Journal of Systematics and Evolution 46 (3): 405–423 (2008)
(formerly Acta Phytotaxonomica Sinica)
doi: 10.3724/SP.J.1002.2008.08031
http://www.plantsystematics.com

Long-term maintenance of stable copy number in the eukaryotic *SMC* family: origin of a vertebrate meiotic *SMC1* and fate of recent segmental duplicates

[1,4]Alexandra SURCEL [1,2]Xiaofan ZHOU* [2,3]Li QUAN* [1,2]Hong MA**

[1](*The Intercollege Graduate Program in Cell and Developmental Biology, The Huck Institutes of the Life Sciences, The Pennsylvania State University,* University Park, PA 16802, USA)

[2](*Department of Biology, The Pennsylvania State University,* University Park, PA 16802, USA)

[3](*The Huck Institutes of the Life Sciences, The Pennsylvania State University,* University Park, PA 16802, USA)

[4](Present address: *Johns Hopkins University School of Medicine,* Baltimore, MD 21205, USA)

Abstract Members of the *S*tructural *M*aintenance of *C*hromosome (SMC) family have long been of interest to molecular and evolutionary biologists for their role in chromosome structural dynamics, particularly sister chromatid cohesion, condensation, and DNA repair. SMC and related proteins are found in all major groups of living organisms and share a common structure of conserved N and C globular domains separated from the conserved hinge domain by long coiled-coil regions. In eukaryotes there are six paralogous proteins that form three heterodimeric pairs, whereas in prokaryotes there is only one SMC protein that homodimerizes. From recently completed genome sequences, we have identified *SMC* genes from 34 eukaryotes that have not been described in previous reports. Our phylogenetic analysis of these and previously identified *SMC* genes supports an origin for the vertebrate meiotic *SMC1* in the most recent common ancestor since the divergence from invertebrate animals. Additionally, we have identified duplicate copies due to segmental duplications for some of the *SMC* paralogs in plants and yeast, mainly *SMC2* and *SMC6*, and detected evidence that duplicates of other paralogs were lost, suggesting differential evolution for these genes. Our analysis indicates that the *SMC* paralogs have been stably maintained at very low copy numbers, even after segmental (genome-wide) duplications. It is possible that such low copy numbers might be selected during eukaryotic evolution, although other possibilities are not ruled out.

Key words cohesin, condensin, meiosis, segmental duplication, *SMC*.

During eukaryotic cell division, chromosomes must be distributed correctly into daughter cells. Improper chromosome segregation results in cell death or aneuploidy, which is the cause of such disorders as Down's Syndrome, Cornelia de Lange syndrome, and tumorigenesis (Pati et al., 2002; Gilliland & Hawley, 2005; Ren et al., 2005; Musio et al., 2006; Deardorff et al., 2007; Ohbayashi et al., 2007). To ensure that each daughter cell receives a complete set of chromosomes, two chromosomal processes are crucial: sister chromatid cohesion and chromosome condensation.

*S*ister *C*hromatid *C*ohesion (SCC) refers to the close association of replicated sister chromatids along the entire length of the chromosome (for a review of cohesion and condensation, see Nasmyth & Haering, 2005). SCC is in place as sister chromatids are formed during the S phase of the cell cycle and it is maintained while chromosomes condense and shorten along their axes, until the anaphase-metaphase transition in mitosis. SCC provides the counter-force to amphitelic attachments of microtubules originating from the two poles of the spindle to the kinetochores. At the completion of amphitelic attachment, SCC is resolved to allow for the separation of sister chromatids, which are pulled by the spindle to the poles. The reduction of chromosome length during condensation is in part due to the formation of solenoidal chromatin loops that form rosettes along a central axis (Paulson & Laemmli, 1977; Laemmli, 1978; Marsden & Laemmli, 1979; Maeshima & Laemmli, 2003). Condensation is a prerequisite for mitosis progression and is inter-dependent on SCC (Nasmyth, 2005; Nasmyth & Haering, 2005).

Despite the early identification of the ubiquitous nature of both condensation and cohesion among eukaryotes, the key players involved in these processes have only been identified relatively recently from biochemical and genetic studies (for review see

Received: 5 March 2008 Accepted: 22 April 2008

* These authors contributed equally to this work.

** Author for correspondence. E-mail: hxm16@psu.edu; Tel.: (814) 863-6414.

Abbreviations: SMC, structural maintenance of chromosome; NJ, neighbor-joining; BLAST, basic local alignment search tool; ML, maximum likelihood.

Hirano, 2002). Both cohesion and condensation are mediated by multimeric, protein complexes known respectively as cohesin and condensin. These complexes are responsible for modulating chromatin dynamics; furthermore, they share an interesting characteristic: they both contain proteins belonging to the *S*tructural *M*aintenance of *C*hromosome (SMC) family.

Members of this evolutionarily conserved ATP-binding protein family are involved in maintaining chromosome integrity and in DNA metabolism (Hirano, 2002). In eukaryotes, there are six paralogous proteins that form three distinctive heterodimers, as part of protein complexes mainly responsible for essential chromatin maintenance—SMC1 and SMC3 are part of cohesin, SMC2 and SMC4 are part of condensin, and SMC5 and SMC6 are part of a DNA repair complex (Hirano, 2002). In archaea and bacteria, there is a single SMC protein that homodimerizes and participates in chromosome dynamics, similar to the eukaryotic cohesin and condensin complexes (Volkov et al., 2003). Some bacteria, such as *Escherichia coli,* contain MukB, a protein that has relatively low levels of sequence similarity, yet a high degree of secondary structural similarity, to that of SMC proteins (Niki et al., 1992; Melby et al., 1998; Cobbe & Heck, 2004). Members of the most abundant group of archaea, Crenarchaeota, which live in extreme hot and cold environments, have neither SMC nor MukB orthologs, but they do contain orthologs of the Rad50 protein, a more distant eukaryotic relative of SMC proteins (Soppa, 2001).

In addition to being involved in genome stability, members of the SMC family of proteins and the related MukB protein all share a characteristic structure consisting of five well-defined domains (Melby et al., 1998). They have conserved globular N- and C-terminal domains that are attached, respectively, to two long coiled-coil regions, which are separated by a flexible hinge domain (Fig. 1: A). The N terminal domains contain a Walker A motif (consensus G-X-S/T-G-X-G-K-S/T-S/T) characteristic of ATP-binding proteins, but surprisingly do not contain a complementary Walker B motif (h-h-h-h-D, where h is a hydrophobic residue), which is usually found in ATP-binding proteins containing a Walker A motif. In a novel configuration, the Walker B motif is located over 1000 amino acid residues away from the Walker A motif in the C-terminally located DA box—a highly conserved 35-amino acid stretch with alanine and aspartic acid residues. The N- and C-terminal domains at either end of the long SMC protein interact intramolecularly to form a functional ATPase similar to those of the *A*TP-*B*inding *C*assette (ABC) transporter proteins (Saitoh et al., 1995). This physical interaction of the two end domains is facilitated by the formation of an intramolecular antiparallel coiled-coil domain of approximately 50 nm in length (Fig. 1: B), which is made possible by a fold in the flexible hinge domain. Furthermore, the hinge domain is responsible for the heterodimerization between components of each dimer pair (Fig. 1: C).

This conserved domain structure, coupled with conserved functions in major eukaryotic lineages for the six SMC paralogs, suggests that they may have arisen from a single ancestor. A few phylogenetic analyses have been conducted on the SMC family, with two different interpretations on the origin of the six paralogs. Using maximum-likelihood analysis, Cobbe and Heck investigated the relationship between *SMCs* from both prokaryotes and eukaryotes and included the *Rad50* and *MukB* families encoding ABC ATPase as outgroups (Cobbe & Heck, 2004). Their phylogenetic tree shows that eukaryotic *SMC* genes evolved from several ancient gene duplication events. In addition, the DNA repair genes, *SMC5* and *SMC6*, form a separate branch from *SMC1–4*, coding for condensins and cohesins. Moreover, genes for the

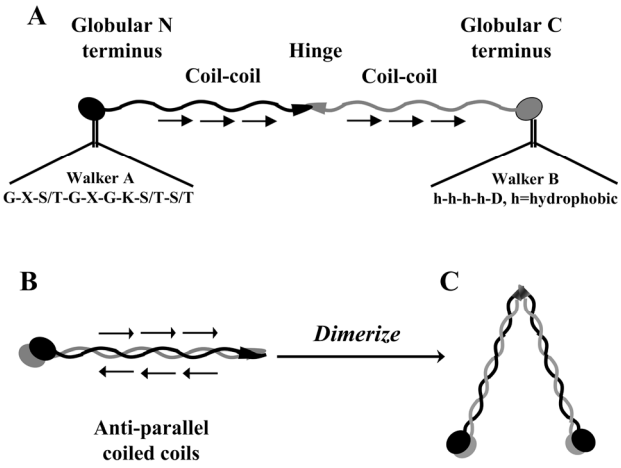


Fig. 1. General structure of proteins from the SMC family. **A,** domain structure of SMC proteins. The N and C terminal globular domains are each connected to long coiled-coil regions separated by a flexible hinge region. The N terminal domain contains the Walker A sequence that acts as a functional ATPase in conjunction with the Walker B motif, present in the C terminal domain. These two domains are brought in close proximity (**B**) when the hinge region folds. The folding of an SMC protein brings together the coiled-coil domains that are now in anti-parallel orientation to each other (shown by the black arrows). **C,** Dimerization between SMC partners occurs at the hinge domain. Interactions with other proteins of the complexes occur at the ends where the N and C termini are, also known as the head domain (interactions not shown here).

larger cohesin and condensin subunits (*SMC1* and *SMC4*) formed a clade separate from those for the smaller subunits (*SMC3* and *SMC2*). They also showed a strong correlation between the terminal domains of the same SMC, but low correlation between the N- and C-terminal domains of the paralogs that form the heterodimer, consistent with intramolecular N-C interaction. Finally, their consensus tree showed that plant and animal *SMC* genes group together with fungi as the outgroup, in contradiction to rRNA trees that show plants being outside of the animal-fungi clade.

In addition to the DNA and protein sequences, information about the secondary structures of proteins can also be used for phylogenetic analyses. Studies on the secondary structure of the long SMC arms containing coiled-coil domains using the COILS program indicate that each SMC arm contained two or more coiled-coil regions, with "breaks" in between (Beasley et al., 2002). For each pair of two paralogous proteins that form a heterodimer, the patterns of these breaks are different. In SMC1, SMC4, and SMC5 proteins, one break occurred in the coiled-coil arm between the N-terminal domain and hinge (left arm) and two occurred in the coiled-coil between the hinge and the C-terminal domain (right arm), whereas for SMC2, SMC3, and SMC6, two breaks were found in the left arm and one in the right arm. The patterns of the secondary structure of the coiled-coil arms suggest that unlike the tree generated by Cobbe and Heck (2004), the *SMC* genes could be categorized into two groups, implying that an early gene duplication event in the ancestral *SMC* gave rise to two heterodimer partners. Subsequent duplication events then produced three copies of each subunit of the heterodimer, allowing divergence to fulfill various functions.

Additionally, a study by Liu and Wang using hydropathy profiles of amino acids generated phylogenetic relationships different than those using sequence information alone (Liu & Wang, 2006). In this study, *SMC4* and *SMC1* form a clade with archael *SMC* and eubacterial *SMC*, separate from the *SMC2* and *SMC3* clade. *SMC6* still forms a clade separate from the cohesin and condensin SMCs, but its hydropathy similarities with eukaryotic Rad50 joins these two disparate groups together (Liu & Wang, 2006). The various topologies of *SMC* trees from different studies suggest that the phylogenetic relationship among *SMCs* is still uncertain.

