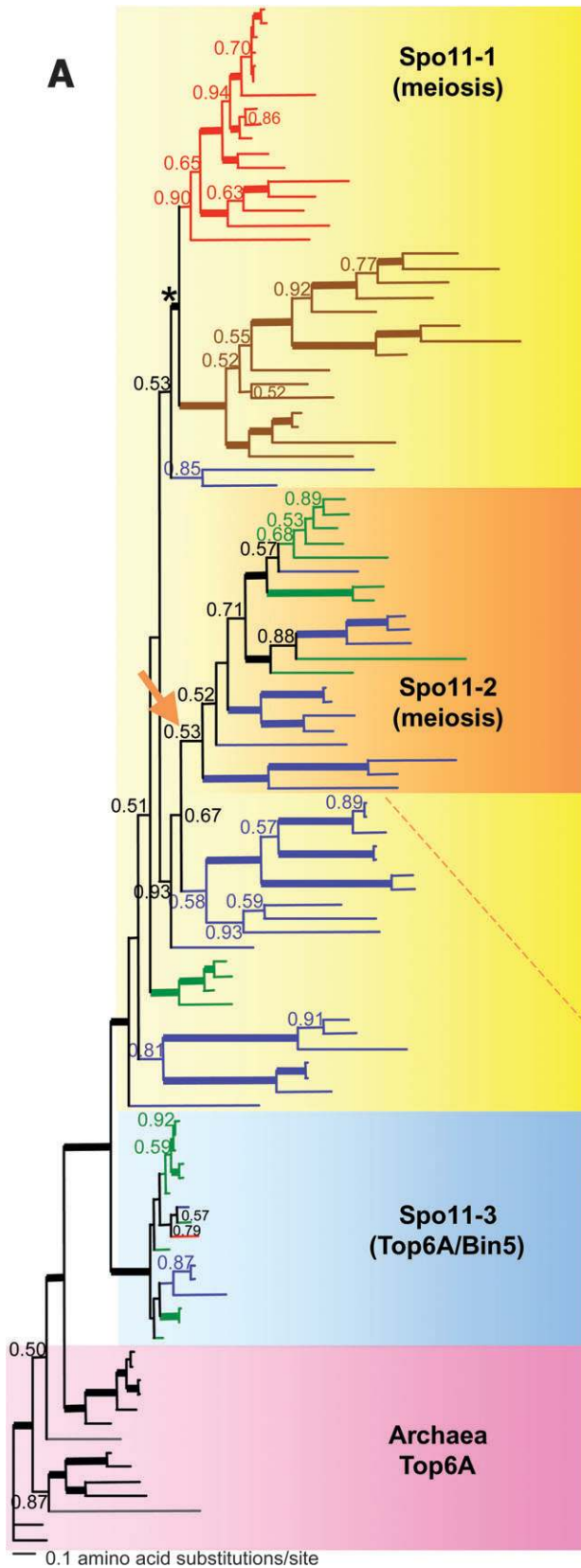
The Bayesian consensus tree topology shown for Spo11-1 and Spo11-2 homologs was compared with 8 alternate tree topologies that varied the placement of *Acanthamoeba* and *Entamoeba* Spo11 homologs. Alternate tree topologies were generated using MacClade 4.08 (Maddison WP and Maddison DR 2006). The site-by-site likelihoods for each of the trees were calculated using Tree-Puzzle 5.2 (Schmidt et al. 2002) with the WAG + I + 8G substitution model and using the mean  $\alpha$  and pI values that were calculated by MrBayes. This information was input into CONSEL (Shimodaira and Hasegawa 2001), which was used to compare the 9 tree topologies with the approximately unbiased (AU) test (Shimodaira 2002), and with unweighted and weighted Kishino–Hasegawa tests (Kishino and Hasegawa 1989; Goldman et al. 2000; Shimodaira and Hasegawa 2001) and Shimodaira–Hasegawa tests (Shimodaira and Hasegawa 2001).

## Results and Discussion

### Gene Discovery

*Spo11* homologs were isolated by degenerate PCR from *T. vaginalis*, *T. cruzi*, and *L. major* and by exact match PCR from *T. brucei*, *T. cruzi*, *C. parvum*, *E. histolytica*, *A. castellani*, and *M. jakobiformis*. Matching open reading frames exist in the genome or EST sequencing projects of these organisms, which are summarized in supplementary table S5, Supplementary Material online. Our *Malawimonas Spo11-3 (Top6A)* fragment, cloned from total DNA using PCR primers designed from an EST, has 4 introns, consistent with the intron density reported in other *Malawimonas* protein-coding genes (Archibald et al. 2002). All sequences have been deposited in GenBank, accession numbers EF199879–EF199888, as shown in detail in supplementary table S5, Supplementary Material online. Orthologs of Top6A and Top6B could not be found in eubacterial genome sequences, except for the planctomycete *Blastopirellula* and the delta-proteobacteria *Bdellovibrio* and *Anaeromyxobacter*. The notable absence of cyanobacterial, plastid, alpha-proteobacterial, mitochondrial, or any other eubacterial homologs of Top6A or Top6B make it unlikely that eukaryotes acquired these genes from eubacteria by primary endosymbiotic gene transfer.

No homologs of *Spo11*, *Top6A*, or *Top6B* were detected in the genome sequence of *Dictyostelium discoideum* strain AX4 (Eichinger et al. 2005), which lives mainly as haploid cells reproducing mitotically with occasional diploidy and parasexual mating (King and Insall 2006). Strain AX4 is derived from strain AX3 that was subject to mutagenesis for axenic cultivation (Kessin 2006). This is the only organism in our survey with a complete ( $>8\times$  coverage) genome sequence that is missing a meiosis-specific *Spo11* homolog. Synaptonemal complexes have been observed in a related species, *Dictyostelium mucoroides* (Macinnes and Francis 1974), and we did find Spo11 homologs in the more distantly related “amoebozoans” *Acanthamoeba* and *Entamoeba*,



which suggests that meiosis-specific *Spo11* homologs were present in the common ancestor of “Amoebozoa” and lost secondarily in a recent ancestor of *D. discoideum* AX4. The absence of these genes might instead be due to incomplete coverage of the genome sequence, in which some gaps remain, or it could reflect that this strain is actually ameiotic, possibly as a result of mutagenesis during axenic cultivation.

The conserved regions of aligned homologs of meiosis-specific Spo11-1 found in this study among animals, fungi, plants, and diverse protists are highlighted in fig. 1. Several conserved motifs identified previously (Bergerat et al. 1997; Diaz et al. 2002) are indicated, as well as additional conserved residues that warrant further functional investigation. Arrows on the alignment indicate conserved regions corresponding to degenerate oligonucleotide primers for PCR that are summarized in supplementary table S3, Supplementary Material online.

