Curr Genet (1997) 31: 144–157

ORIGINAL PAPER

Natalie Yeager Stassen · John M. Logsdon Jr
Gaurav J. Vora · Hildo H. Offenberg
Jeffrey D. Palmer · Miriam E. Zolan

Isolation and characterization of *rad51* orthologs from *Coprinus cinereus* and *Lycopersicon esculentum*, and phylogenetic analysis of eukaryotic *recA* homologs

Received: 20 July / 25 October 1996

Abstract In eubacteria, the *recA* gene has long been recognized as essential for homologous recombination and DNA repair. Recent work has identified *recA* homologs in archaebacteria and eukaryotes, thus emphasizing the universal role this gene plays in DNA metabolism. We have isolated and characterized two new *recA* homologs, one from the basidiomycete *Coprinus cinereus* and the other from the angiosperm *Lycopersicon esculentum*. Like the *RAD51* gene of *Saccharomyces cerevisiae*, the *Coprinus* gene is highly induced by gamma irradiation and during meiosis. Phylogenetic analyses of eukarotic *recA* homologs reveal a gene duplication early in eukaryotic evolution which gave rise to two putatively monophyletic groups of *recA*-like genes. One group of 11 characterized genes, designated the *rad51* group, is orthologous to the *Saccharomyces RAD51* gene and also contains the *Coprinus* and *Lycopersicon* genes. The other group of seven genes, designated the *dmc1* group, is orthologous to the *Saccharomyces DMC1* gene. Sequence comparisons and phylogenetic analysis reveal extensive lineage- and gene-specific differences in rates of RecA protein evolution. Dmc1 consistently evolves faster than Rad51, and fungal proteins of both types, especially those of *Saccharomyces*, change rapidly, particularly in comparison to the slowly evolving vertebrate proteins. The *Drosophila* Rad51 protein has undergone remarkably rapid sequence divergence.

N. Y. Stassen · J. M. Logsdon Jr.[1] · G. J. Vora[2] · J. D. Palmer
M. E. Zolan (✉)
Department of Biology, Indiana University,
Bloomington, IN 47405, USA

H. H. Offenberg
Department of Genetics, Agricultural University,
Wageningen, The Netherlands

Present addresses:
[1] Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada
[2] Department of Microbiology, University of Massachussetts, Amherst, MA 01003, USA

Communicated by B. G. Turgeon

Key words *Coprinus cinereus · Lycopersicon esculentum · RAD51* homologs · Gene duplication · Unequal evolutionary rates · Molecular phylogeny

Introduction

The *recA* gene of *Escherichia coli* is necessary for homologous recombination and DNA repair (West 1992; Clark and Sandler 1994; Camerini-Otero and Hsieh 1995). Biochemical studies have shown that the RecA protein promotes identification and exchange of regions of homology (for reviews see Roca and Cox 1990; Kowalczykowski 1991; Clark and Sandler 1994; Camerini-Otero and Hsieh 1995), and the RecA protein is also central to the bacterial recombinational DNA-repair system, which responds to DNA damage (Cox 1993). *recA* genes have been isolated and studied from a large number (>65) of diverse eubacteria (Eisen 1995) and, recently, from three diverse archaebacteria (Sandler et al. 1996). In the last few years, multiple *recA*-like genes have been sequenced from several animals, fungi, and plants (for references, see Materials and methods).

In *Saccharomyces cerevisiae* four *recA* homology have been identified. Three of these genes, *RAD51*, *RAD55* and *RAD57*, belong to the *RAD52* epistasis group for DNA repair. All three of these genes are involved in recombination and recombinational repair (Game 1993); mutants in these genes are sensitive to DNA-damaging agents such as methyl methanesulfonate and ionizing radiation, and are defective in meiosis. The fourth *recA* homolog of *Saccharomyces*, *DMC1*, is required for meiosis, but does not appear to be involved in DNA repair (Bishop et al. 1992). The Rad51 protein can form a helical filament on both single- and double-stranded DNA, although the polarity of this filament is opposite to that formed by *recA* (Ogawa et al. 1993a; Sung and Robberson 1995). Also, like RecA, the Rad51 protein can catalyze homologous DNA pairing and strand exchange in an ATP-dependent manner (Sung 1994). The *RAD51* transcript is inducible during meiosis

(Shinohara et al. 1992), by irradiation with ultraviolet light, gamma rays (Aboussekhra et al. 1992) or X-rays (Basile et al. 1992), and by treatment with methyl methanesulfonate (Shinohara et al. 1992). The *DMC1* transcript is induced during meiosis (Bishop et al. 1992), although not by these DNA-damaging agents, and the Dmc1 and Rad51 proteins co-localize during meiosis (Bishop 1994).