During meiosis, a diploid cell undergoes one round of DNA replication followed by two rounds of chromosome segregation, homologous chromosomes in meiosis I and sister chromatids in meiosis II, generating four haploid cells. SCC and recombinational cross-overs maintain homolog association from late prophase I to the onset of anaphase I, when cohesin is removed along the chromosome arms to allow for the separation of homologs, but not sister chromatids (Siomos et al., 2001; Yu & Koshland, 2007). This preferential dissolution of cohesin on the chromosome arms is in part accomplished by the presence of meiosis-specific isoforms of cohesin proteins. In mouse and human, two *SMC1* isoforms exist—*SMC1α* (or *SMC1L1* for *SMC1*-like 1) and *SMC1β* (or *SMC1L2* for *SMC1*-like 2) (Revenkova et al., 2004). *SMC1β*-deficient mice are sterile and defective in cohesin maintenance, chromosome recombination, and synapsis (Revenkova et al., 2004; Revenkova & Jessberger, 2005; Hodges et al., 2005). In addition, meiosis-specific cohesin isoforms of SCC1 and SCC3 (the two proteins that form the cohesin multiprotein complex with SMC1 and SMC3), have also been identified (Eijpe et al., 2000; Hodges et al., 2005; Revenkova & Jessberger, 2005). The presence of these meiotic cohesin proteins supports the idea that the regulation of cohesin in meiosis may be different than that in mitosis. However, it was not clear whether other vertebrates and invertebrates also have two *SMC1* paralogs, and what the evolutionary relationship of the *SMC1* genes is. In addition, it was not known whether any other *SMC* paralogs also have two or more forms in some lineages. In particular, genome duplication events have been proposed for plants and the budding yeast (Simillion et al., 2002; Kellis et al., 2004); the relationship between such genome duplication and *SMC* gene family evolution has not been addressed.

The steady increase in sequenced genomes in recent years has provided a wealth of information available to address these questions. Here, we report the identification and prediction through extensive data mining of one or more additional *SMC* genes from 34 species. Our phylogenetic analysis of the *SMC* genes results in phylogenetic trees with similar topologies to those published previously (Melby et al., 1998; Cobbe & Heck, 2004). Detailed analysis of the mitotic and meiotic *SMC1* isoforms suggests that the gene duplication event responsible for the meiotic isoform occurred in early vertebrate evolution and that the two isoforms have been subjected to differential selective pressure. Additionally, we provide support to a hypothesis that *SMCs* from plant genomes show a stable copy number with the exception of very recent duplications. Our genome analysis of these plant

sequences, as well as those of budding yeast, supports the hypothesis that losses of additional copies of *SMCs* occurred after recent genome duplications.

1 Material and Methods

1.1 Phylogenetic analysis of *SMC* genes

Sequences of *SMC* genes were obtained from public databases (TAIR, NCBI, TIGR, and JGI) initially using BLAST with each SMC protein sequence from *Saccharomyces cerevisiae* as query; a BLAST score cutoff of 22% similarity was used to avoid sequences of other ABC ATPases. Additional SMC genes were identified from genomic sequences using tBLASTn with human and *Arabidopsis* sequences as queries for animal and plant genomes, respectively, and predicted manually. The protein sequences of the SMC homologs were aligned using MUSCLE version 3.6 (Edgar, 2004) with the default settings, followed by manual adjustment using GeneDoc V.2.6.002 software (http://www.nrbsc.org/gfx/genedoc/index.html, Nicholas et al., 1997). Neighbor joining trees were constructed using MEGA 4.0 (Tamura et al., 2007). The reliability of internal branches was calculated with 1000 bootstrap pseudoreplicates using the "pairwise deletion option" of amino acid sequences. Maximum likelihood analysis was performed by using PHYML 2.4.4 (Guindon & Gascuel, 2003) using the WAG model with gamma correction; bootstrap support was obtained using 100 replicates. Maximum parsimony (MP) analysis was carried out using PAUP* 4.0 beta 10 (Swofford, 2001) with default settings, and bootstrap support was determined by using 100 replicates.

1.2 Detection of segmental duplication

To find evidence for the ancient segmental duplications in the *SMC* genes in yeast, we checked the available information about the yeast whole genome duplication analysis through the supplemental material of Kellis et al. (2004) available at http://www.nature.com/nature/journal/v428/n6983/extref/nature02424-s1.htm.

Additionally, to search for recent segmental duplication evidence for the *SMC* genes in *Arabidopsis* and poplar, we collected 50kb genomic DNA sequences both upstream and downstream of all of the existing *SMC* genes from both species. Among the 8 *AtSMC*s and 7 *PtSMC*s, *AtSMC2a* & *AtSMC2b* are the pair that has been retained since the genome duplication, as well as *AtSMC6a* & *AtSMC6b* and *PtSMC2a* &*PtSMC2b*. We compared the 100kb regions of each of the gene pair using the DotPlot function of the PipMaker program (Schwartz et al., 2000) at http://pipmaker.bx.psu.edu/pipmaker/. For the rest of the *SMC* genes that lost the other duplicated copies, we used the 100kb regions of the *SMC* genes to run BLAST searches against NCBI *Arabidopsis* (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlasGen.cgi?taxid=3702) and JGI poplar (http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Poptr1_1&advanced=1) databases to look for the duplication evidence of the flanking regions near *SMC* genes. Upon the BLAST results, we picked another 100kb sequences from the subject regions that have the best hits with the query flanking sequences. Afterwards, we used the Pipmaker program to compare the two corresponding 100kb regions where we found segmental duplication evidence.

2 Results and Discussion

2.1 Identification of *SMC* genes

We performed numerous BLAST searches for homologs to each of the six SMC proteins from animals, plants and fungi, with an emphasis on those with completely sequenced genomes. In total, 273 *SMC* genes were collected from 43 species (Tables 1 and 2; complete sequences will be provided upon request). In addition to 92 known eukaryotic *SMC* sequences identified in previous phylogenetic studies

Table 1 Numbgr of *SMC* genes in this study

Organisms: number of species	*SMC1*		*SMC2*	*SMC3*	*SMC4*	*SMC5*	*SMC6*
	α	*β*					
Vertebrates: 15 (7)[1)]	14 (7)	14 (3)	14 (5)	14 (7)	14 (5)	14 (3)	14 (3)
Invertebrates: 11 (3)	11 (3)		11 (3)	11 (3)	13 (4)[2)]	11 (3)	12 (4)[2)]
Plants: 7 (2)	7 (1)		9 (3)[3)]	7 (2)	7 (2)	7 (2)	9 (3)[3)]
Fungi: 11 (4)	11 (4)		11 (3)	11 (3)	11 (3)	11 (4)	11 (3)
Total: 44 (16)	279 (87)						

1) Numbers in parentheses indicate the numbers of species or sequences which have been reported previously. 2) *Caenorhabditis elegans* has two copies of *SMC4* and *SMC6*. 3) *Arabidopsis* and poplar each have two copies of *SMC2*; *Arabidopsis* and *Selaginella moellendorffii* each have two copies of *SMC6*.

Table 2 *SMC* genes included in this study

Common name	Scientific name	*SMC1*		*SMC2*	*SMC3*	*SMC4*	*SMC5*	*SMC6*
		α	*β*					
Human	*Homo sapiens*	x	x	x	x	x	x	x
Chimpanzee	*Pan troglodytes*	x	x	x	x	x	x	x
Mouse	*Mus musculus*	x	x	x	x	x	x	x
Rat	*Rattus norvegicus*	x	x	x	x	x	x	x
Dog	*Canis familiaris*	x	x	x	x	x	x	x
Cat	*Felis catus*	x	x					
Cow	*Bos taurus*	x	x	x	x	x	x	x
Opossum	*Monodelphis domestica*	x	x	x	x	x	x	x
Pig	*Sus scrofa*			x	x	x	x	x
Macaque	*Macaca mulatta*	x	x	x	x	x	x	x
Horse	*Equus caballus*	x	x	x	x	x	x	x
Chicken	*Gallus gallus*	x	x	x	x	x	x	x
Frog	*Xenopus tropicalis*	x	x	x	x	x	x	x
Zebrafish	*Danio rerio*	x	x	x	x	x	x	x
Pufferfish	*Takafugu rubripes*	x	x	x	x	x	x	x
Yellow fever mosquito	*Aedes aegypti*	x		x	x	x	x	x
Malaria mosquito	*Anopheles gambiae*	x		x	x	x	x	x
Honey bee	*Apis mellifera*	x		x	x	x	x	x
Nematode	*Caenorhabditis elegans*	x		x	x	x[1)]	x	x[1)]
	Caenorhabditis briggsae	x		x	x	x[1)]	x	x
Fruit fly	*Drosophila melanogaster*	x		x	x	x	x	x
	Drosophila pseudoobscura	x		x	x	x	x	x
Beetle	*Tribolium castaneum*	x		x	x	x	x	x
Sea urchin	*Strongylocentrotus purpuratus*	x		x	x	x	x	x
Sea squirt	*Ciona intestinalis*	x		x	x	x	x	x
Sea anemone	*Nematostella vectensis*	x		x	x	x	x	x
Arabidopsis	*Arabidopsis thaliana*	x		x[1)]	x	x	x	x[1)]
Rice	*Oryza sativa*	x		x	x	x	x	x
Poplar	*Populus trichocarpa*	x		x[1)]	x	x	x	x
Moss	*Physcomitrella patens*	x		x	x	x	x	x
Spikemoss	*Selaginella moellendorffii*	x		x	x	x	x	x[1)]
Wine grape	*Vitis vinifera*	x		x	x	x	x	x
–	*Ostreococcus lucimarinus*	x		x	x	x	x	x
–	*Aspergillus fumigatus*	x		x	x	x	x	x
–	*Aspergillus nidulans*	x		x[2)]	x	x	x	x
–	*Aspergillus oryzae*	x		x	x[3)]	x[3)]	x	x
–	*Candida albicans*	x		x	x	x	x	x
–	*Candida glabrata*	x		x	x	x	x	x
–	*Eremothecium gossypii*	x		x	x	x	x	x
–	*Gibberella zeae*	x		x	x	x	x	x
–	*Kluyveromyces lactis*	x		x	x	x[4)]	x	x
–	*Neurospora crassa*	x		x	x	x	x	x
Budding yeast	*Saccharomyces cerevisiae*	x		x	x	x	x	x
Fission yeast	*Schizosaccharomyces pombe*	x		x	x	x	x	x

Available common names are shown next to the scientific names.

1) These species have two copies of each of these genes. 2) The *Aspergillus nidulans SMC2* gene was previously mistakenly annotated as *SMC3*. 3) The *Aspergillus oryzae SMC3* and *SMC4* genes were previously mistakenly annotated as *SMC2* and *SMC1*, respectively. 4) The *Kluyveromyces lactis SMC4* gene was previously mistakenly annotated as *SMC1*.

(Melby et al., 1998; Cobbe & Heck, 2004), we identified and/or predicted 181 new *SMC* genes from ten vertebrates, seven invertebrates, seven plants, and six fungi. In most cases, we found one copy of each of the six *SMC* paralogs; however, we recovered two copies of *SMC1* in 14 vertebrates, two copies of *SMC4* and *SMC6* in the nematode *Caenorhabditis elegans* and additional copies of *SMC2* and *SMC6* in some plants. Furthermore, we were able to detect several pseudogenes in plants, such as a *SMC2* pseudogene in grapevine and a *SMC3* pseudogene in *Populus* (not shown). These pseudogenes lack half of the coding region and have accumulated stop codons in the remaining half.