### Evolution of *Spo11* Homologs by Gene Duplications

We identified homologs of *Spo11* by PCR in some protists and bioinformatically in animals, fungi, plants, and representatives of diverse protist lineages (figs. 2 and 3). Eukaryotic *Spo11* homologs fell into 3 distinct groups, shown in figure 2A, representing the paralogous genes *Spo11-1*, *Spo11-2*, and *Spo11-3* (also see supplementary fig. S2, Supplementary Material online). Phylogenetic analyses in figures 2A and B further delineate eukaryotic orthologs of meiosis-specific *Spo11-1* and *Spo11-2* (Grelon et al. 2001; Stacey et al. 2006). Notably, diverse eukaryotic *Spo11-2* orthologs are shown to be distinct from *Spo11-3* orthologs in figure 2A. Figure 2B reinforces the result seen in figure 2A that *Spo11-2* orthologs are also distinct from *Spo11-1*, when more distant *Spo11-3* and *Top6A* homologs are not considered, making more amino acid sites available for phylogenetic analysis. Figure 3 illustrates a phylogenetic analysis of meiotic Spo11-1 orthologs considered alone without other Spo11 paralogs or prokaryotic orthologs. Figure 2A also indicates that *Spo11-3* (*Top6A*), also known as *Bin5* or *Rhl2*, evolved by an early eukaryotic gene duplication event prior to the divergence of *Spo11-2* and *Spo11-1* genes. Spo11-3 (*Top6A*) proteins interact with topoisomerase VIB (*Bin3* [*Top6B*], fig. 4) in a nonmeiotic role in *Arabidopsis* and *Oryza* (Hartung et al. 2002; Sugimoto-Shirasu et al. 2002; Yin et al. 2002; Jain et al. 2006). In addition to

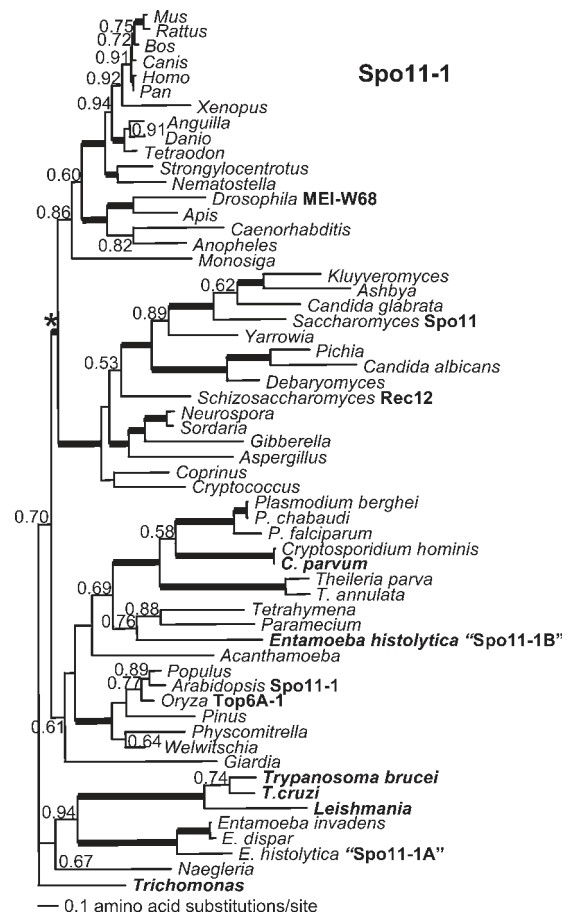


FIG. 3.—Unrooted phylogenetic tree of Spo11-1 homologs. This tree is the consensus topology of the best 850 trees estimated by Bayesian inference from 165 aligned amino acids. An asterisk marks the topological constraint on the node that unites animals and Fungi (opisthokonts). Thickened lines represent posterior probabilities of 0.95–1.00. Numbers at the nodes correspond to Bayesian posterior probabilities  $\geq 0.50$ . The names of organisms that are highlighted in bold represent PCR products isolated in this study. The unconstrained tree topology is shown in supplementary figure S6, Supplementary Material online, with sequences identified according to their NCBI GI or locus identification numbers in supplementary tables S2 and S5, Supplementary Material online. LnL =  $-13,584.69$ ,  $\alpha = 1.86$  ( $1.42 < \alpha < 2.32$ ),  $pI = 0.054$  ( $0.0015 < pI < 0.12$ ).

figure 2A, analyses of Spo11-3 orthologs that employ more amino acid sites by excluding Spo11-1 and Spo11-2 orthologs place eukaryotic orthologs in a distinct group separate from prokaryotic orthologs (fig. 2C), rather than nested

FIG. 2.—Phylogenetic analyses of all Spo11 and Top6A homologs. Trees were estimated from aligned amino acids by Bayesian inference, with an asterisk indicating that the node uniting animals and fungi (opisthokonts) is constrained. Thickened lines represent posterior probabilities of 0.95–1.00. Numbers at the nodes correspond to Bayesian posterior probabilities  $\geq 0.50$  and in (B), the percent bootstrap support  $\geq 50\%$  from 100 replicates of PROML, with the lack of bootstrap support indicated by a dash (-). Organism names highlighted in bold represent PCR products isolated in this study. Sequences are identified in supplementary figure S3 and supplementary tables S2 and S5, Supplementary Material online according to their NCBI GI or locus identification numbers. Unconstrained tree topologies are shown in supplementary figures S3–S5, Supplementary Material online. Animals and *Monosiga* are highlighted in red, Fungi in brown, “Archaeplastida” in green, and protists in blue. A yellow box highlights meiosis-specific orthologs, with Spo11-2 orthologs further highlighted by an orange box, arrows and dotted lines, Spo11-3 orthologs in a blue box, and prokaryotic orthologs in a pink box. (A) The consensus topology of the 850 best trees rooted with archaeobacterial Top6A homologs, inferred from 180 residues. LnL =  $-25,215.35$ ,  $\alpha = 1.42$  ( $1.18 < \alpha < 1.71$ ),  $pI = 0.037$  ( $0.0036 < pI < 0.097$ ). (B) The consensus topology of the 650 best trees of Spo11-1 homologs rooted arbitrarily with Spo11-2, inferred from 162 residues. LnL =  $-18,408.17$ ,  $\alpha = 1.26$  ( $1.004 < \alpha < 1.57$ ),  $pI = 0.035$  ( $0.0016 < pI < 0.098$ ). (C) The consensus topology of the 980 best trees of Spo11-3 (*Top6A*/*Bin5*/*Rhl2*) homologs rooted with archaeobacterial Top6A homologs, inferred from 333 residues. LnL =  $-11,029.89$ ,  $\alpha = 2.30$  ( $1.77 < \alpha < 2.94$ ),  $pI = 0.034$  ( $0.010 < pI < 0.063$ ). LnL, log-likelihood.