We are currently studying the processes of meiosis and DNA repair in the basidiomycete *Coprinus cinereus.* This organism is particularly well suited for studies of meiosis, because its meiotic cycle is long and naturally synchronous (Raju 1972; Pukkila et al. 1984). Each mushroom cap contains 10^6–10^7 meiotic cells, which can be studied by light (Pukkila and Lu 1985; Zolan et al. 1988) and electron microscopy (Holm et al. 1981). We are using two complementary approaches to the identification of *Coprinus* genes which are involved in meiosis and DNA repair. The first approach is to isolate mutants defective in both processes. Using this strategy we have identified four genes necessary both for the survival of gamma irradiation and for meiosis (Zolan et al. 1988; Valentine et al. 1995). The second approach is to use the polymerase chain reaction (PCR) to amplify the *Coprinus* homologs of genes, such as *recA*, known to be important in processes of meiosis and DNA repair in other organisms.

In this study we used PCR to identify and clone a *Coprinus recA* homolog which shares extensive sequence identity with other fungal *rad51* orthologs. Consistent with a predicted role in both DNA repair and meiosis, the expression of *Coprinus rad51* is induced after gamma irradiation of tissue and during meiosis. In order to further pursue a phylogenetic study of eukaryotic *recA* homologs, we used a hybridization approach to isolate a *rad51* ortholog from tomato (*Lycopersicon esculentum*). This gene represents the first plant *rad51*-like gene isolated, and its inclusion in formal phylogenetic analyses facilitates the recognition of two ancient, widely-distributed families of *recA*-like genes in eukaryotes, orthologous to the *RAD51* and *DMC1* genes of *Saccharomyces*. These analyses also reveal considerable variation in rates of RecA protein sequence evolution among eukaryotes, with fungal and *Drosophila* sequences changing rapidly, and vertebrate sequences notably slowly.

Materials and methods

Strains. The primary *C. cinereus* strain employed in this study, used both for genomic DNA isolation (Zolan and Pukkila 1986) and for the Northern analysis shown in Fig. 4A, was Okayama-7 (O7; Wu et al. 1983). Chromosome plugs were made as described (Zolan et al. 1992) from O7, Java-6 [J6, a wild-type strain (Binninger et al. 1987)], and strain 218 (Binninger et al. 1987). A wild-type dikaryon (Valentine et al. 1995) was the source of the dikaryotic and meiotic RNA used for Fig. 4B. *L. esculentum* cv Evita seed was obtained from De Ruiterseeds, Bleiswijk, The Netherlands; anthers from 3–4 mm-long buds were used as a source of RNA for the construction of an anther cDNA library.

Gene amplification. Amino-acid sequences of eukaryotic *recA* homologs were aligned using the Clustal W program (Thompson et al. 1994) and adjusted by visual inspection. Degenerate primers were designed to amino-acid residues 125–131 [E(I/L/M)FGEFR; primer sequence = GAGCTGTTCGGCGA(G/A)TT(T/C)(C/A)G] and 165–160 [GETDIY; primer sequence = GAACGTTCCCTC (A/G/T) GT(A/G)TC]. Ten-microliter PCR reactions were performed with 15 ng of template (*Coprinus* O7 genomic DNA), 1 mM $MgCl_2$, 0.5 μM of each of the above primers, and 0.25 mM dNTPs in an Idaho Technology Air Thermo-cycler. The PCR conditions were: 4 min at 94 °C; followed by 40 cycles of 94 °C for 10 s, 55 °C for 20 s and 72 °C for 60 s; and ending with 6 min at 72 °C. Products were run on a 1.2% agarose gel, and major bands of sizes predicted from the protein sequences (assuming no introns or approximately one 50-bp intron per 200 bp of coding sequence; Skrzynia et al. 1989), were isolated using the Wizard PCR preps kit (Promega) and cloned into the pCRII vector (Invitrogen). Several clones of each PCR product were sequenced. A blast search at the NCB1 blast E-mail server (Altschul et al. 1990) was then used to identify clones that were similar to *recA* and its eukaryotic homologs.