2.2 Phylogenetic analysis of the *SMC* family

Alignments of SMC sequences showed the expected highly conserved regions of the terminal globular domains and the hinge domain (the alignment will be provided upon request). The coiled-coil or arm domains are more divergent, even among members of the same clade, although *SMC1* sequences showed more conservation than the other SMC proteins within the arm domains. Using the three conserved regions, we generated Neighbor-Joining (NJ) and Maximum Likelihood (ML) trees for the *SMC* family (see Fig. 2 for an ML tree with bootstrap support from ML and NJ analyses). The results show strong support for the six clades, for the *SMC1–6* paralogs, respectively, and for the relationship among these six clades. Consistent with the previous analysis of eukaryotic genes, those for the larger subunits of cohesin and condensin, *SMC1* and *SMC4*, respectively are sister clades, as are the *SMC2* and *SMC3* genes encoding smaller subunits, while *SMC5* and *SMC6* form a separate clade. Compared with previous reports, the trees in this study have higher bootstrap values. Our results support the idea that the common ancestor of *SMC5* and *SMC6* was separated from the ancestor of the other 4 paralogs due to a duplication event in early eukaryotes, although the possibility of long-branch attraction could not be ruled out.

Our phylogenetic analyses also showed that several genes were previously annotated incorrectly (Table 2). The previously designated *SMC2* and *SMC1* from the filamentous fungus *Aspergillus oryzae* were in fact *SMC3* and *SMC4*, respectively. Likewise, the putative *SMC1* sequence for the yeast *Kluyveromyces lactis* was in fact *SMC4* and the putative *SMC3* sequence for the filamentous fungus *Aspergillus nidulans* was its *SMC2* paralog. A recent review of these BLAST searches reveals that the correct annotation for the *K. lactis* and the *A. nidulans* sequences have been assigned in the interim, while the *A. oryzae* genome has been removed from the NCBI database and will be restored upon completion of genome assembly.

2.3 Origin of *Smc1β* and the evolution of meiotic cohesin function

Previous studies of *SMC1* genes have reported a meiotic isoform for *SMC1*, called *SMC1β*, in addition to the non-specific *SMC1α*, in human, mouse, and a fish (Cobbe & Heck, 2004; Revenkova et al., 2004). It is possible that the two *SMC1* isoforms originated in the most recent common ancestor of vertebrates; alternatively the two *SMC1* isoforms were produced by a duplication event in an earlier ancestor, but the meiotic isoform has been lost in non-vertebrate organisms. To investigate the origin of the vertebrate meiosis-specific isoform *SMC1β*, we performed phylogenetic analyses on the *SMC1* sequences that we retrieved, including both the *SMC1α* and *SMC1β* isoforms from 14 vertebrates (10 mammals, chicken, frog, and two fish), and the single copy *SMC1* genes from 27 other organisms (Tables 1 and 2). Vertebrates are members of the chordates, which also include urochordates and cephalochordates; chordates and echinoderms represent two major lineages of deuterostomes, distinct from the protostomes, including insects and nematodes. To address the origin of the *SMC1* isoforms, we have identified *SMC1* sequences from the recently sequenced genomes of sea squirt (a urochordate) and sea urchin (an echinoderm), as well as from a sea anemone, which is a basal metazoan.

Because there is greater sequence conservation among *SMC1* genes than between the six *SMC* paralogs, in addition to the conserved domains (N and C termini and the hinge domain) used in the above phylogenetic analyses, the less conserved coiled-coil regions were also included in the analyses. We have performed phylogenetic analyses using ML and NJ methods with most of the sequences, as well as a third analysis using NJ with the coiled coil domains; these analyses resulted in very similar topologies (Fig. 3). These analyses indicate that (1) all animal *SMC1* sequences form a single clade with 100%/100%/99% bootstrap support; (2) all vertebrate *SMC1* genes form a monophyletic group with strong bootstrap support (100%/80%/96%); and (3) the two *SMC1* isoforms form respective clades with 100%/100%/100% support. Therefore, the *SMC1*α and *SMC1*β isoforms were likely the result of a duplication that occurred in the most recent common ancestor of vertebrates, since the divergence of vertebrates from urochordates and echinoderms.

Fig. 2A

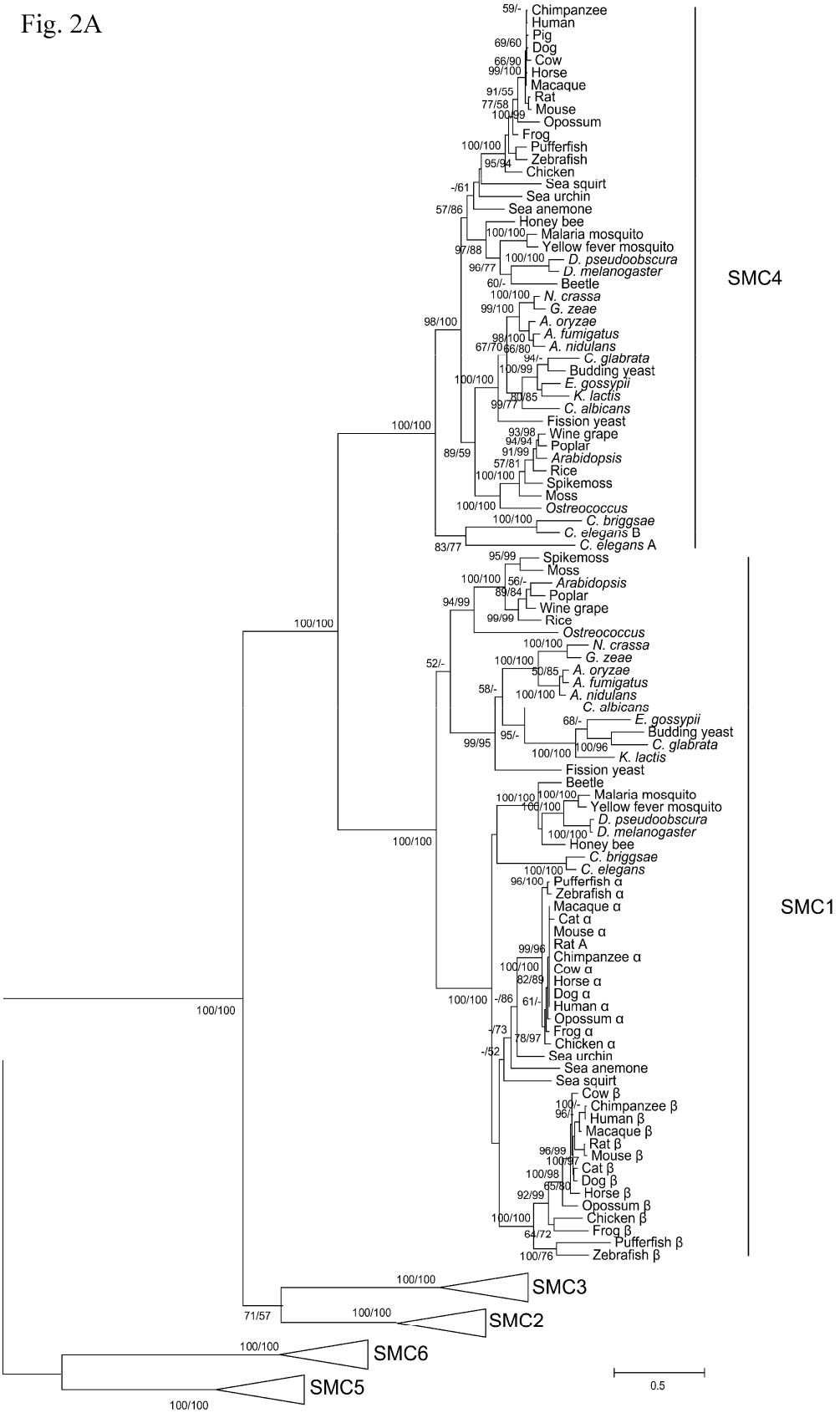
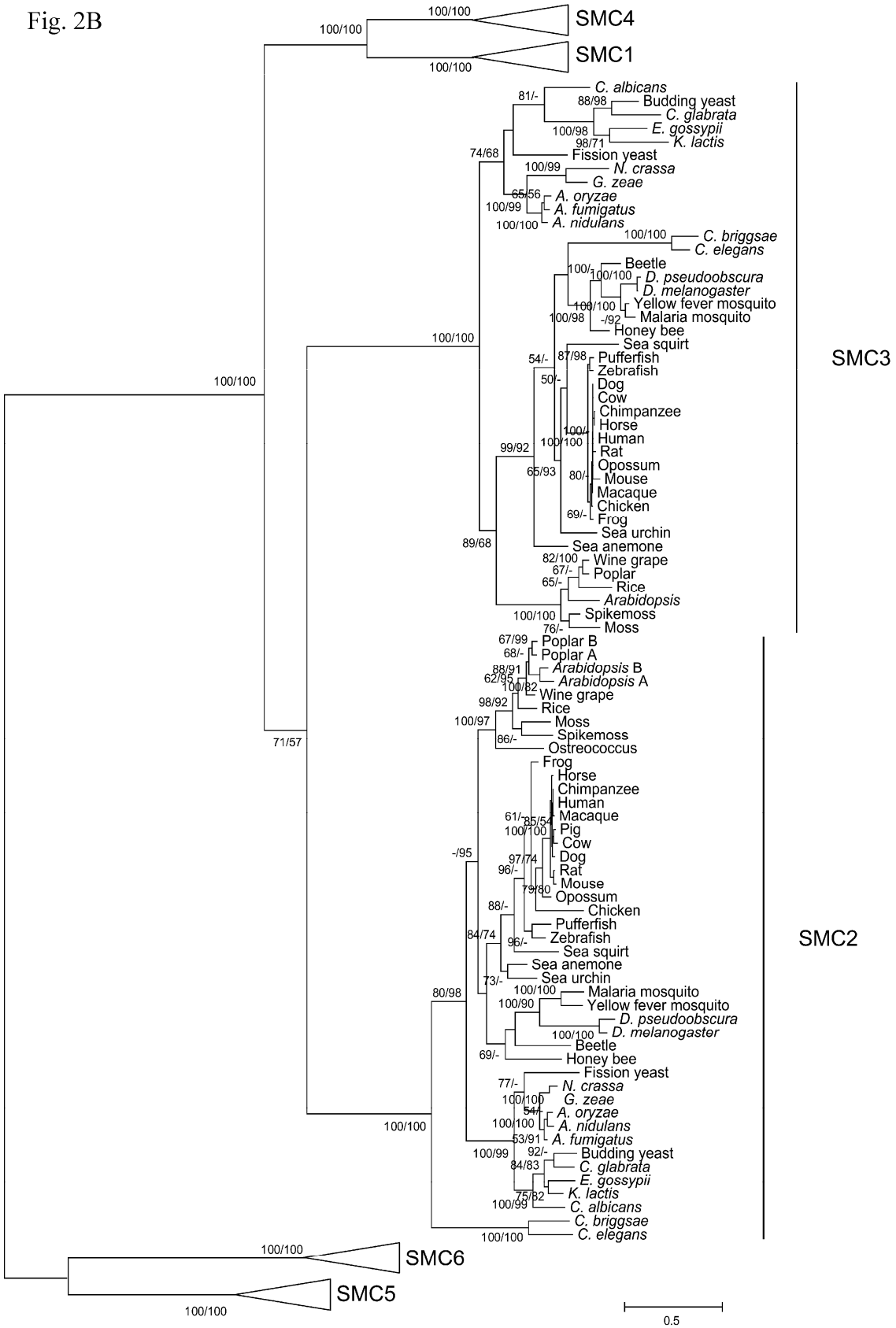