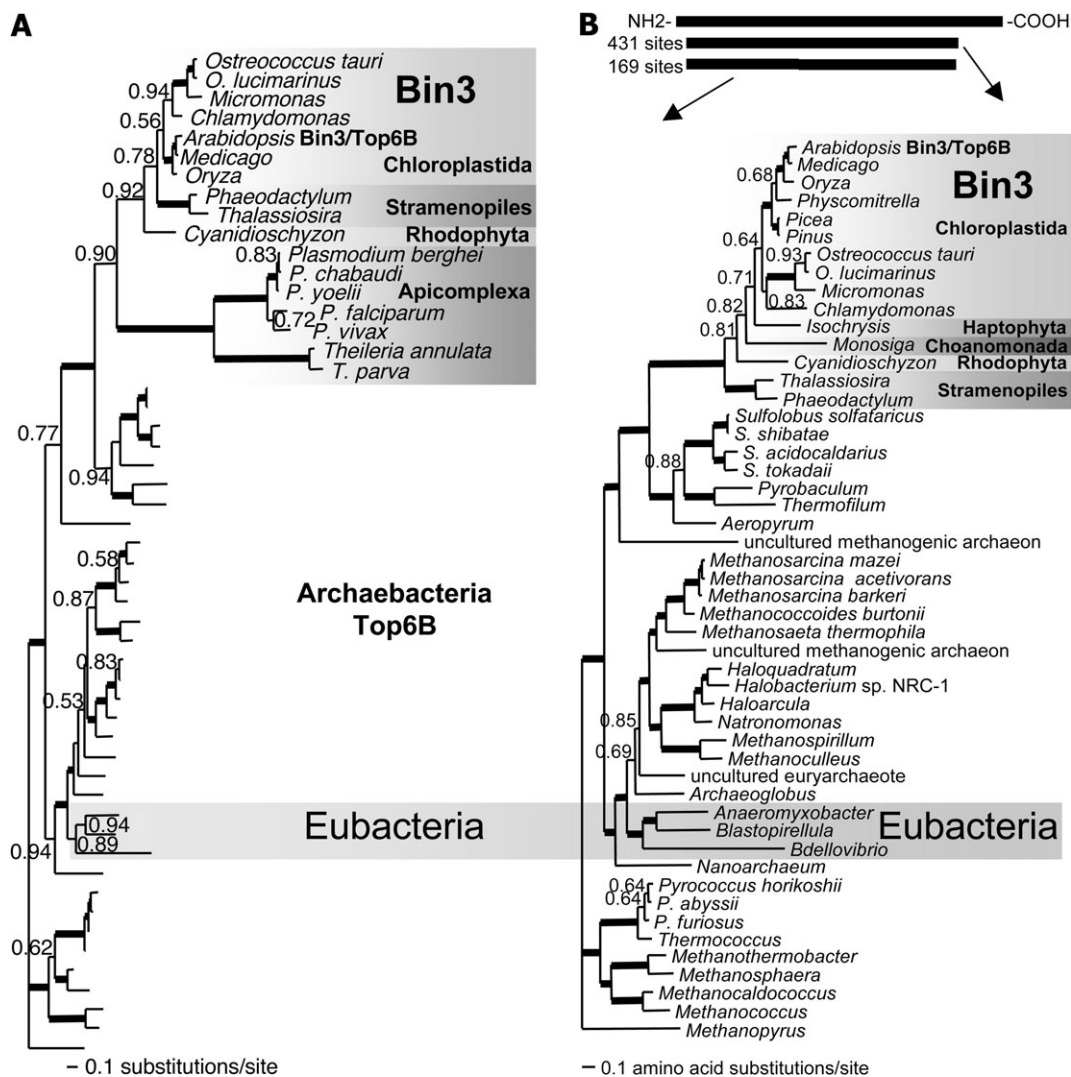


FIG. 4.—Phylogenetic analyses of Bin3 homologs rooted with archaeobacterial Top6B orthologs. Trees were estimated from aligned amino acids by Bayesian inference. Numbers at the nodes correspond to Bayesian posterior probabilities. Thickened lines represent posterior probabilities of 0.95–1.00. Sequences are identified in supplementary figure S7 and supplementary table S2, Supplementary Material online according to their NCBI GI or locus identification numbers. For conserved amino acids and motifs in representative organisms, see supplementary figure S8, Supplementary Material online. (A) The consensus topology of the 950 best trees inferred from 169 residues at the N-terminal domain. LnL =  $-9,368.45$ ,  $\alpha = 1.28$  ( $0.97 < \alpha < 1.67$ ),  $pI = 0.052$  ( $0.0039 < pI < 0.11$ ). (B) The consensus topology of the 900 best trees inferred from 431 residues with *Monosiga* and without Apicomplexa. LnL =  $-24,375.90$ ,  $\alpha = 1.98$  ( $1.67 < \alpha < 2.31$ ),  $pI = 0.062$  ( $0.038 < pI < 0.090$ ). LnL, log-likelihood.

within any prokaryotic group. This demonstrates clearly that eukaryotic and prokaryotic *Spo11* (*Top6A*) homologs descended from the last common ancestor of eukaryotes and archaeobacteria and are not related by lateral gene transfer from archaeobacteria to eukaryotes. Similarly, figure 4 shows that eukaryotic homologs of Bin3 (*Top6B*) also descended from the last common ancestor of eukaryotes and archaeobacteria and are also not related by lateral gene transfer from archaeobacteria to eukaryotes. As expected, we found *Spo11-3* (*Top6A*) and *Top6B* homologs to be distributed in similar groups of organisms. Our analyses indicate that eukaryotic *Top6A* homologs evolved by a gene duplication event prior to the divergence of extant eukaryotes that led to the *Spo11-3* (*Top6A*) (shown in blue in fig. 2) and 2 meiosis-specific *Spo11* lineages (i.e., *Spo11-1* and

*Spo11-2*, shown in yellow in fig. 2A, with *Spo11-2* indicated in orange). Phylogenetic analyses of eukaryotic *Spo11-3* (*Top6A*) and Bin3 (*Top6B*) homologs alone (fig. 5A and B) in the absence of prokaryotic orthologs do not indicate any specific relationships between green or red algae and *Monosiga*, *Malawimonas*, diatoms, and haptophytes. This tentatively suggests that *Spo11-3* (*Top6A*) and Bin3 (*Top6B*) orthologs in protists are not derived by secondary endosymbiotic gene transfer from red or green algae and that *Spo11-3* (*Top6A*) and Bin3 (*Top6B*) orthologs were lost from several eukaryotic lineages for which complete genome sequence data are available. A possibility requiring further investigation is that the root of the eukaryotic tree lies between the “Archaeplastida” and the common ancestor of the rest of eukaryotes (which