DNA and genomic clone isolation. Poly A^+ RNA was isolated from mushroom caps at 1 h before and 1 h after karyogamy (see below). From a mixture of these samples, a cDNA library was constructed in the Uni-Zap XR vector (Stratagene). A lambda ZAP cDNA library of mRNA from *Lycopersicon* anthers was made using the cDNA library construction kit (Stratagene). A *Coprinus* cDNA library was screened by the method detailed in the Uni-Zap XR library instruction manual (Stratagene), using the insert portion of the cloned *Coprinus rad51* PCR product as the probe. The same insert was used to screen a genomic cosmid library of O7 DNA (May et al. 1991) using standard procedures for colony hybridizations (Sambrook et al. 1989), and genomic subclones were constructed using pBluescript (SK^+; Stratagene). The *Lycopersicon* cDNA library was screened with a probe derived from *Saccharomyces DMC1* (kindly provided by D.K. Bishop; Bishop et al. 1992). For screening, hybridization was done at 56 °C in 0.5 M $NaPO_4$, 7% SDS and 1 mM EDTA (Church and Gilbert 1984). Washes were performed at the same temperature with 0.75 M NaCl, 0.1 M Tris pH 7.8, 5 mM EDTA, 0.1% SDS. Among 1.5×10^5 phage screened, 11 positive clones were identified, all of which had overlapping restriction-enzyme maps. One clone was chosen for sequence analysis. For this purpose, several restriction enzyme fragments were subcloned in pBluescript (SK^+; Stratagene).

DNA sequencing. Sequencing of the *Coprinus* cDNA clone and genomic subclones was carried out on an automated sequencer (Li-Cor model 4000) following reaction preparation with SequiTherm thermostable DNA polymerase (Epicentre Technologies). Sequences were assembled using the DNAsis program (Hitachi Software Engineering Co.). The DyeDeoxy Terminator Cycle sequencing kit from Perkin Elmer was used for the automated sequencing of the *Lycopersicon* cDNA subclones, and the cDNA sequence was assembled using the University of Wisconsin GCG (version 7.0) sequence analysis package.

RNA isolation. For the examination of *Coprinus rad51* RNA levels after gamma irradiation, 25 ml of liquid YMG medium (Rao and Niederpruem 1969) was inoculated with small chunks of O7 tissue. After 2 days of growth at 37 °C with shaking, the culture was ground in a sterilized blender and added to 100 ml of fresh liquid YMG, allowed to grow for 2 days, then ground again and added to 875 ml of liquid YMG. After an additional 2 days of growth, the tissue was harvested onto Whatman filter paper by filtration through a Buchner funnel. An unirradiated sample was immediately frozen in liquid nitrogen and the rest of the tissue was irradiated with 40 krad using a ^{137}Cs irradiator (J.L. Shephard and Associates, model Mark I-68A). The tissue was then returned to liquid YMG medium, and aliquots were harvested at 1 h intervals for 6 h. Tissue was immediately frozen in liquid nitrogen upon harvesting and stored at –80 °C. For meiotic RNA samples, the cap tissue was separated from the stipe, and samples were immediately frozen in liquid nitrogen and stored at –80 °C. Frozen tissue was ground to a fine powder in a coffee grinder which had been pre-chilled with dry ice. Approximately 5 g of tis-

sue was used for each haploid time point, and 0.5–1.0 g of cap or stipe tissue was used for the meiotic time points. The extraction buffer was made from the following RNase-free components: 0.5 M NaCl, 0.2 M Tris-HCl, pH 7.5, 0.1 M EDTA, 1% SDS, 0.5% β-mercaptoethanol and 0.5% diethylpyrocarbonate (DEPC). A volume of extraction buffer equal to the volume of tissue was mixed with the tissue in sterile Oak Ridge tubes. An equal weight of glass beads (0.5–0.7 mm), which had previously been washed in nitric acid, was added to the Oak Ridge tube along with a volume of phenol equal to the volume of extraction buffer used. This mixture was vigorously vortexed for 2 min. Another equal volume of phenol was then added along with an equal volume of SEVAG (chloroform:isoamyl alcohol, 24:1) and the tube was vortexed briefly. The tube was then centrifuged at 15 000 rpm, in a SS34 rotor, for 15 min at 4 °C. The aqueous phase was re-extracted with equal volumes of a 1:1 phenol:SEVAG mixture until no white pellicle was seen at the interface. A final extraction was then performed with an equal volume of SEVAG. The RNA was then precipitated with 2 volumes of 100% ethanol and 0.05% DEPC for at least 30 min at –20 °C. The RNA was collected by centrifugation and washed with 70% ethanol and 0.05% DEPC. The RNA was re-suspended in RNase-free TE (10 mM Tris, pH 8.0, 1 mM EDTA) and stored at –80 °C. Poly A^+ RNA was isolated by the Poly AT-tract mRNA isolation system (Promega). Frozen anthers (stored under liquid nitrogen) from *Lycopersicon* were powdered in a Mikro-desmembrator II (Braun) for 1 min. The powder was used for RNA isolation by a guanidine isothiocyanate/LiCl method (Cathala et al. 1983). Poly $(A)^+$ RNA was purified by affinity chromatography on oligo (dT)-cellulose (Aviv and Leder 1972).