Chimpanzee
Human
Pig
Dog
Cow
Horse
Macaque
Rat
Mouse
Opossum
Frog
Pufferfish
Zebrafish
Chicken
Sea squirt
Sea urchin
Sea anemone
Honey bee
Malaria mosquito
Yellow fever mosquito
D. pseudoobscura
D. melanogaster
Beetle
N. crassa
G. zeae
A. oryzae
A. fumigatus
A. nidulans
C. glabrata
Budding yeast
E. gossypii
K. lactis
C. albicans
Fission yeast
Wine grape
Poplar
Arabidopsis
Rice
Spikemoss
Moss
Ostreococcus
C. briggsae
C. elegans B
C. elegans A
SMC4
Spikemoss
Moss
Arabidopsis
Poplar
Wine grape
Rice
Ostreococcus
N. crassa
G. zeae
A. oryzae
A. fumigatus
A. nidulans
C. albicans
E. gossypii
Budding yeast
C. glabrata
K. lactis
Fission yeast
Beetle
Malaria mosquito
Yellow fever mosquito
D. pseudoobscura
D. melanogaster
Honey bee
C. briggsae
C. elegans
Pufferfish α
Zebrafish α
Macaque α
Cat α
Mouse α
Rat A
Chimpanzee α
Cow α
Horse α
Dog α
Human α
Opossum α
Frog α
Chicken α
Sea urchin
Sea anemone
Sea squirt
Cow β
Chimpanzee β
Human β
Macaque β
Rat β
Mouse β
Cat β
Dog β
Horse β
Opossum β
Chicken β
Frog β
Pufferfish β
Zebrafish β
SMC1
SMC3
SMC2
SMC6
SMC5
0.5

Fig. 2B


SMC4
SMC1
C. albicans
Budding yeast
C. glabrata
E. gossypii
K. lactis
Fission yeast
N. crassa
G. zeae
A. oryzae
A. fumigatus
A. nidulans
C. briggsae
C. elegans
Beetle
D. pseudoobscura
D. melanogaster
Yellow fever mosquito
Malaria mosquito
Honey bee
Sea squirt
Pufferfish
Zebrafish
Dog
Cow
Chimpanzee
Horse
Human
Rat
Opossum
Mouse
Macaque
Chicken
Frog
Sea urchin
Sea anemone
Wine grape
Poplar
Rice
Arabidopsis
Spikemoss
Moss
Poplar B
Poplar A
Arabidopsis B
Arabidopsis A
Wine grape
Rice
Moss
Spikemoss
Ostreococcus
Frog
Horse
Chimpanzee
Human
Macaque
Pig
Cow
Dog
Rat
Mouse
Opossum
Chicken
Pufferfish
Zebrafish
Sea squirt
Sea anemone
Sea urchin
Malaria mosquito
Yellow fever mosquito
D. pseudoobscura
D. melanogaster
Beetle
Honey bee
Fission yeast
N. crassa
G. zeae
A. oryzae
A. nidulans
A. fumigatus
Budding yeast
C. glabrata
E. gossypii
K. lactis
C. albicans
C. briggsae
C. elegans
SMC6
SMC5
SMC3
SMC2
0.5

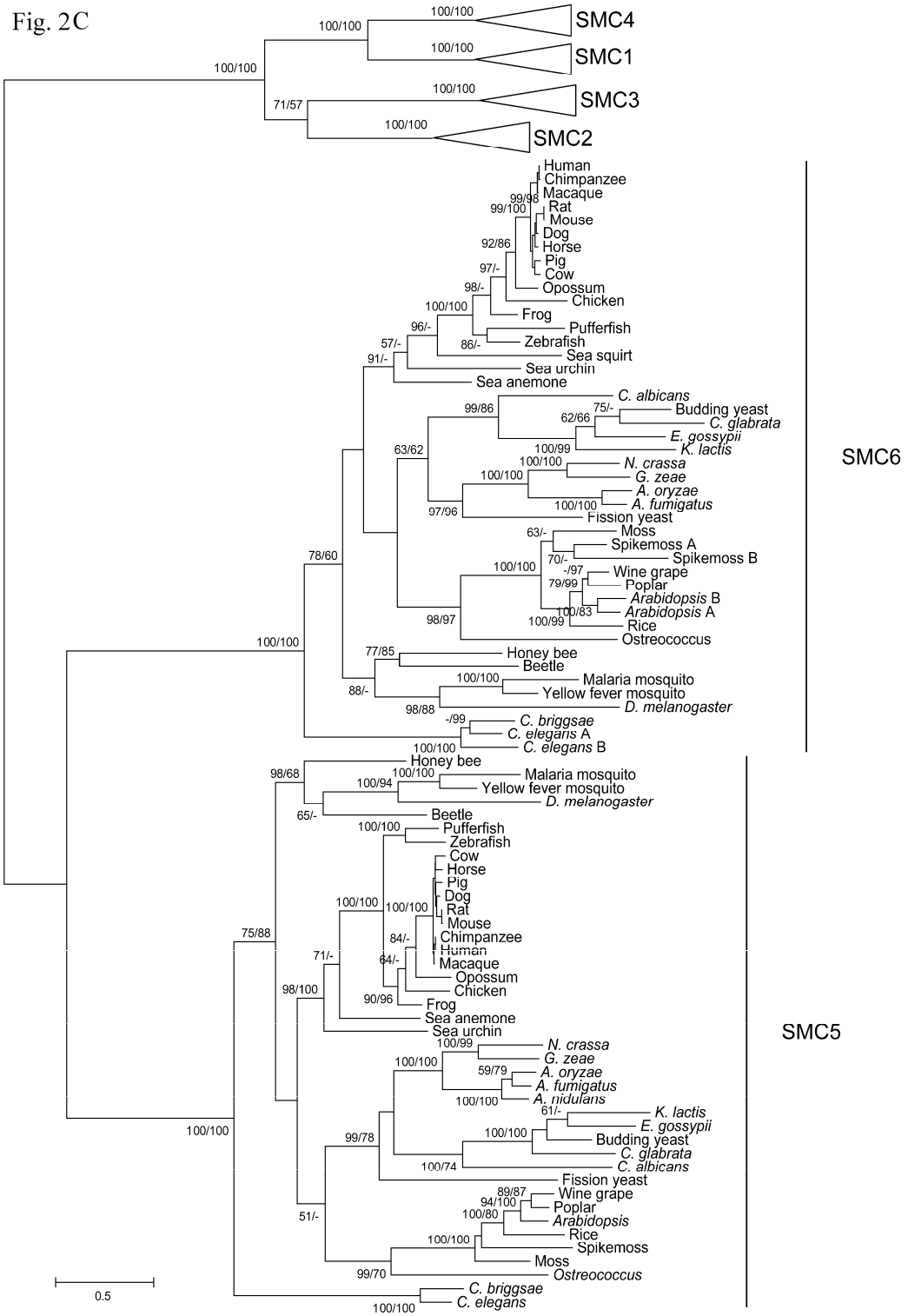


Fig. 2. A phylogenetic tree for the *SMC* family, shown in three parts. **A.** Details for clades containing the genes for the large subunits—SMC1 and SMC4—of cohesins and condensins, respectively. **B.** Details for clades containing the genes for the smaller SMC subunits of the cohesin and condensin complexes—SMC3 and SMC2, respectively. **C.** Details of the clades for SMC5 and SMC6, which are subunits of a DNA repair complex. Bootstrap supports are shown for ML/NJ analyses on a ML tree. Some of the taxa represented here are: frog, *Xenopus tropicalis*; zebrafish, *Danio rerio*; pufferfish, *Takafugu rubripes*; sea squirt, *Ciona intestinalis*; beetle, *Tribolium castaneum*; moss, *Physcomitrella patens*; spikemoss, *Selaginella moellendorffii*. For complete taxon information, please see Table 2.

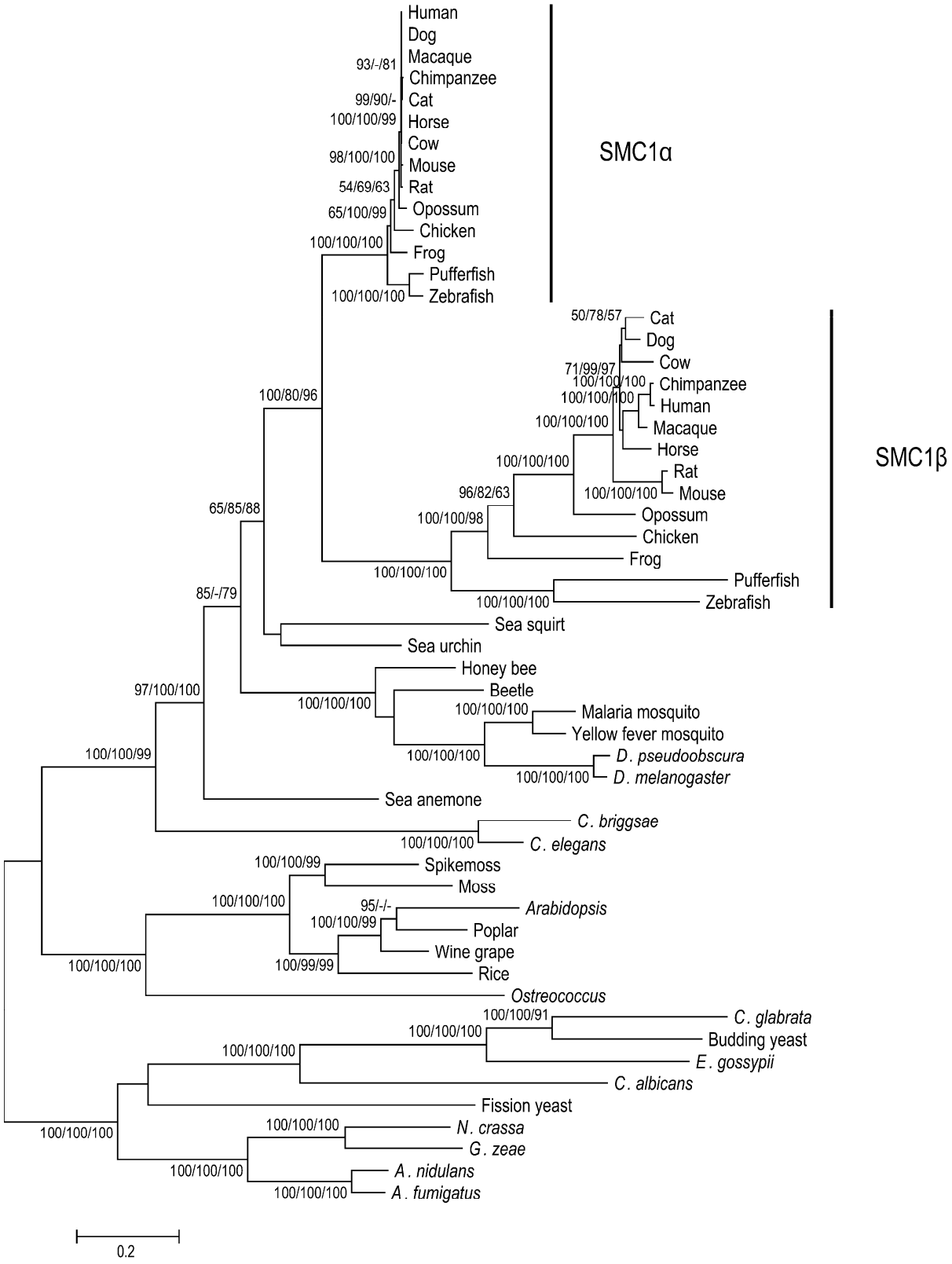


Fig. 3. A phylogenetic tree for the *SMC1* genes. Shown here is an ML-tree with all *SMC1* sequences, including both *SMC1α* and *SMC1β* from vertebrates, as well as *SMC1* from other animals, plants, and fungi. Bootstrap values are from analyses using ML/NJ/(NJ with the coiled-coil regions).