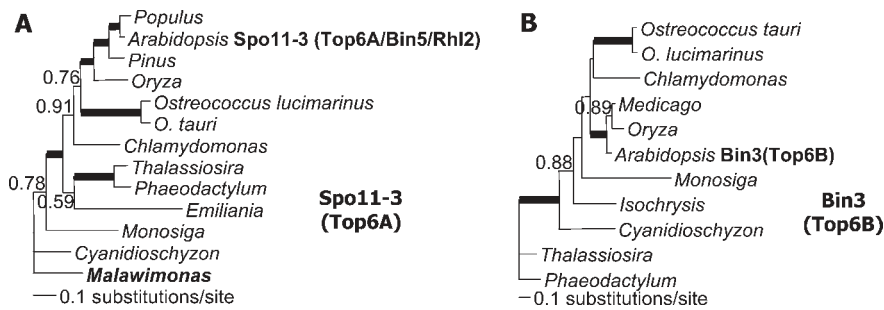


FIG. 5.—Phylogenetic analyses of eukaryotic Spo11-3 (Top6A) and Bin3 (Top6B) homologs. Trees were estimated from aligned amino acids by Bayesian inference. Numbers at the nodes correspond to Bayesian posterior probabilities. Thickened lines represent posterior probabilities of 0.95–1.00. Sequences are identified in supplementary figure S9 and supplementary table S2, Supplementary Material online according to their NCBI GI or locus identification numbers. For phylogenetic trees including *Monosiga* and rooted with prokaryotic orthologs, see figure 2C and supplementary figure S4C, Supplementary Material online. (A) The consensus topology of the 900 best trees inferred from 338 residues of Spo11-3 (Top6A) homologs. LnL =  $-3,891.07$ ,  $\alpha = 0.94$  ( $0.57 < \alpha < 1.60$ ),  $pI = 0.15$  ( $0.012 < pI < 0.27$ ). (B) The consensus topology of the 950 best trees inferred from 482 residues of Bin3 (Top6B) homologs. LnL =  $-6,531.56$ ,  $\alpha = 1.32$  ( $0.89 < \alpha < 1.92$ ),  $pI = 0.081$  ( $0.0073 < pI < 0.16$ ). LnL, log-likelihood.

lost *Spo11-3* [*Top6A*] and *Bin3* [*Top6B*]), and both *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) were then laterally transferred from “Archaeplastida” independently to *Monosiga*, *Malawimonas*, diatoms, and haptophytes.

Within the meiosis-specific *Spo11* lineage, the *Spo11-1* and *Spo11-2* genes are separated by another gene duplication (fig. 2B). Because a well-supported clade consistently recovered in our analyses (some data not shown) separated duplicate Spo11 homologs from trypanosomes, plant Spo11-2, apicomplexans, *Acanthamoeba*, and *Naegleria*, we concluded that all of the members of this clade are orthologs of Spo11-2 (orange box, fig. 2), whether or not they have extant duplicates in the Spo11-1 clade. Notably, all eukaryotic lineages with complete genome sequences included in our study have orthologs of *Arabidopsis* meiosis-specific genes *Spo11-1*, *Spo11-2*, or both (table 1 and figs. 2B and 6). However, *Spo11-1* may have also subsequently evolved by lineage-specific gene duplication in *Entamoeba* (fig. 3), which lacks a *Spo11-2* ortholog in all of our analyses.

The AU test significantly rejects ( $P < 0.05$ ) that *E. histolytica* Spo11-1B is a sister of *Acanthamoeba* Spo11-1 or *Acanthamoeba* Spo11-2 or that it could be a Spo11-2 (supplementary table S6, Supplementary Material online). Alternate topologies placing either of the *E. histolytica* duplicates in the Spo11-2 clade, placing *Acanthamoeba* Spo11-2 in the Spo11-1 clade, or placing *Acanthamoeba* Spo11-1 in the Spo11-2 clade were significantly rejected ( $P < 0.05$ ) by the AU test (supplementary table S6, Supplementary Material online). The AU test supported the topology shown in figure 2B as the best ( $P = 0.903$ ), with the second-best tree topology having *Acanthamoeba* Spo11-1 as a sister to *E. histolytica* Spo11-1A ( $P = 0.298$ ). In other words, it seems likely that *Entamoeba* did have a lineage-specific gene duplication of *Spo11-1*. All of the results from the AU test and other topology tests run in CONSEL are summarized in supplementary table S6, Supplementary Material online.

The gene duplication separating *Spo11-1* and *Spo11-2* likely occurred early in eukaryotic evolution (fig. 6). *Spo11-2* homologs are present in land plants, green algae, and red algae, and among protists, in stramenopiles, api-

complexans, kinetoplastids, *Naegleria*, and *Acanthamoeba* (figs. 2B and 6). However, *Spo11-2* is not found in animals, *Monosiga*, fungi, *Entamoeba*, ciliates, *Cryptosporidium*, *Trichomonas*, and *Giardia*, and *Spo11-1* is not found in green and red algae, and stramenopiles. This pattern indicates that the gene duplication from which *Spo11-1* and *Spo11-2* originated preceded the divergence of any of these eukaryotic lineages and was then followed by some lineage-specific losses (fig. 6). The *Spo11-1/Spo11-2* gene duplication was likely preceded by (or coincident with) the origin of meiosis because both *Arabidopsis* *Spo11-1* and *Spo11-2* are meiosis-specific. Consistent with this hypothesis, mRNA of a *Spo11-2* homolog was found in the “sexual reproductive stage” of a green alga, *Closterium*, as noted in its GenBank entry (GI no. 40000781). Thus, we infer from our phylogenetic analysis that the close relatives of plant and green algal Spo11-2 are also meiosis specific (fig. 2B).

*Arabidopsis* *Spo11-1* and *Spo11-2* were recently shown to genetically interact together in meiosis because double mutant heterozygotes exhibit nonallelic noncomplementation, and both genes are required to achieve wild-type levels of meiotic recombination (Stacey et al. 2006). Models of Spo11-1 activity inferred by comparison to the crystal structure of *M. jannaschii* Top6A (Nichols et al. 1999) and from molecular genetic analysis of *Saccharomyces* *Spo11-1* (Neale et al. 2005; Sasanuma et al. 2007) indicate that 2 Spo11-1 molecules interact together during the creation of meiosis-specific DSBs. Based on these data, we propose that organisms having both *Spo11-1* and *Spo11-2* genes evolved meiotic heterodimeric activity of Spo11-1 and Spo11-2 proteins during meiosis by gene duplication and divergence early during the evolution of meiosis, though the proteins may act as both homodimers and heterodimers. If either of these hypotheses is true, we argue that the heterodimeric activity would have evolved after the origin of meiosis and the initial evolution of meiotic Spo11 homologs. An interpretation of the phylogenetic pattern (figs. 2B and 3 and table 1) of protist Spo11 homologs is that the meiosis-specific Spo11 homologs act together in complexes, either as 1) a Spo11-1 homodimer in animals, fungi, *Trichomonas*, and *Giardia*, which we infer to be the

**Table 1**  
**Distribution of eukaryotic *Spo11* (*Top6A*) and *Bin3* (*Top6B*) homologs across representatives of 5 of the 6 eukaryotic supergroups**