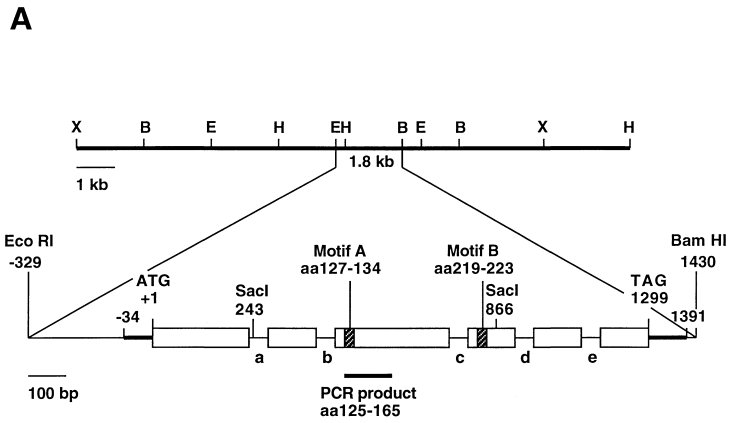
Electrophoresis and hybridizations. 2.5 µg of *Coprinus* O7 genomic DNA was digested with the indicated restriction enzymes, and the fragments were separated on a 0.8% agarose gel in TBE buffer (90 mM Trizma-base, 90 mM Boric acid, and 2 mM EDTA). Individual chromosomes were separated using 1% LE agarose (Beckman) for small chromosomes or 0.9% chromosomal grade agarose (BioRad) for large chromosomes (Zolan et al. 1992). Fragments or chromosomes were transferred to Magnagraph nylon membranes (MSI) and fixed by exposure to ultraviolet light. Membranes were pre-washed for 1 h or more in 1×SSC, 0.5% SDS at 65 °C. Pre-hybridization and hybridization were carried out in 4×SSC, 1% SDS, and 0.5% non-fat dry milk. Probes were made by random priming (Sambrook et al. 1989) using exonuclease-minus Klenow (Stratagene); specific activities were greater than 1×10^9 µg. Membranes were hybridized for at least 16 h and then washed twice at room temperature for 5 min in 2×SSC, 0.5% SDS and then twice for 1 h at 65 °C, in 0.2×SSC, 0.1% SDS. For the *Coprinus* Northerns, each lane contained 25 µg of glyoxal-treated total RNA, run on a 1% agarose gel in 10 mM of sodium phosphate buffer (Sambrook et al. 1989). RNA was transferred to a Magnagraph (MSI) membrane in 20×SSC and fixed to the membrane by UV. Membranes were prewashed for more than 1 h in 1×SSC, 0.5% SDS at 65 °C. Pre-hybridization and hybridization were carried out in 1 M NaCl, 1% SDS, 5% dextran sulfate. Probes were made using the same procedure used to generate probes for Southerns. Blots were hybridized for 48 h or more and then washed twice at room temperature for 5 min in 2×SSC, 0.5% SDS and then once in the same solution at 65 °C for at least 1 h. A final wash was then done with 1×SSC, 0.1% SDS for at least 1 h at 65 °C.