The *SMC1* genes of the two nematodes, *C. elegans* and *C. briggsae*, form a basalmost group outside of all other animal sequences. This can be explained by long-branch attraction because of the rapid evolution of the nematode genes, as observed for other gene families, such as the highly conserved *recA/RAD51* gene family (Lin et al., 2006). All three analyses (Fig. 3) also support the monophyly (bootstrap values of 65%/85%/88%) of the genes from vertebrates, sea squirt and sea urchin, all of which are deuterostomes, supporting the idea that they are more closely related than those from other animals. The sea anemone *SMC1* gene was placed outside the combined clade of deuterostome and insect *SMC1* genes. To investigate the relationship of other invertebrate *SMC* genes further, we also conducted phylogenetic analyses on the other five *SMC* subfamilies (Figs. 4–6). In the NJ trees of the *SMC2*, *SMC3*, and *SMC4* subfamilies (Figs. 4 and 5), the sea squirt is placed outside of the sea urchin, whereas the opposite is observed in the *SMC6* subfamily, although with low support (Fig. 6). In the *SMC5* subfamily, the sea squirt sequence is separated from all of the other genes (Fig. 6). Therefore, the phylogenetic positions of the *SMC* genes from sea squirt, sea urchin and sea anemone are not consistent.

To investigate the relationship of the *SMC1* genes further, we compared their sequences in detail at the residue level using a multi-sequence alignment. As an example, Fig. 7A shows the portion of the alignment for the hinge with conserved residues highlighted. We found that 10 amino acid residues were identical for the vertebrate *SMC1α* genes and the sea squirt and sea urchin *SMC1* genes (Fig. 7: B). In addition, 37 amino acid residues were conserved in vertebrate *SMC1α* genes and invertebrate *SMC1* genes (Fig. 7: B), whereas only 12 residues were shared in vertebrate *SMC1α* and *SMC1β* sequences, but not in invertebrate *SMC1* genes (not shown). Most of the 37 residues conserved among the *SMC1α* genes and invertebrate *SMC1* genes are in the conserved regions, while the majority of the twelve residues conserved among the vertebrate sequences occur in the less conserved domains—seven in the coiled-coil domains, one in the N terminus, one in the hinge, and three in the C-terminal domain. Because both the vertebrate *SMC1α* and the invertebrate *SMC1* genes have conserved functions in both mitosis and meiosis, the relatively large number of residues shared among these genes suggests that these residues might be important for conserved functions. The small number of residues that are in common between the *SMC1α* and *SMC1β* genes suggests that these two paralogs have divergent protein activities.

It has been shown that another cohesion subunit, Scc1, has a meiotic isoform, known as Rec8, found in fungal, animal and plant species (Parisi et al., 1999; Watanabe & Nurse, 1999; Pasierbek et al., 2001; Dong & Makaroff, 2001; Wang et al., 2003; Zhang et al., 2006). Its wide distribution suggests that meiosis-specific cohesin machinery originated early in eukaryotic evolution. An early-animal origin of the meiotic isoform *SMC1β* would be consistent with the distribution of conserved residues shown in Fig. 7B; however, the absence of such a meiotic isoform in all examined invertebrate animals, as well as in fungi and plants, unlike the situation for Scc1, makes it unlikely that the duplication for *SMC1α* and *SMC1β* occurred before the divergence of animals.

2.4 Divergence of vertebrate *SMC1* isoforms

In the tree shown in Fig. 3, the *SMC1β* genes have longer branches than the *SMC1α* genes, suggesting that the *SMC1β* genes might have evolved more rapidly. To further investigate the evolution of the meiosis-specific *SMC1* isoforms, we performed dN/dS analyses for two vertebrate pairs—human vs. mouse and human vs. chicken (Fig. 8). Our analysis suggests that *SMC1α* genes have been under purifying selection with dN/dS values lower than 0.1 (Fig. 8: A), consistent with the findings that *SMC1α* are more similar to the single copy *SMC1* genes in invertebrates. Although the dN/dS values were generally low across the *SMC1α* genes, there is a region near the 3′ end with slightly higher dN/dS values, suggesting potentially relaxed selection (Fig. 8: A). Moreover, the dN/dS values were higher for the human/chicken pair than for the human/mouse pair, perhaps reflecting functional divergence between human and chicken. To examine the divergence of *SMC1α* genes further, we performed pair-wise comparisons among vertebrate *SMC1α* genes (Table 3). The very low dN/dS ratios indicate that the vertebrate *SMC1α* genes have been under purifying selection, with mammalian genes generally having experienced the greatest pressure.

Compared with those of the *SMC1α* genes, the dN/dS values for *SMC1β* pairs were greater, between 0.1 and 0.4 (see Fig. 8B), suggesting that the meiotic isoform has been under reduced selection pressure. Also, the dN/dS values were similar between the pair of human/chicken and the pair of human/mouse, suggesting a possible acceleration of divergence between human and mouse because these two mammals have separated more recently than human and

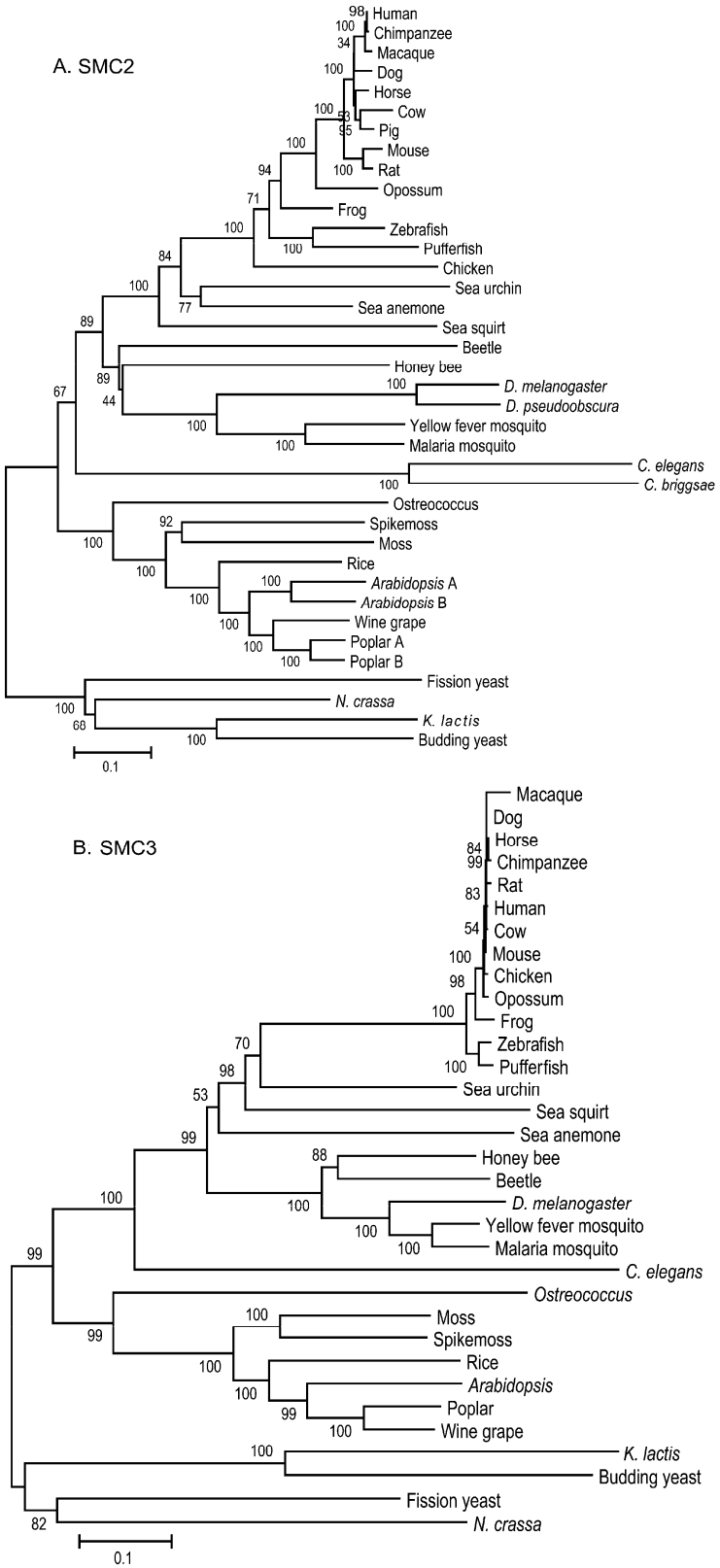


Fig. 4. NJ trees for the *SMC2* and *SMC3* subfamilies. **A,** A tree for the *SMC2* subfamily, with bootstrap values from NJ analysis. **B,** A tree for the *SMC3* subfamily, with bootstrap values from NJ analysis.

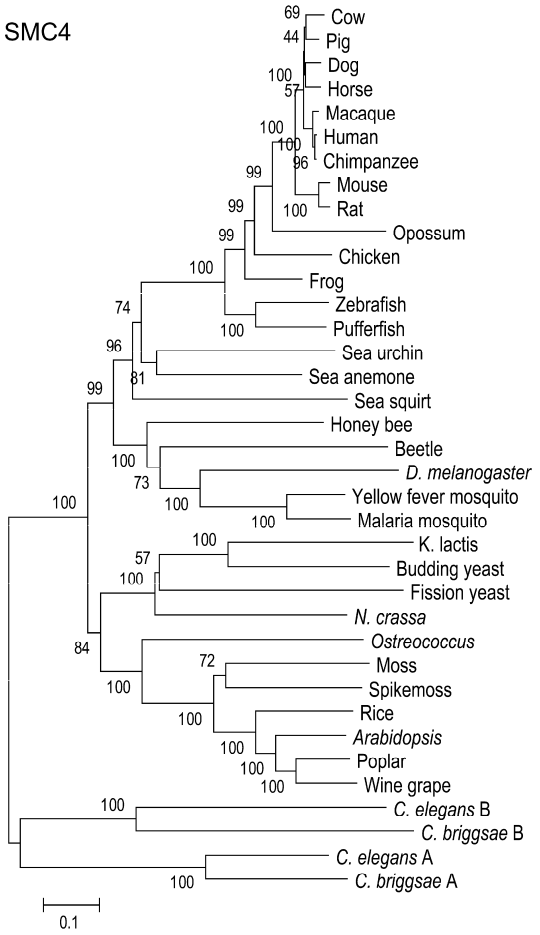


Fig. 5. An NJ tree for the *SMC4* subfamily, with bootstrap values from NJ analysis.

chicken. These results suggest that following gene duplication, the *SMC1α* isoform has been highly conserved under selection for its essential role in mitosis, whereas the *SMC1β* isoform has been allowed to diversify because it is only needed for meiosis, which represents a subset of the function of the ancestral *SMC1* gene.