Archaea Eukaryotes						<i>Top6A</i>			<i>Top6B</i>		
Supergroup	Subgroup	Subgroup	Subgroup	Subgroup	Representative Genera	<i>Spo11-1</i>	<i>Spo11-2</i>	<i>Spo11-3</i> ( <i>Bin5/Rhl2</i> )	( <i>Bin3</i> )		
Opisthokonta		Metazoa	Cnidaria	Anthozoa	<i>Nematostella</i>	+	–	–	–		
			Bilateria		<i>Homo, Apis, Caenorhabditis</i> , etc.	+	–	–	–		
		Choanomonada			<i>Monosiga</i>	+	–	+	+		
	Fungi		Basidiomycota		<i>Cryptococcus, Coprinus</i>	+	–	–	–		
			Ascomycota		<i>Candida, Neurospora</i> , etc.	+	–	–	–		
Microsporidia				<i>Encephalitozoon</i>	+	–	–	–			
“Amoebozoa”		Entamoebida		<b><i>Entamoeba</i></b>	++	–	–	–			
		Acanthamoebidae		<b><i>Acanthamoeba</i></b> (GSS)	+	+	–	–			
		Eumycetozoa		<i>Dictyostelium discoideum</i> strain AX4	–	–	–	–			
“Excavata”	Metamonada		Eopharyngia	Diplomonadida	<i>Giardia</i>	+	–	–	–		
			Trichomonadida		<b><i>Trichomonas</i></b>	+	–	–	–		
	Discicristata		Euglenozoa	Trypanosomatida	<b><i>Trypanosoma, Leishmania</i></b>	+	+	–	–		
			Heterolobosea		<i>Naegleria</i>	+	+	–	–		
			Malawimonadidae		<b><i>Malawimonas</i></b>	–	–	–	–		
“Chromalveolata”	“Chromista”	Stramenopiles	Bacillariophyta	Mediophyceae	<i>Thalassiosira</i>	–	+	+	+		
				Bacillariophyceae	<i>Phaeodactylum</i>	–	+	+	+		
				Oomycetes	Peronosporales	<i>Phytophthora</i>	–	+	–	–	
		Haptophyta		Pymnesiophyceae	Isochrysidales	<i>Emiliana</i> (EST)	–	–	+	–	
					<i>Isochrysis</i> (EST)	–	–	–	+		
		Alveolata		Ciliophora	Oligohymenophorea	<i>Tetrahymena, Paramecium</i>	+	–	–	–	
					Apicomplexa	Aconoidasida	<i>Plasmodium, Theileria</i>	+	+	–	+
“Archaeplastida”		Chloroplastida		Conoidasida	<b><i>Cryptosporidium</i></b>	+	–	–	–		
				Plantae	<i>Arabidopsis, Physcomitrella</i> , etc.	+	+	+	+		
					<i>Ostreococcus</i>	–	+	+	+		
		Chlorophyta				<i>Micromonas</i> (EST)	–	–	–	+	
						<i>Chlamydomonas</i>	–	–	–	+	
						Zygnemophyceae	<i>Closterium</i> (EST)	–	+	–	–
						Trebouxiophyceae	<i>Prototheca</i> (EST)	–	–	–	+
Rhodophyceae				<i>Cyanidioschyzon</i>	–	+	+	+			
				<i>Galdieria</i>	+	+	–	–			

NOTE.—Gene names are assigned according to the phylogenetic analyses shown in figures 2–5 and supplementary figures S2–S7, Supplementary Material online. The *Spo11-1* and *Spo11-2* columns represent meiosis-specific homologs. Presence and absence of genes from completed genome projects are shown (+, –), missing data remain blank. Although the formal supergroup assignment of assemblages of eukaryotes (Adl et al. 2005) remains controversial (Parfrey et al. 2006), the controversy does not change our conclusions. Rather, we use these names to emphasize the breadth of diverse eukaryotic lineages sampled in our study. Genera from which *Spo11* homologs were identified by PCR in this study are given in bold, and ESTs and GSS are indicated parenthetically. Apicomplexan *Bin3* (*Top6B*) homologs are only conserved relative to other organisms at the N-terminal domain.

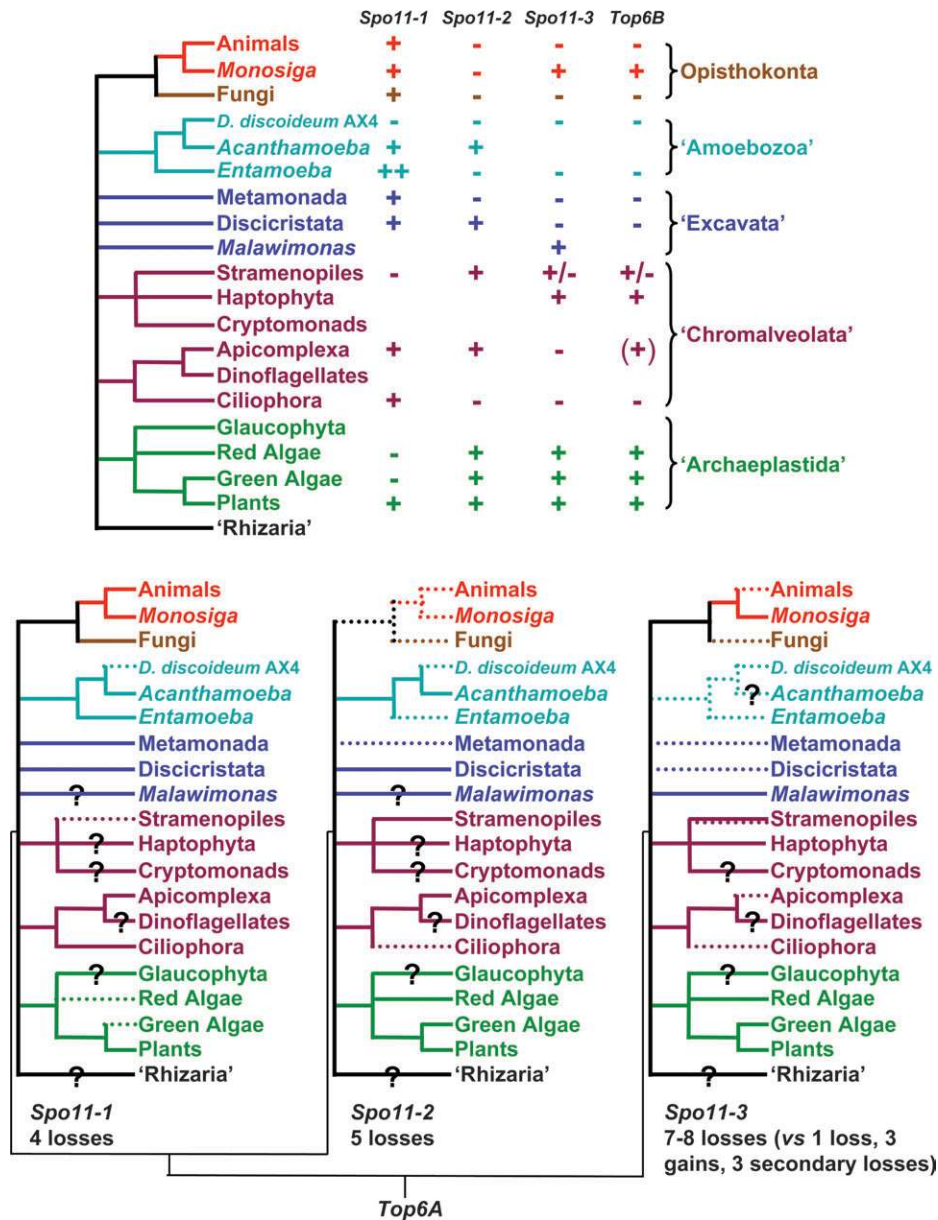


FIG. 6.—Hypotheses for the evolution of *Spo11* and *Top6B* homologs in eukaryotes. The upper panel integrates our findings of the presence of homologs with a consensus summary of the phylogeny of eukaryotes (Simpson and Roger 2004; Adl et al. 2005; Parfrey et al. 2006; Simpson et al. 2006). The lower panel illustrates various hypotheses for the pattern of gene duplications early during eukaryotic evolution and gene losses that explain the observed presence (solid lines) and absence (dotted lines) of genes. Absences were only noted for those organisms with completely sequenced genomes (i.e., 8× coverage); question marks (?) are indicated for organisms without complete genome sequence data, in which the gene has not been found. We propose that *Bin3* (*Top6B*) exhibits a pattern of gene loss similar to *Spo11-3* (*Top6A*), except for the presence of partly homologous *Bin3* (*Top6B*) genes in Apicomplexa indicated in parentheses (see text).