Sequence comparisons and phylogenetic analyses. Sources of the 23 RecA sequences analyzed in this study are as follows (in order from top-to-bottom as presented in Figs. 6–8, and using the *original* protein names; cf. to Fig. 7): *Coprinus cinereus* Rad51, this report; *Saccharomyces cerevisiae* Rad51, Basile et al. (1992), Shinohara et al. (1992); *Schizosaccharomyces pome* Rad51, Muris et al. (1993), Shinohara et al. (1993), Jang et al. (1994); *Neurospora crassa* Mei3, Cheng et al. (1993), Hatakeyama et al. (1995); *Homo sapiens* Rad51, Shinohara et al. (1993), Yoshimura et al. (1993); *Mus musculus* Rad51, Morita et al. (1993), Shinohara et al. (1993); *Gallus gallus* Rad51, Bezzubova et al. (1993); *Xenopus laevis* Rad51.1 and Rad51.2, Maeshima et al. (1995); *Drosophila melanogaster* Dmr1, Akaboshi et al. (1994), McKee et al. (1996); *Lycopersicon esculentum* Rad51, this report; *Candida albicans* Dlh1, Diener and Fink (1996); *Saccharomyces cerevisiae* Dmc1, Bishop et al. (1992); *Schizosaccharomyces pombe* Dmc1, A. Shinohara and A. Yamazaki, personal communication; *Homo sapiens* Dmc1/Lim15, Sato et al. (1995a), Habu et al. (1996); *Mus musuclus* Dmc1/Lim15, Sato et al. (1995b), Habu et al. (1996); *Arabidopsis thaliana* Lim15, Sato et al. (1995c); *Lilium longiflorum* Lim15, Kobayashi et al. (1994); *Methanococcus jannaschii* RadA and *Sulfolobus sofataricus* RadA, Sandler et al. (1996); *Escherichia coli* RecA, *Bacillus subtilis* RecA, and *Thermus aquaticus* RecA, Eisen (1995). The RecA alignment shown in Fig. 6 was generated by starting with the alignment of Sandler et al. (1996), which included a total of seven representative eubacterial, archaebacterial and eukaryotic sequences, and manually adding the additional sequences included here. This resulted in an alignment which is virtually identical to that of Sandler et al. (1996). Sequences of the RecA-homologous *Saccharomyces* proteins Rad55 and Rad57 were excluded from the alignments since they are highly divergent (Lovett 1994) and therefore extremely difficult to align with confidence. Similarly, the Rec2 protein of *Ustilago maydis* was excluded since its similarity to other RecA homologs is extremely limited (Heyer 1994). A *recA*-like, expressed-sequence-tag cDNA sequence from *Caenorhabditis elegans* (clone CEMSE, McCombie et al. 1992) was excluded because it is grossly incomplete (approximately 40% of the vertebrate *rad51* gene length). The termini of eubacterial RecA sequences do not align with the eukaryotic sequences and were therefore excluded. The regions of the RecA amino-acid alignment marked in Fig. 6 were used for phylogenetic analyses with maximum parsimony and neighbor joining, while the first and second codon positions of the corresponding nucleotide alignment were used in analyses with these two methods and with maximum likelihood. All phylogenetic analysis were carried out using PAUP version 4.Od47 (Swofford 1996). Default parameters were used on all analyses, except that the parsimony trees were built using the steepest-descent option. For bootstrap re-sampling (Felsenstein 1985; Swofford et al. 1996), 100 replicates were performed with ten random taxa additions per replicate.

Results and discussion

Isolation of a *recA* homolog from *Coprinus*

A conserved domain of RecA and its homologs, called the "homologous core" (Ogawa et al. 1993b), includes portions of the RecA protein known to be involved in UV-resistance, recombination, active oligomer formation, and ATP binding (Ogawa et al. 1992). In order to isolate *Coprinus recA* homologs, two degenerate PCR primers (see Materials and methods) were used in PCR reactions with *Coprinus* genomic DNA as template. This primer pair produced a 123-bp product which showed strong sequence similarity to *rad51*-type *recA* homologs. This PCR product was used as a probe to isolate cDNA and genomic clones of the complete *Coprinus rad51* gene by hybridization screening of cDNA and genomic libraries. The two cDNA clones isolated contained 1.2-kb inserts, one of which was completely sequenced. Four overlapping genomic clones were identified, and the genomic region surrounding the *rad51* gene was mapped (Fig. 1A). Using the cDNA clone as a probe, we identified a 1.8-kb *Eco*RI-*Bam*HI genomic fragment which contains the entire *Coprinus rad51* gene. This 1.8-kb region was subcloned into three *Sac*I fragments (Fig. 1A), which were completely sequenced (Fig. 1B, GenBank accession number U21905).