2.5 Plants and nematodes have additional copies of some *SMC* genes

To understand the evolutionary history of *SMC2*–*SMC6* genes, we have performed phylogenetic analysis for each of these five paralogous sets (Figs. 4–6). To obtain stronger supports, we generated trees for representative subsets of sequences using three methods (Fig. 9). We found several relatively recent duplicates in some of these groups, including two copies of *SMC2* in both poplar and *Arabidopsis*, two copies of *SMC4* in nematodes, and two copies of *SMC6* in nematodes, *Arabidopsis* and spikemoss (*Selaginella*

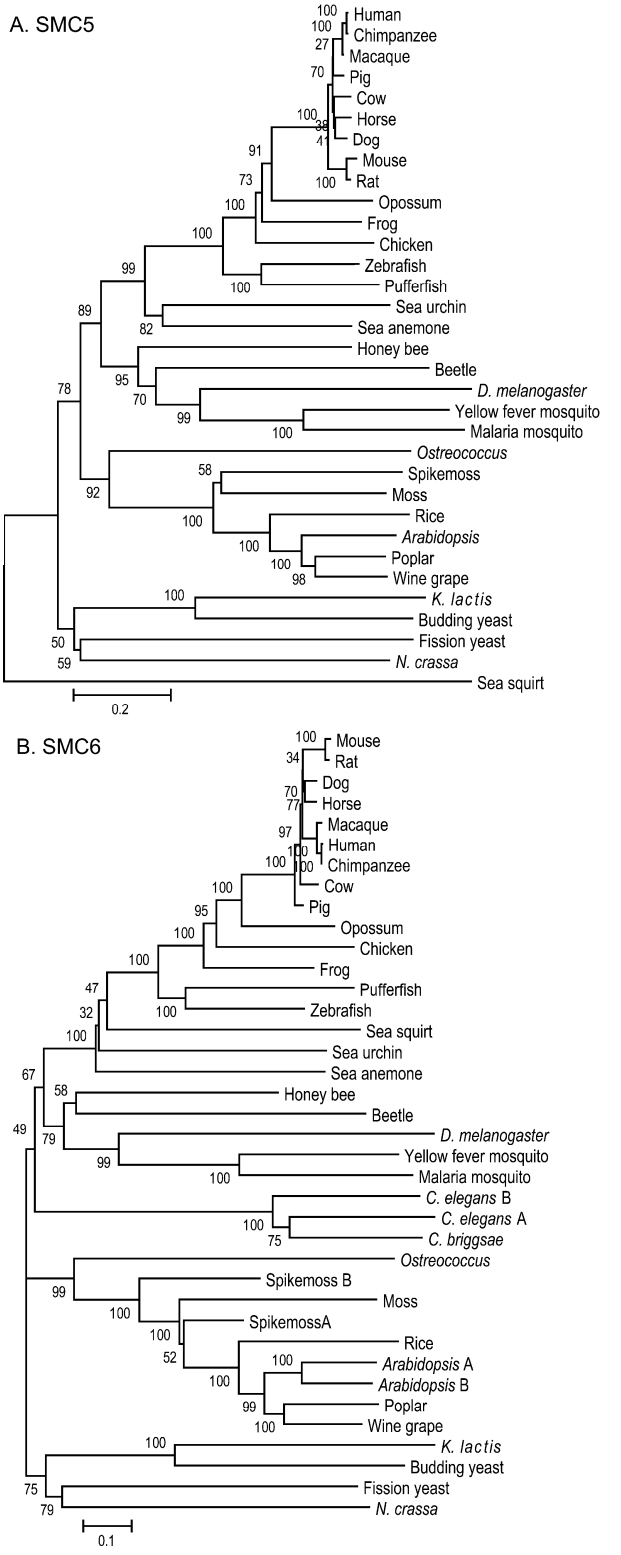


Fig. 6. NJ trees for the *SMC5* and *SMC6* subfamilies. **A,** A tree for the *SMC5* subfamily, with bootstrap values from NJ analysis. **B,** A tree for the *SMC6* subfamily, with bootstrap values from NJ analysis.

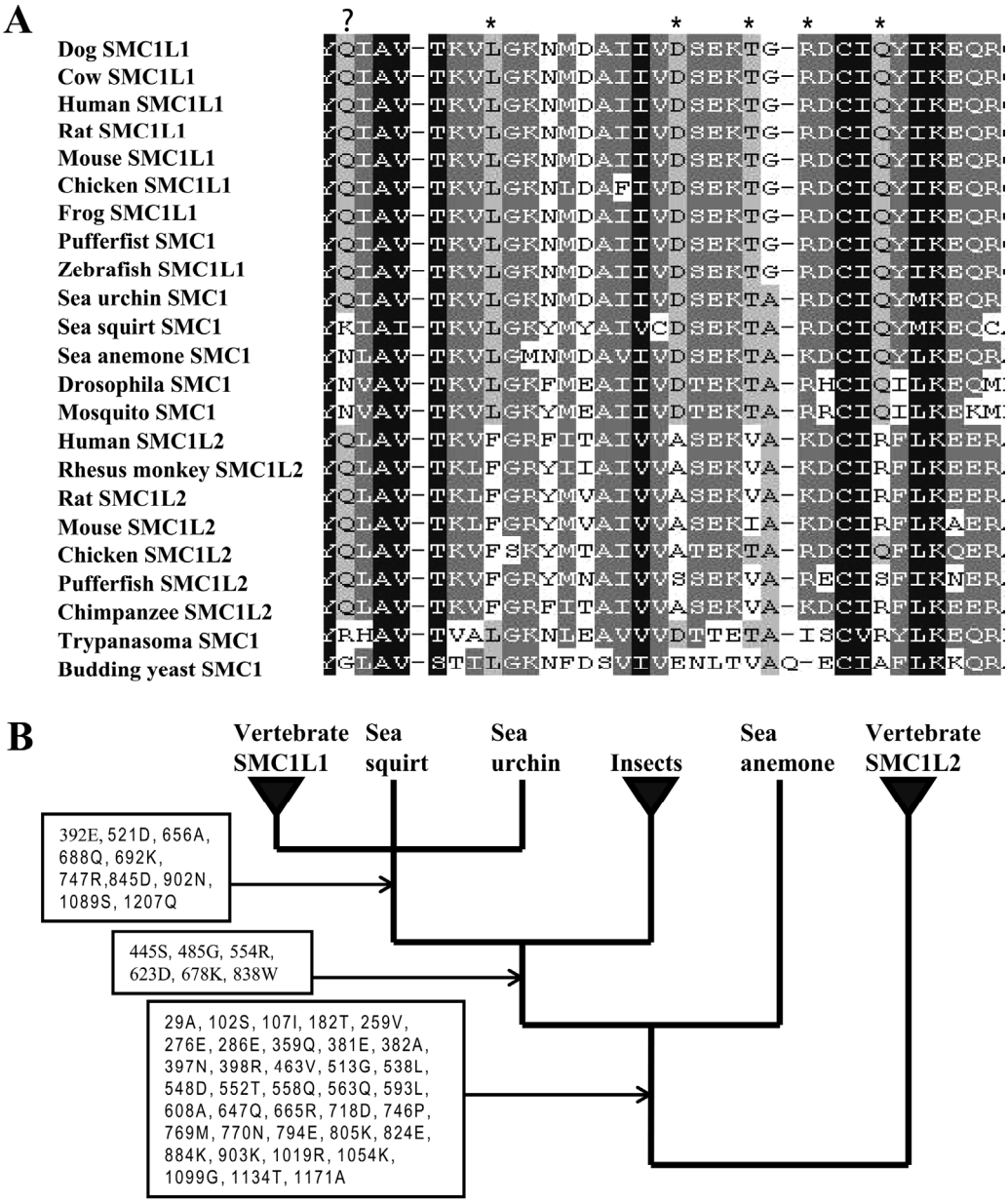


Fig. 7. Sequence analysis of SMC1 proteins. **A,** An alignment of Smc1α and Smc1β residues in the hinge domain. The residue marked by the question mark supports the node joining the vertebrate animal SMC1α and SMC1β clades. Residues marked by an asterisk (*) support the grouping of vertebrate *SMC1α* with the insect SMC1. **B,** A generalized phylogeny of the *SMC1* family with conserved amino acid residues. Residue numbers are from the human *SMC1β*.

moellendorffii) (Tables 1 and 2). An NJ tree of the *SMC2* and *SMC3* clades with plant sequences shows that the two copies of *SMC2* in *Arabidopsis* and poplar (Fig. 9, middle) were produced by two independent recent duplication events after the divergence of *Arabidopsis* and poplar. We noticed that grape had only one copy of *SMC2*, but as mentioned above, we identified a *SMC2* psedogene in the grape genome, demonstrating that grape also had a second copy of *SMC2* until recently. An NJ tree of the *SMC5* and *SMC6* groups with plant sequences shows that duplicated *Arabidopsis SMC6* copies form a clade (Fig. 9, bottom), indicating that they are recent duplicates. The other duplicated *SMC6* copy from spikemoss does not group together, suggesting that an ancient gene duplication mechanism occurred in the early ancestor of plants and that the *SMC6B* is only retained in spikemoss, although other permutations are also

Table 3 Substitution rates between vertebrate animal *SMC1α* genes

	Human	Chimpanzee	Monkey	Mouse	Rat	Dog	Cow	Horse	Opossum	Chicken	Frog	Zebrafish	Pufferfish
Human		0.009	0.035	0.362	0.381	0.233	0.252	0.211	1.169	1.097	1.443	1.429	1.563
Chimpanzee	0.189		0.029	0.349	0.365	0.226	0.249	0.202	1.164	1.075	1.448	1.417	1.531
Monkey	0	0.062		0.362	0.391	0.222	0.247	0.197	1.142	1.070	1.447	1.447	1.511
Mouse	0.006	0.011	0.006		0.131	0.455	0.471	0.438	1.406	1.557	1.668	1.577	1.724
Rat	0.004	0.010	0.004	0.026		0.464	0.457	0.443	1.380	1.570	1.752	1.640	1.704
Dog	0	0.008	0	0.007	0.006		0.225	0.193	1.142	1.088	1.413	1.434	1.632
Cow	0	0.007	0	0.006	0.006	0		0.206	1.203	1.118	1.418	1.454	1.537
Horse	0	0.009	0	0.007	0.006	0	0		1.148	1.110	1.410	1.481	1.479
Opossum	0.011	0.013	0.011	0.012	0.011	0.011	0.010	0.011		1.572	1.620	1.910	1.779
Chicken	0.037	0.040	0.039	0.028	0.028	0.038	0.037	0.038	0.030		1.651	1.848	1.309
Frog	0.023	0.025	0.024	0.022	0.021	0.024	0.024	0.025	0.025	0.038		1.811	1.725
Zebrafish	0.043	0.045	0.043	0.041	0.040	0.044	0.043	0.042	0.034	0.044	0.041		1.609
Pufferfish	0.041	0.043	0.043	0.039	0.040	0.039	0.042	0.043	0.038	0.061	0.043	0.024	

Synonymous substitution rates (dS) are shown above the diagonal (upper right) and the ratios of non-synonymous to synonymous substitution rates (dN/dS) are shown below the diagonal (lower left).

Human, *Homo sapiens*; Chimpanzee, *Pan troglodytes*; Monkey, *Macaca mulatta*; Mouse, *Mus musculus*; Rat, *Rattus norvegicus*; Dog, *Canis familiaris*; Cow, *Bos taurus*; Horse, *Equus caballus*; Opossum, *Monodelphis domestica*; Chicken, *Gallus gallus*; Frog, *Xenopus tropicalis*; Zebrafish, *Danio rerio*; Pufferfish, *Takafugu rubripes*.