ancestral state, or as 2) a *Spo11-1* heterodimer in *E. histolytica* that has 2 *Spo11-1* homologous genes, 3) a heterodimer of *Spo11-1* and *Spo11-2* in plants, kinetoplastids, *Naegleria*, and *Acanthamoeba* (as is likely the case in *Arabidopsis* [Stacey et al. 2006]), or 4) a *Spo11-2* homodimer in green and red algae and the stramenopiles *Thalassiosira*, *Phaeodactylum*, and *Phytophthora* organisms that apparently lack *Spo11-1*.

The results of our analyses do not support an earlier proposal made on the basis of a few shared intron positions that *Spo11-3* evolved by a retrotransposition of *Spo11-2*

(Hartung et al. 2002). The pattern of broad phylogenetic distribution of each paralog across diverse eukaryotic groups (fig. 2), the patterns of conserved amino acids among and between paralogs (supplementary fig. S2, Supplementary Material online), and the relationship of *Spo11-1* and *Spo11-2* as sisters (fig. 2B) are together evidence that is inconsistent with the origin of *Spo11-3* by retrotransposition of *Spo11-2*. Instead, these data together favor the explanation that *Spo11-1*, *Spo11-2*, and *Spo11-3* evolved by early eukaryotic gene duplications followed by lineage-specific losses of paralogs in some organisms (fig. 6).

## Origin of Eukaryotic *Spo11/Top6A* and *Top6B* Homologs—Their Relationships to Prokaryotes

Our analyses indicate that *Top6A* and *Top6B* homologs were present in the common ancestor of archaeobacteria and eukaryotes. Prokaryotic orthologs of *Top6A* and *Top6B* are distributed widely in archaeobacteria but are generally not found in eubacteria, with the exception of a few eubacterial homologs we identified (figs. 2C and 4). These few eubacterial sequences nestled within groups of archaeobacterial *Top6A* or *Top6B* sequences, indicating that they were laterally transferred from archaeobacteria at least twice to eubacteria: once to planctomycetes and once to delta-proteobacteria. Our analyses indicate that eukaryotic *Spo11* (*Top6A*) and *Bin3* (*Top6B*) homologs most likely evolved in the last common ancestor of eukaryotes and archaeobacteria (figs. 2 and 4), contrary to suggestions made in the absence of diverse protist data that either *Top6A* and *Top6B* were transferred laterally from an archaeobacterium to plants or other eukaryotes (Hartung and Puchta 2001; Corbett and Berger 2003a; Jain et al. 2006; Stacey et al. 2006) or *Spo11-2*, *Spo11-3*, and archaeobacterial *Top6A* are more closely related to each other than to eukaryotic meiosis-specific *Spo11-1* orthologs (Yin et al. 2002; Jain et al. 2006; Stacey et al. 2006).

Furthermore, recent studies indicate that *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) proteins have an N-terminal nuclear localization signal and are located in the nucleus of *Arabidopsis* rather than being targeted to the chloroplasts or mitochondria (Hartung and Puchta 2001; Hartung et al. 2002; Sugimoto-Shirasu et al. 2002). This also supports the inference that primary endosymbiotic gene transfer from plastids or mitochondria is not responsible for the origin and evolution of eukaryotic *Spo11* (*Top6A*) and *Bin3* (*Top6B*) homologs, a hypothesis that is further supported by the paucity of eubacterial homologs of either gene and their absence from all sequenced cyanobacterial and alpha-proteobacterial genomes. Both *Spo11* (*Top6A*) and *Bin3* (*Top6B*) genes are orthologous to their archaeobacterial homologs, rather than being related by horizontal gene transfer. If these genes were acquired by horizontal gene transfer from a specific archaeobacterial lineage, eukaryotic homologs would be specifically related to a particular group of archaeobacteria, similar to the eubacterial *Top6A* and *Top6B* homologs that are found nested within archaeobacterial homologs in figures 2C and 4. Our data are inconsistent with this scenario and instead are completely consistent with the presence of *Top6A* and *Top6B* in a common ancestor of eukaryotes and archaeobacteria.

## Loss of Eukaryotic *Spo11/Top6A* and *Top6B* Homologs

Orthologs of *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) occur primarily in eukaryotes whose common ancestors hypothetically had plastids derived by either primary or secondary endosymbiosis: 1) in the “Archaeplastida” (Adl et al. 2005), represented in our data by plants, and green and red algae, and 2) in the “Chromista” (Cavalier-Smith 2002; Cavalier-Smith 2003b), represented in our data by the stramenopiles *Thalassiosira*, *Phaeodactylum*, and

*Phytophthora* and the haptophytes *Isochrysis* and *Emiliania*. These data are shown in figures 2C, 4, and 5 and summarized in table 1 and figure 6. However, *Malawimonas*, an excavate not known to be a close relative of “Archaeplastida” or “Chromista,” also expresses an mRNA for *Spo11-3* (*Top6A*), shown in figure 2C. We confirmed that this single EST is indeed from *Malawimonas* by isolating it from genomic DNA by PCR. Another curious result is that apicomplexans apparently lack *Spo11-3* (*Top6A*) (fig. 2) and yet their *Bin3* (*Top6B*) homolog is only conserved in the N-terminal domain (fig. 4A) and has sequence that is apparently nonhomologous to *Bin3* (*Top6B*) in the rest of the protein. These data suggest that apicomplexans have lost *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) may have evolved in this lineage to function independently of *Top6A*. The draft genome sequence of *Monosiga*, a protist relative of animals, is curiously unlike animals in that it reveals the apparent presence of *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) (figs. 2, 4, and 5 and table 1). We can only speculate that lateral gene transfer from a green alga could explain the presence of topoisomerase VI subunits in *Monosiga*, and we lack additional data to support this speculation.