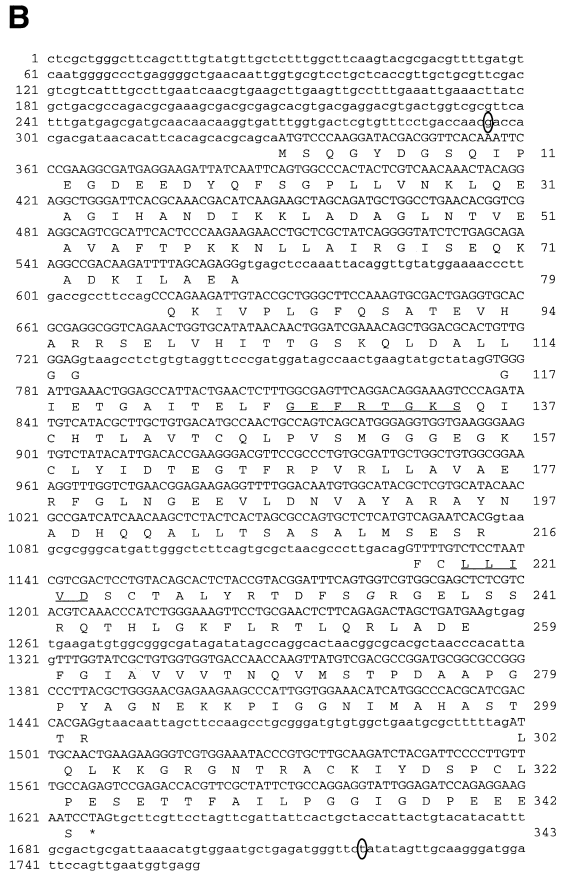


Fig. 1 Genomic structure (**A**) and sequence (**B**) of the *Coprinus rad51* gene (GenBank accession number U21905). **A** *thin lines* represent flanking and intron sequences, and *thick lines* represent 5′ and 3′ UTR sequences in the sequenced cDNA clone. Exons are depicted as *open boxes*, the five introns are numbered *a–e*, and the two ATP-binding motifs are shown as *hatched boxes.* All *numbers* correspond to the nucleic acid and amino-acid (aa) numbers in **B**. **B** The positions of the beginning and end of the cDNA clone sequence are *circled.* The coding sequence is in *capital letters*, and the flanking and intron sequences are in *small letters.* ATP-binding motifs are *underlined*, and the stop codon is denoted by an *asterisk.* The *numbers on the left* indicate the cumulative length in nucleotides, while the *numbers on the right* indicate the length in amino acids

The *Coprinus rad51* cDNA sequence contains a 1032-bp open reading frame (Fig. 1B) which, according to our sequence alignments and phylogenetic analyses (see below), encodes the *Coprinus* ortholog of *Saccharomyces RAD51*. Comparison of *Coprinus rad51* cDNA and genomic sequences revealed that the coding region of the gene is interrupted by five introns (Fig. 1A). The first of these introns is identical in position to an intron in the *Neurospora crassa rad51* (*mei-3*) gene (Cheng et al. 1993; Hatakeyama et al. 1995), while the second *Coprinus* intron is identical in position to the sixth intron in *Arabidopsis dmc1* (Sato et al. 1995c). The exons of *Coprinus rad51* range in size from 126 to 301 bp, while the introns range in size from 49 to 66 bp. The presence of numerous short introns is consistent with the structure of five other *Coprinus* genes (*Ccbetal-1*, Tymon et al. 1992; *Ccras*, Genbank X70789; *cip*, Genbank D13295 and D13226; *tpi*, Logsdon et al. 1995; *trp1*, Skrzynia et al. 1989). The 43 introns in these six genes range in size from 44 to 172 bp (average size = 59 bp), and the coding regions of the exons are 9 to 577 bp in size (average size = 177 bp). The 5′ and 3′ intron boundaries of the *Coprinus rad51* introns (Fig. 1B) agree well with the consensus sequence seen for the other *Coprinus* genes, $5'\text{-}G_{70}/G_{100}T_{100}R_{86}M_{77}G_{95}T_{67}\cdots Y_{95}A_{100}G_{100}/\text{–}3'$ (where for X_{NNN}, the subscript is the percentage of introns with the given nucleotide X). This consensus sequence is consistent with those previously found in other fungi (Gurr et al. 1987; Rymond and Rosbash 1992; Edelmann and Staben 1994) and in mammals (Rymond and Rosbash 1992).

Chromosomal location and copy number of *Coprinus rad51*

Haploid strains of *C. cinereus* have 13 chromosomes, which range in size from 1 to 5 megabases (Pukkila and Casselton 1991). The *Coprinus rad51* gene was mapped to chromosome 1 by hybridization of a genomic *rad51* clone to Southern blots of pulsed-field gels resolving either small or large chromosomes (the latter separation is shown in Fig. 2). In previous work, we identified by mutagenesis four genes, *rad3*, *rad9*, *rad11* and *rad12*, which are part of the same pathway for the survival of gamma irradiation and which are also required for meiosis in *Coprinus* (Zolan et al. 1988; Valentine et al. 1995; Zolan et al. 1995). One of these genes, *rad9*, has been sequenced (Seitz et al. 1996) and all four have been mapped (Zolan et al. 1992; 1993). None of these *rad* genes maps to chromosome 1, and there is no sequence similarity between *rad9* and *rad51*. Therefore, *rad51* is a fifth *Coprinus* gene likely to be required for both DNA repair and meiosis. The construction of a mutant of the *Coprinus rad51* gene will allow us to determine whether it is indeed part of the same