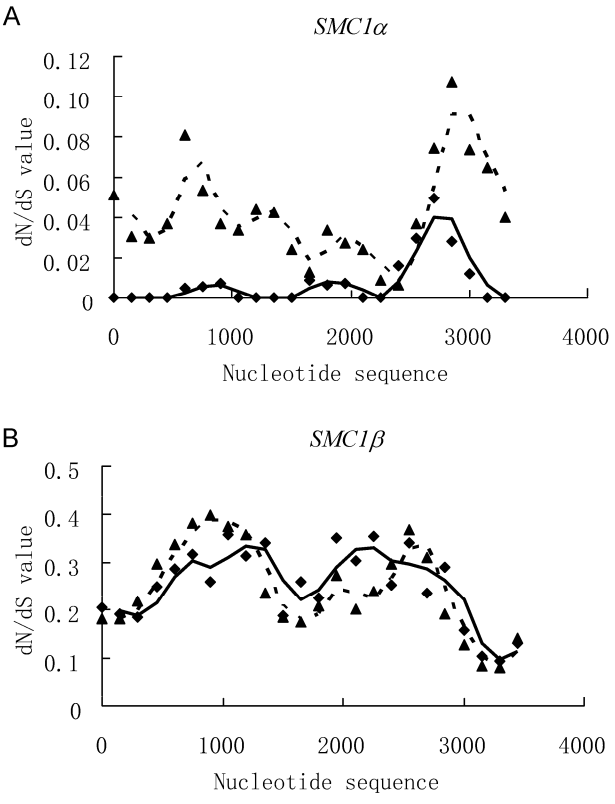


Fig. 8. dN/dS analysis of *Smc1* from human, chicken, and mouse sequences. **A,** Comparison of *SMC1α* (Smc1L1) between human and mouse (diamonds) and between human and chicken (triangles). **B,** Comparison of *SMC1β* (Smc1L2) between human and mouse (diamonds) and between human and chicken (triangles).

possible. Furthermore, the nematodes *C. elegans* and *C. briggsae* also have two *SMC4* genes and *C. elegans* has two *SMC6* genes. These copies were the result of duplication event(s) that occurred before the divergence of two species (Fig. 5). Our results indicate that *SMC2*, *SMC4*, and *SMC6* genes are sometimes duplicated due to lineage-specific recent duplications.

2.6 Effect of segmental duplications on *SMC* gene copy number

Our phylogenetic analysis showed that in general the *SMC* paralogs have remained stable over much of eukaryotic history, although a few duplication events have resulted in two copies for some paralogs in specific lineages. The chromosomal positions of the duplicated copies indicate that they are not likely the result of tandem duplication events. In plants, it is thought that large segmental (possibly genome-wide) duplication is the most common duplication event because most plant species are diploidized polyploids and contain many duplicated chromosome blocks in the genomes (Adams & Wendel, 2005). It has been proposed that *Arabidopsis* underwent three genome-wide duplication events, of which the most recent occurred 75±22 million years ago (Simillion et al., 2002). Similarly, the poplar genome is believed to have been duplicated more recently, about 8–13 million years ago (Tuskan et al., 2006; Jansson & Douglas, 2007). The genome of *S. cerevisiae* is also believed to have experienced an ancient duplication event (Kellis et al., 2004).

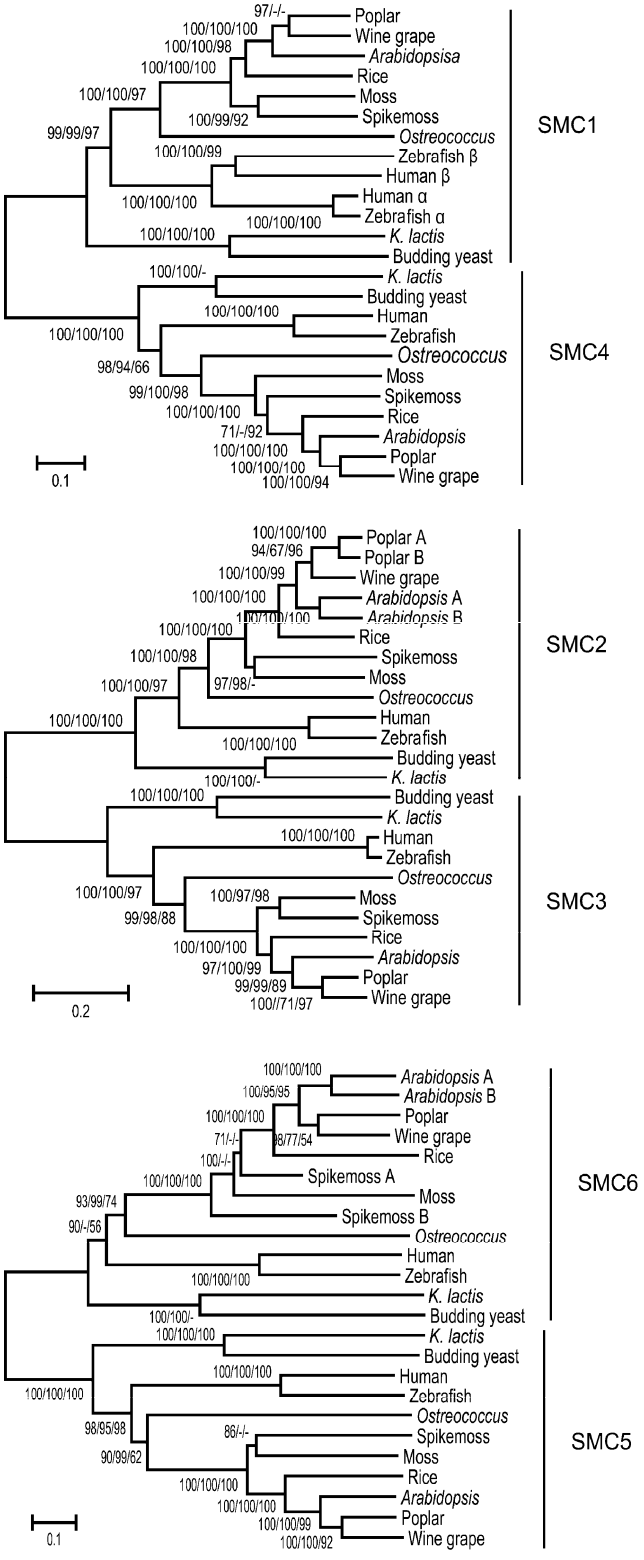


Fig. 9. Phylogenetic analysis of *SMC* genes from plants and other selected taxa. (top) An NJ tree of *SMC1* and *SMC4*; (middle) An NJ tree for *SMC2* and *SMC3*; (bottom) An NJ tree of *SMC5* and *SMC6*. Bootstrap supports are from NJ/ML/MP analyses.

To examine the possible mechanism for the duplication history of the *SMC* genes in *Arabidopsis* and poplar, we looked first for evidence of segmental duplications. Initially, we looked for evidence of genome/segmental duplications in yeast where massive gene loss and specialization took place. Our literature search identified duplicated sister regions in the budding yeast genome that had a 2:1 mapping with a related yeast species, *Kluyveromyces waltii*, which diverged before the ancient genome duplication in budding yeast (Kellis et al., 2004). We found that for *SMC2*, *SMC5*, and *SMC6*, there are three *K. waltii* tiles (genomic regions), tile 117, tile 1, and tile 56, respectively, that each matches to two regions in the budding yeast genome. For each set of three matching regions, *K. waltii* and one of the two budding yeast regions contained an *SMC* gene. The remaining region of the budding yeast genome only contained genes flanking the *SMC* gene, indicating that one of the duplicated *SMC* genes had been lost. Because the yeast genome has undergone frequent and massive chromosome rearrangements, the genome duplication evidence for the other three yeast *SMCs—SMC1*, *SMC3* and *SMC4*—was not clear.

In addition, to test whether segmental duplication was responsible for the duplicated plant *SMC* genes, we performed dot-matrix analyses to compare the genomic regions 50k upstream and downstream of the duplicated *SMC2* and *SMC6* genes, or with chromosomal regions containing genes similar to the genes flanking one of the other *SMC* genes. As shown in Fig. 10, we can see clearly that except for *AtSMC1*, all other *SMC* genes are associated with segmental duplications. Even for regions related to *AtSMC1,* there are also some conserved regions, but with different orientations, suggesting chromosomal rearrangements. Whereas the duplicated copies of *AtSMC2*, *AtSMC6*, and *PtSMC2* were retained, the one copy of the other *SMC* genes was lost, along with a region of approximately 30 kb adjacent to the *SMC* genes. Therefore, following the most recently genome-wide duplication in both *Arabidopsis* and poplar, most duplicated copies were lost, resulting in the retention of a single copy. The retention of duplicate copies of *AtSMC2, AtSMC6* and *PtSMC2* suggests that these genes provide a selective advantage, and that the retention of additional copies of the other paralogs might be deleterious. This supposition is buttressed by the fact that regions that contain *AtSMC1* and *ScSMC1* seem to have more frequent gene loss and genome rearrangement,