The more broad distribution of *Spo11-1* and *Spo11-2* in eukaryotes compared with *Spo11-3* (*Top6A*) supports the idea that the common ancestor of *Spo11-1* and *Spo11-2* evolved from a *Spo11-3* (*Top6A*) ancestor by gene duplication early during eukaryotic evolution, prior to the divergence of opisthokonts, “Amoebozoa,” “Archaeplastida,” “Chromalveolates,” and “Excavates” (fig. 6). If so, both *Spo11-3* (*Top6A*) and its coevolving partner *Bin3* (*Top6B*) were subsequently lost separately multiple times: in the ancestor(s) of animals, Fungi, and Amoebozoa (together called “unikonts” [Cavalier-Smith 2002]), in the common ancestors of *Giardia* and *Trichomonas* (metamonads [Cavalier-Smith 2003a]), and in the common ancestor of the discicristates (Heterolobosea) *Naegleria* and Euglenozoa (*Trypanosoma* and *Leishmania*), as outlined in table 1 and figure 6, based on figures 2 and 4. This hypothesis of multiple gene losses following gene duplication is consistent with our observation that *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) homologs are absent from every organism we searched that had a complete genome project, except for “Archaeplastida,” some “Chromalveolates,” *Monosiga*, and *Malawimonas* (figs. 2, 4–6). An alternate hypothesis that appears less likely is that *Spo11-3* (*Top6A*) and *Bin3* (*Top6B*) may have been lost in the common ancestor of all eukaryotes except for the “Archaeplastida” and then reacquired by several separate eukaryote-to-eukaryote lateral gene transfers. If so, both genes could have been either 1) laterally transferred independently to *Malawimonas*, *Monosiga*, haptophytes, and diatoms or 2) transferred following secondary endosymbiosis directly to the common ancestor of “Chromalveolates” and then lost separately by some members of this group. The details of this multiple gene loss and transfer hypothesis depend largely on gene loss, and the number of hypothetical gene losses vary depending upon the agreed relationships among eukaryotes (Adl et al. 2005), requiring more data from additional diverse organisms for it to be resolved (Parfrey et al. 2006).

## Conclusions

We present a detailed phylogenetic study of Spo11, a protein of central importance in meiotic recombination, and find that it has a complex evolutionary history of gene duplication, followed by loss and further gene duplication in some lineages. The original ancient gene duplications preceded the evolution of major eukaryotic lineages. Together our data suggest that the evolution of land plants, the divergence of the primary photosynthetic green and red algal lineages, and the evolution of most, if not all, eukaryotic lineages were preceded by 1) the origin of eukaryotic Spo11 (*Top6A*) and *Top6B* homologs, 2) the gene duplication separating Spo11-3 (*Top6A*) from meiosis-specific homologs, and 3) the gene duplication separating Spo11-1 and Spo11-2. Furthermore, some Spo11 paralogs were subsequently lost from several lineages after these duplications occurred. Spo11-3 (*Top6A*) and Bin3 (*Top6B*) orthologs were lost in the common ancestors of “unikonts,” metamonads, heteroloboseans, and euglenozoans. Spo11-1 orthologs are absent from the recently completed genome sequence projects of green algae, red algae, and stramenopiles and thus are inferred to be lost. Spo11-2 orthologs were lost in animals, fungi, *Giardia*, *Trichomonas*, *Cryptosporidium*, ciliates, and *Entamoeba*. The ancient nature of the gene duplications and the broad taxonomic representation among eukaryotes shown here both illustrate the unappreciated taxonomic diversity in the presence of these genes and suggest that Spo11-3 (*Top6A*) and Bin3 (*Top6B*) orthologs could be present in other opisthokonts, including animals, fungi, and their protist relatives. The prevalence of Spo11-2, Spo11-3 (*Top6A*), and Bin3 (*Top6B*) homologs in plants and some protists but not animals or fungi indicates that the as-yet unexplored functions of protist Spo11-3 (*Top6A/Bin5*) and Bin3 (*Top6B*) in endoreduplication, cell growth, and proliferation, and of Spo11-1 and Spo11-2 in meiosis may be plant like. Although experiments in several organisms demonstrate the meiosis-specific role of Spo11-1 orthologs, currently there is only experimental evidence from plants demonstrating that Spo11-2 orthologs are meiosis specific. This emphasizes the need for genetic and cytological studies to elucidate the role of Spo11-2 in meiosis in other diverse eukaryotes in which they are also found. These less-studied genes may have broad implications in the developmental biology of diverse protists as well as plants and should also be examined.

## Supplementary Materials

Supplementary tables S1–S6 and figures S1–S9 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

The authors would like to thank two anonymous reviewers, Robert E. Malone, Debashish Bhattacharya, Arthur W. Pightling, Demelza Koehn, Bryant F. McAllister, and Andrew M. Schurko for helpful comments, Elizabeth Reynolds for assistance with database searches, and

Steven B. Thomas, Lauren M. Stefaniak, and Abram Doval for sequencing. This work was supported by a grant from the National Science Foundation (MCB-0437442-0216702) to J.M.L. and by both startup funds from Roanoke College and a Mednick grant from the Virginia Foundation for Independent Colleges to M.A.R. S.-B.M. was supported partly by a University of Iowa Avis Cone Graduate Fellowship.

## Literature Cited

- Adl SM, Simpson AG, Farmer MA, et al. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* 52:399–451.
- Agrawal AF. 2006. Evolution of sex: why do organisms shuffle their genotypes? *Curr Biol.* 16:R696–R704.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped Blast and PSI-Blast: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Archibald JM, O’Kelly CJ, Doolittle WF. 2002. The chaperonin genes of jakobid and jakobid-like flagellates: implications for eukaryotic evolution. *Mol Biol Evol.* 19:422–431.
- Arora C, Kee K, Maleki S, Keeney S. 2004. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Mol Cell.* 13:549–559.
- Atcheson CL, DiDomenico B, Frackman S, Esposito RE, Elder RT. 1987. Isolation, DNA sequence, and regulation of a meiosis-specific eukaryotic recombination gene. *Proc Natl Acad Sci USA.* 84:8035–8039.
- Baudat F, Keeney S. 2001. Meiotic recombination: making and breaking go hand in hand. *Curr Biol.* 11:R45–R48.
- Baudat F, Manova K, Yuen JP, Jasin M, Keeney S. 2000. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol Cell.* 6:989–998.
- Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature.* 386:414–417.
- Bowring FJ, Yeadon PJ, Stainer RG, Catcheside DE. 2006. Chromosome pairing and meiotic recombination in *Neurospora crassa* spo11 mutants. *Curr Genet.* 50:115–123.
- Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol.* 52:297–354.
- Cavalier-Smith T. 2003a. The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, *Carpodiemonas*, Eopharyngia) and Loukozoa emend. (Jakobea, *Malawimonas*): their evolutionary affinities and new higher taxa. *Int J Syst Evol Microbiol.* 53:1741–1758.
- Cavalier-Smith T. 2003b. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philos Trans R Soc Lond B Biol Sci.* 358:109–133 discussion 133–104.
- Celerin M, Merino ST, Stone JE, Menzie AM, Zolan ME. 2000. Multiple roles of Spo11 in meiotic chromosome behavior. *EMBO J.* 19:2739–2750.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31:3497–3500.
- Corbett KD, Benedetti P, Berger JM. 2007. Holoenzyme assembly and ATP-mediated conformational dynamics of topoisomerase VI. *Nat Struct Mol Biol.* 14:611–619.