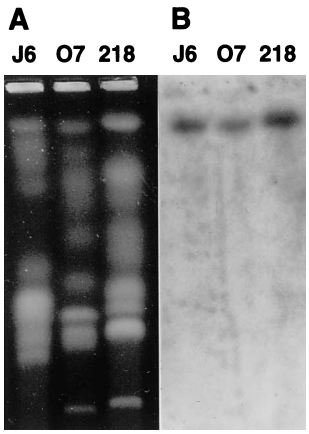


Fig. 2A, B Chromosomal location of *Coprinus rad51*. **A** a 0.9% chromosomal-grade agarose gel was run at 60 V for 144 h with a pulse time of 22 min. **B** the gel was blotted and hybridized with a *Coprinus rad51* genomic clone

DNA repair and meiotic pathways as the *rad* genes previously identified.

Coprinus rad51 appears to be a single-copy gene. DNA from the strain used for cloning was digested with three enzymes which do not cut within the gene. A Southern blot of the gel was hybridized with the *rad51* cDNA, and in each case one band of hybridization was seen (Fig. 3). Therefore, if other *recA* homologs are present within the *Coprinus* genome, they are divergent enough to be undetectable using the high-stringency conditions we employed.

Expression pattern of *Coprinus rad51*

Since *Coprinus rad51* is orthologous to the *Saccharomyces RAD51* gene, we predicted that its expression would be induced after gamma irradiation and during meiosis, as found for the yeast gene (Shinohara et al. 1992). Time-course studies of tissue after irradiation and during meiosis showed a dramatic induction of *rad51* expression in *Coprinus* (Fig. 4). We probed Northern blots of RNA isolated from irradiated and meiotic tissue and observed a 30-fold induction of a 1.2-kb transcript after 2 h of growth following irradiation (Fig. 4A) and a greater than 50-fold induction, of a transcript of the same size, at 6 h post-karyogamy (Fig. 4B). A second transcript, of 1.5 kb, was detected in irradiated but not meiotic tissue when the Northerns were probed with the complete cDNA. We believe the smaller, 1.2-kb, transcript is the *Coprinus rad51* message for three reasons. First, the 1.2-kb transcript is induced in both irradiated and meiotic tissue, as expected for a *RAD51* homolog. Second, the 1.5-kb transcript was not seen when

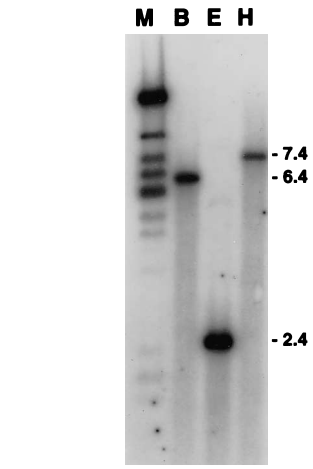


Fig. 3 Southern analysis of *Coprinus* O7 genomic DNA. A filter-blot containing total DNA digests with *Bam*HI (*B*), *Eco*RI (*E*), or *Hin*dIII (*H*) was probed with the *Coprinus rad51* cDNA clone. The sizes of the genomic fragments are indicated on the left and were determined by comparison to lambda *Eco*RI-*Hin*dIII double-digest fragments (*M*)