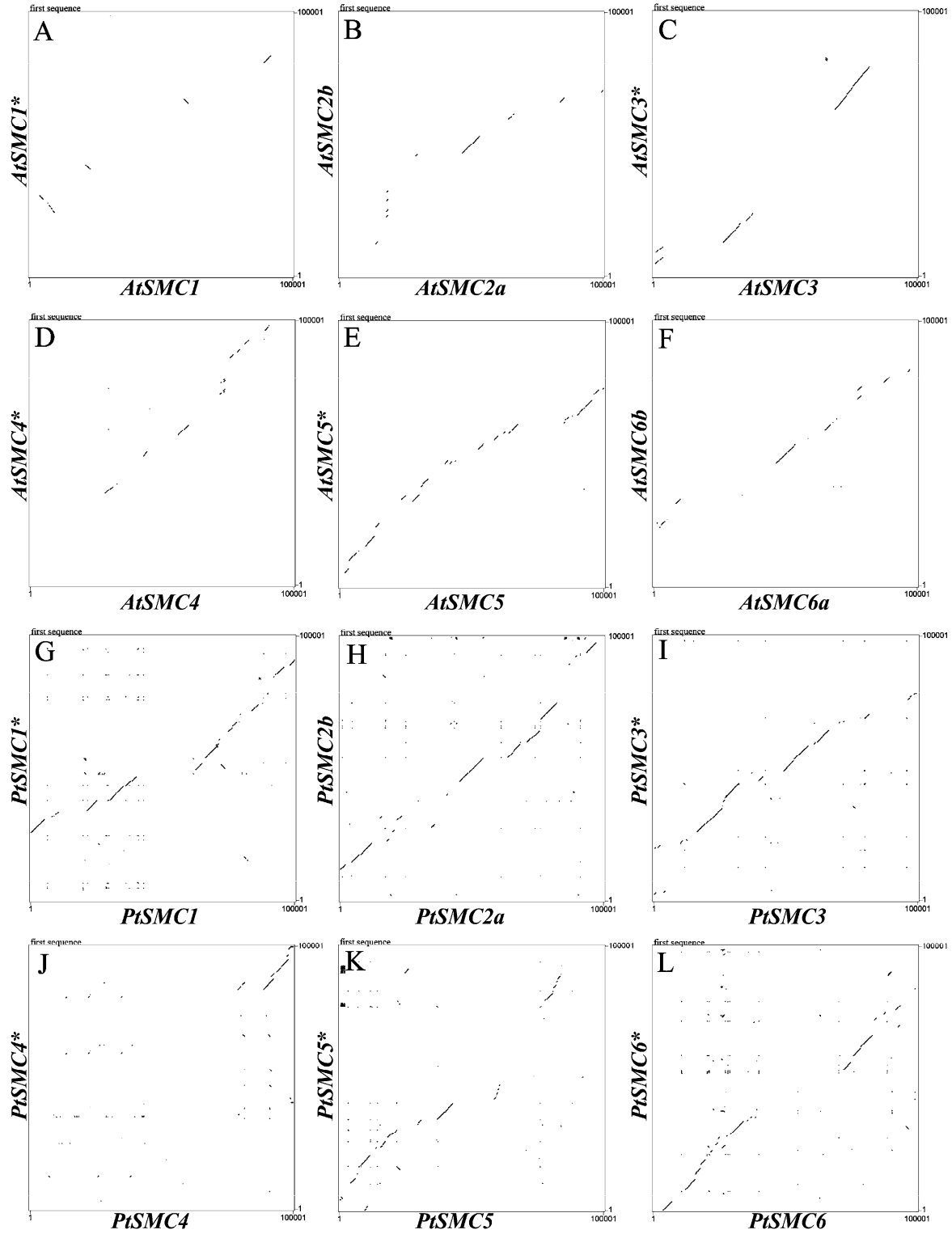


Fig. 10. Comparisons of the flanking genomic regions (50 kb on both upstream and downstream) of *AtSMC* and *PtSMC* genes. **A,** *AtSMC1* vs. *AtSMC1**; **B,** *AtSMC2a* vs. *AtSMC2b*; **C,** *AtSMC3* vs. *AtSMC3**; **D,** *AtSMC4* vs. *AtSM4**; **E,** *AtSMC5* vs. *AtSMC5**; **F,** *AtSMC6a* vs. *AtSMC6b*; **G,** *PtSMC1* vs. *PtSMC1**; **H,** *PtSMC2a* vs. *PtSMC2b*; **I,** *PtSMC3* vs. *PtSMC3**; **J,** *PtSMC4* vs. *PtSMC4**; **K,** *PtSMC5* vs. *PtSMC5**; **L,** *PtSMC6* vs. *PtSMC6**. All *Arabidopsis* and poplar *SMCs* can be explained by segmental duplication and gene loss events. * The hypothesized *SMC* genes that were lost during evolution.

exhibited by losses of long segments flanking *SMC1* and apparent chromosome rearrangements. The specific loss of *SMC1* and *SMC3*, both members of the cohesin complex, implies that sister chromatid cohesion may be a more tightly regulated process than either chromosome condensation (with *SMC2* and *SMC4*) or DNA repair (with *SMC5* and *SMC6*).

3 Conclusions

The *SMC* family is an excellent example of how the ever-growing body of genome sequence information can add to our understanding of the evolution of a family of proteins. We demonstrate that the origin of the meiotic isoform of *SMC1* is likely in the most recent common ancestor of vertebrates and that it has experienced divergence under relaxed selection. Additionally, we have shown that most of the duplicate copies of *SMC* genes in plants and yeast due to genome-wide duplication events were lost, with preferential retention of *SMC2*, *SMC4*, and *SMC6*. These results suggest that the condensin and DNA repair pathways are more flexible and able to accommodate multiple copies of their respective SMC proteins, whereas sister chromatid cohesion does not tolerate higher doses of *SMC1* and *SMC3*.

Acknowledgements A.S. was supported by the Huck Institutes of the Life Sciences and the Department of Biochemistry and Molecular Biology at the Pennsylvania State University. X.Z. was supported by the Huck Institutes of the Life Sciences. L.Q. was supported by the Department of Biology and by funds from Rijk Zwaan, the Netherlands. This work was supported by a grant from the US Department of Energy (DE-FG02-03ER15-448), and by funds from the Huck Institutes of the Life Sciences and from the Department of Biology, The Pennsylvania State University.

References

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. Current Opinions in Plant Biology 8: 135–141.

Beasley M, Xu H, Warren W, McKay M. 2002. Conserved disruptions in the predicted coiled-coil domains of eukaryotic SMC complexes: implications for structure and function. Genome Research 12: 1201–1209.

Cobbe N, Heck MM. 2004. The evolution of SMC proteins: phylogenetic analysis and structural implications. Molecular Biology and Evolution 21: 332–347.

Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, Pie J, Gil-Rodriquez C, Arnedo M, Loeys B, Kline AD, Wilson M, Lillquist K, Siu V, Ramos FJ, Musio A, Jackson LS, Dorsett D, Krantz ID. 2007. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. The American Journal of Human Genetics 80: 485–494.

Dong F, Cai X, Makaroff CA. 2001. Cloning and characterization of two *Arabidopsis* genes that belong to the RAD21/REC8 family of chromosome cohesin proteins. Gene 271: 99–108.

Edger RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity [online]. BMC Bioinformatics 5: 113. doi:10.1186/1471-2105-5-113. Available from http://www.biomedcentral.com [accessed 5 March 2008].

Eijpe M, Heyting C, Gross B, Jessberger R. 2000. Association of mammalian SMC1 and SMC3 proteins with meiotic chromosomes and synaptonemal complexes. Journal of Cell Science 113: 673–683.

Gilliland WD, Hawley RS. 2005. Cohesin and the maternal age effect. Cell 123: 371–373.

Guindon S, Gascuel O. 2003. PhyML—A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704.

Hirano T. 2002. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. Genes & Development 16: 399–414.

Hodges CA, Revenkova E, Jessberger R, Hassold TJ, Hunt PA. 2005. SMC1beta-deficient female mice provide evidence that cohesins are a missing link in age-related nondisjunction. Nature Genetics 37: 1351–1355.

Jansson S, Douglas CJ. 2007. Populus: a model system for plant biology. Annual Review of Plant Biology 58: 435–458.

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

Laemmli UK. 1978. Levels of organization of the DNA in eucaryotic chromosomes. Pharmacological Reviews 30: 469–476.

Lin Z, Kong H, Nei M, Ma H. 2006. Origins and evolution of the *recA/RAD51* gene family: Evidence for ancient gene duplication and endosymbiont gene transfer. Proceedings of the National Academy of Sciences USA 103: 10328–10333.

Liu N, Wang T. 2006. Protein-based phylogenetic analysis by using hydropathy profile of amino acids. FEBS Letters 580: 5321–5327.

Maeshima K, Laemmli UK. 2003. A two-fold scaffolding model for mitotic chromosome assembly. Developmental Cell 4: 467–480.

Marsden MP, Laemmli UK. 1979. Metaphase chromosome structure: evidence for a radial loop model. Cell 17: 849–858.

Melby TE, Ciampaglio CN, Briscoe G, Erickson HP. 1998. The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. Journal of Cell Biology 142: 1595–1604.

Musio A, Selicorni A, Focarelli ML, Gervasini C, Milani D, Russo S, Vezzoni P, Larizza L. 2006. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nature Genetics 38: 528–530.

Nasmyth K. 2005. How might cohesin hold sister chromatids together? Philosophical Transactions of the Royal Society 360: 483–496.

Nasmyth K, Haering CH. 2005. The structure and function of SMC and kleisin complexes. Annual Review of Biochemistry 74: 595–648.

Nicholas KB, Nicholas HB, Jr, Deerfield DW, II. 1997. GeneDoc: analysis and visualization of genetic variation. EMBNET News 4: 1–4.

Niki H, Imamura R, Kitaoka M, Yamanaka K, Ogura T, Hiraga S. 1992. *E. coli* MukB protein involved in chromosome partition forms a homodimer with a rod-and-hinge structure having DNA binding and ATP/GTP binding activities. EMBO Journal 11: 5101–5109.

Ohbayashi T, Oikawa K, Yamada K, Nishida-Umehara C, Matsuda Y, Satoh H, Mukai H, Mukai K, Kuroda M. 2007. Unscheduled overexpression of human WAPL promotes chromosomal instability. Biochemical and Biophysical Research Communications 356: 699–704.

Pasierbek P, Jantsch M, Melcher M, Schleiffer A, Schweizer D, Loidl J. 2001. A *Caenorhabditis elegans* cohesion protein with functions in meiotic chromosome pairing and disjunction. Genes and Development 15: 1349–1360.

Parisi S, McKay MJ, Molnar M, Thompson MA, van der Spek PJ, van Drunen-Schoenmaker E, Kanaar R, Lehmann E, Hoeijmakers JH, Kohli J. 1999. Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21 family conserved from fission yeast to humans. Molecular Cell Biology 19: 3513–3528.

Pati D, Zhang N, Plon S-E. 2002. Linking sister chromatid cohesion and apoptosis: role of Rad21. Molecular Cell Biology 22: 8267–8277.

Paulson JR, Laemmli UK. 1977. The structure of histone-depleted metaphase chromosomes. Cell 12: 817–828.

Ren QH, Yang H, Rosinski M, Conrad MN, Dreser ME, Guacci V, Zhang Z. 2005. Mutation of the cohesin related gene PDS5 causes cell death with predominant apoptotic features in *Saccharomyces cerevisiae* during early meiosis. Mutation Research 570: 163–173.

Revenkova E, Eijpe M, Heyting C, Hodges CA, Hunt PA, Lieve B, Scherthan H, Jessberger R. 2004. Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. Nature Cell Biology 6: 552–562.

Revenkova E, Jessberger R. 2005. Keeping sister chromatids together: cohesins in meiosis. Reproduction 130: 783–790.

Saitoh N, Goldberg I, Earnshaw WC. 1995. The SMC proteins and the coming of age of the chromosome scaffold hypothesis. Bioessays 17: 759–766.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. 2000. PipMaker—a web server for aligning two genomic DNA sequences. Genome Research 10: 577–586.

Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences USA. 99: 13627–13632.

Siomos MF, Badrinath A, Pasierbek P, Livingstone D, White J, Glotzer M, Nasmyth K. 2001. Separase is required for chromosome segregation during meiosis I in *Caenorhabditis elegans*. Current Biology 11: 1825–1835.

Soppa J. 2001. Prokaryotic structural maintenance of chromosomes (SMC) proteins: distribution, phylogeny, and comparison with MukBs and additional prokaryotic and eukaryotic coiled-coil proteins. Gene 278: 253–264.

Swofford DL. 2001. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0 Beta. Sunderland, MA: Sinauer Associates.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596–1599.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson D-R, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604.

Volkov A, Mascarenhas J, Andrei-Selmer C, Ulrich H-D, Grauman PL. 2003. A prokaryotic condensin/cohesin-like complex can actively compact chromosomes from a single position on the nucleoid and binds to DNA as a ring-like structure. Molecular Cell Biology 23: 5638–5650.

Wang F, Yoder J, Antoshechkin I, Han M. 2003. *Caenorhabditis elegans* EVL-14/PDS-5 and SCC-3 are essential for sister chromatid cohesion in meiosis and mitosis. Molecular and Cellular Biology 23: 7698–7707.

Watanabe Y, Nurse P. 1999. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. Nature 400: 461–464.

Yu H-G, Koshland D. 2007. The Aurora kinase Ipl1 maintains the centromeric localization of PP2A to protect cohesin during meiosis. Journal of Cell Biology 176: 911–918.

Zhang L, Tao J, Wang S, Chong K, Wang T. 2006. The rice OsRad21-4, an orthologue of yeast Rec8 protein, is required for efficient meiosis. Plant Molecular Biology 60: 533–554.