- Corbett KD, Berger JM. 2003a. Emerging roles for plant topoisomerase VI. *Chem Biol.* 10:107–111.
- Corbett KD, Berger JM. 2003b. Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *EMBO J.* 22:151–163.
- Corbett KD, Berger JM. 2005. Structural dissection of ATP turnover in the prototypical GHF ATPase TopoVI. *Structure.* 13:873–882.
- Dacks JB, Doolittle WF. 2001. Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell.* 107:419–425.
- Dernburg AF, McDonald K, Moulder G, Barstead R, Dresser M, Villeneuve AM. 1998. Meiotic recombination in *C. elegans* initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell.* 94:387–398.
- Diaz RL, Alcid AD, Berger JM, Keeney S. 2002. Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. *Mol Cell Biol.* 22:1106–1115.
- Eichinger L, Pachebat JA, Glockner G, et al. (97 co-authors). 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature.* 435:43–57.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool.* 27:401–410.
- Felsenstein J. 2005. PHYLIP, phylogeny inference package. Seattle: University of Washington.
- Gadelle D, Filee J, Buhler C, Forterre P. 2003. Phylogenomics of type II DNA topoisomerases. *Bioessays.* 25:232–242.
- Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst Biol.* 49:652–670.
- Grelon M, Vezon D, Gendrot G, Pelletier G. 2001. AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J.* 20:589–600.
- Hartung F, Angelis KJ, Meister A, Schubert I, Melzer M, Puchta H. 2002. An archaeobacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr Biol.* 12:1787–1791.
- Hartung F, Blattner FR, Puchta H. 2002. Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucleic Acids Res.* 30:5175–5181.
- Hartung F, Puchta H. 2000. Molecular characterization of two paralogous SPO11 homologues in *Arabidopsis thaliana*. *Nucleic Acids Res.* 28:1548–1554.
- Hartung F, Puchta H. 2001. Molecular characterization of homologues of both subunits A (SPO11) and B of the archaeobacterial topoisomerase 6 in plants. *Gene.* 271:81–86.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.
- Jain M, Tyagi AK, Khurana JP. 2006. Overexpression of putative topoisomerase 6 genes from rice confers stress tolerance in transgenic *Arabidopsis* plants. *FEBS J.* 273:5245–5260.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 8:275–282.
- Keeney S. 2001. Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol.* 52:1–53.
- Keeney S, Baudat F, Angeles M, Zhou ZH, Copeland NG, Jenkins NA, Manova K, Jasin M. 1999. A mouse homolog of the *Saccharomyces cerevisiae* meiotic recombination DNA transesterase Spo11p. *Genomics.* 61:170–182.
- Keeney S, Giroux CN, Kleckner N. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell.* 88:375–384.
- Keeney S, Neale MJ. 2006. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem Soc Trans.* 34:523–525.
- Kessin RH. 2006. The secret lives of *Dictyostelium*. In: Eichinger L, Rivero F, editors. *Dictyostelium discoideum* protocols. Methods in molecular biology. Vol. 346. Totowa (NJ): Humana Press. p. 1–13.
- King J, Insall R. 2006. Parasexual genetics using axenic cells. In: Eichinger L, Rivero F, editors. *Dictyostelium discoideum* protocols. Methods in molecular biology. Vol. 346. Totowa (NJ): Humana Press. p. 125–135.
- Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol.* 29:170–179.
- Lichten M. 2001. Meiotic recombination: breaking the genome to save it. *Curr Biol.* 11:R253–R256.
- Lin Y, Smith GR. 1994. Transient, meiosis-induced expression of the *rec6* and *rec12* genes of *Schizosaccharomyces pombe*. *Genetics.* 136:769–779.
- Macinnes MA, Francis D. 1974. Meiosis in *Dictyostelium mucoroides*. *Nature.* 251:321–324.
- Maddison WP, Maddison DR. 2006. MacClade. Sunderland (MA): Sinauer Associates.
- Maleki S, Neale MJ, Arora C, Henderson KA, Keeney S. 2007. Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in *Saccharomyces cerevisiae*. *Chromosoma.* doi 10.1007/s00412-007-0111-y
- Malone RE, Bullard S, Hermiston M, Rieger R, Cool M, Galbraith A. 1991. Isolation of mutants defective in early steps of meiotic recombination in the yeast *Saccharomyces cerevisiae*. *Genetics.* 128:79–88.
- McKim KS, Green-Marroquin BL, Sekelsky JJ, Chin G, Steinberg C, Khodosh R, Hawley RS. 1998. Meiotic synapsis in the absence of recombination. *Science.* 279:876–878.
- McKim KS, Hayashi-Hagihara A. 1998. mei-W68 in *Drosophila melanogaster* encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev.* 12:2932–2942.
- Merino ST, Cummings WJ, Acharya SN, Zolan ME. 2000. Replication-dependent early meiotic requirement for Spo11 and Rad50. *Proc Natl Acad Sci USA.* 97:10477–10482.
- Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature.* 436:1053–1057.
- Nichols MD, DeAngelis K, Keck JL, Berger JM. 1999. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.* 18:6177–6188.
- Otto SP, Gerstein AC. 2006. Why have sex? The population genetics of sex and recombination. *Biochem Soc Trans.* 34:519–522.
- Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, Patterson DJ, Katz LA. 2006. Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet.* 2:e220.
- Ramesh MA, Malik SB, Logsdon JM Jr. 2005. A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol.* 15:185–191.
- Romanienko PJ, Camerini-Otero RD. 1999. Cloning, characterization, and localization of mouse and human SPO11. *Genomics.* 61:156–169.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19:1572–1574.
- Sandhu GS, Precup JW, Kline BC. 1989. Rapid one-step characterization of recombinant vectors by direct analysis of transformed *Escherichia coli* colonies. *Biotechniques.* 7:689–690.

- Sasanuma H, Murakami H, Fukuda T, Shibata T, Nicolas A, Ohta K. 2007. Meiotic association between Spo11 regulated by Rec102, Rec104 and Rec114. *Nucleic Acids Res.* 35:1119–1133.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics.* 18:502–504.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 51:492–508.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics.* 17:1246–1247.
- Simpson AG, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of “primitive” eukaryotes. *Mol Biol Evol.* 23:615–625.
- Simpson AGB, Roger AJ. 2004. The real ‘kingdoms’ of eukaryotes. *Curr Biol.* 14:R693–R696.
- Stacey NJ, Kuromori T, Azumi Y, Roberts G, Breuer C, Wada T, Maxwell A, Roberts K, Sugimoto-Shirasu K. 2006. *Arabidopsis* SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J.* 48:206–216.
- Storlazzi A, Tesse S, Gargano S, James F, Kleckner N, Zickler D. 2003. Meiotic double-strand breaks at the interface of chromosome movement, chromosome remodeling, and reductive division. *Genes Dev.* 17:2675–2687.
- Sugimoto-Shirasu K, Stacey NJ, Corsar J, Roberts K, McCann MC. 2002. DNA topoisomerase VI is essential for endoreduplication in *Arabidopsis*. *Curr Biol.* 12:1782–1786.
- Villeneuve AM, Hillers KJ. 2001. Whence meiosis? *Cell.* 106:647–650.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 18:691–699.
- Yin Y, Cheong H, Friedrichsen D, Zhao Y, Hu J, Mora-Garcia S, Chory J. 2002. A crucial role for the putative *Arabidopsis* topoisomerase VI in plant growth and development. *Proc Natl Acad Sci USA.* 99:10191–10196.

Andrew Roger, Associate Editor

Accepted September 27, 2007