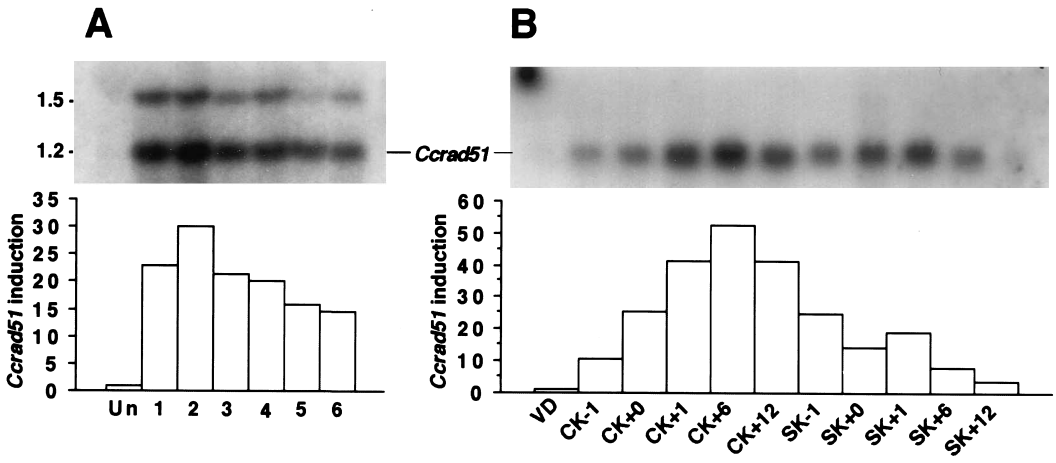


Fig. 4A, B Northern-blot analysis of *Coprinus rad51* RNA. **A** 25 µg of total RNA isolated from strain O7 tissue before (*Un*) or after irradiation and return to growth for 1 to 6 h was probed with the *Coprinus rad51* cDNA clone (*CCrad51*). **B** 25 µg of total RNA isolated from caps and stipes during the course of meiosis was probed with the same clone. Vegetative dikaryon (*VD*) tissue was used as a control for induction of the *rad51* during the course of meiosis in cap (C) tissue and stipe (S) tissue at 1 h before karyogamy (*CK-1* and *SK-1*), at karyogamy (*CK + 0* and *SK + 0*), and at 1, 6 and 12 h after karyogamy. As a loading control, the same blots were hybridized with a probe made by labelling a clone of the *Coprinus* ribosomal RNA gene repeat (Wu et al. 1983). Both hybridizations were quantified by use of a phosphorimager (Molecular Dynamics), and the amount of *rad51* hybridization in each lane was normalized using the control data. Graphs represent the level of *rad51* induction compared to the samples from unirradiated tissue (*Un*) for panel **A**, and compared to a vegetative dikaryon (*VD*) for panel **B**

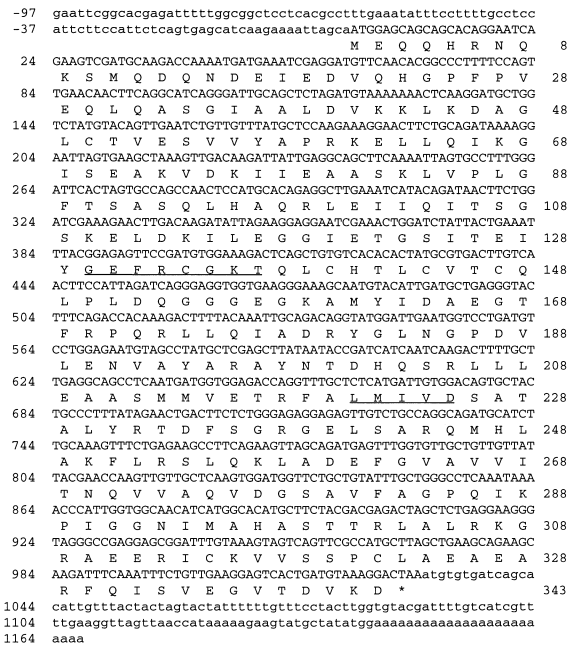


Fig. 5 The nucleotide and amino-acid sequences of *Lycopersicon rad51*, as determined from a cDNA clone (GenBank accession number U22441). The coding sequence is in *capital letters* and the flanking sequences are in *small letters.* ATP-binding motifs are *underlined*, and the stop codon is denoted by an *asterisk.* The *numbers on the left* indicate the cumulative length in nucleotides, while the *numbers on the right* indicate the length in amino acids

the Northern in Fig. 4A was probed with the original, 123-bp, PCR product, which represents the most conserved part of the *recA* homologs (Figs. 1A and 6). Third, a *Sac*I-*Bam*HI clone (Fig. 1A) containing the 3′ end of the *Coprinus rad51* coding region and 131 bp of flanking non-coding sequence hybridized more strongly to the larger, 1.5-kb transcript than to the 1.2-kb transcript, while a 1.4-kb *Bam*HI genomic fragment directly downstream from *rad51* (Fig. 1A) hybridized exclusively (and strongly) to the 1.5-kb transcript (data not shown). Therefore, we hypothesize that the 1.5-kb transcript is derived from a gene which abuts, or perhaps partially overlaps, *rad51* and which is similarly regulated after irradiation but not during meiosis.

Isolation and expression of *Lycopersicon rad51*

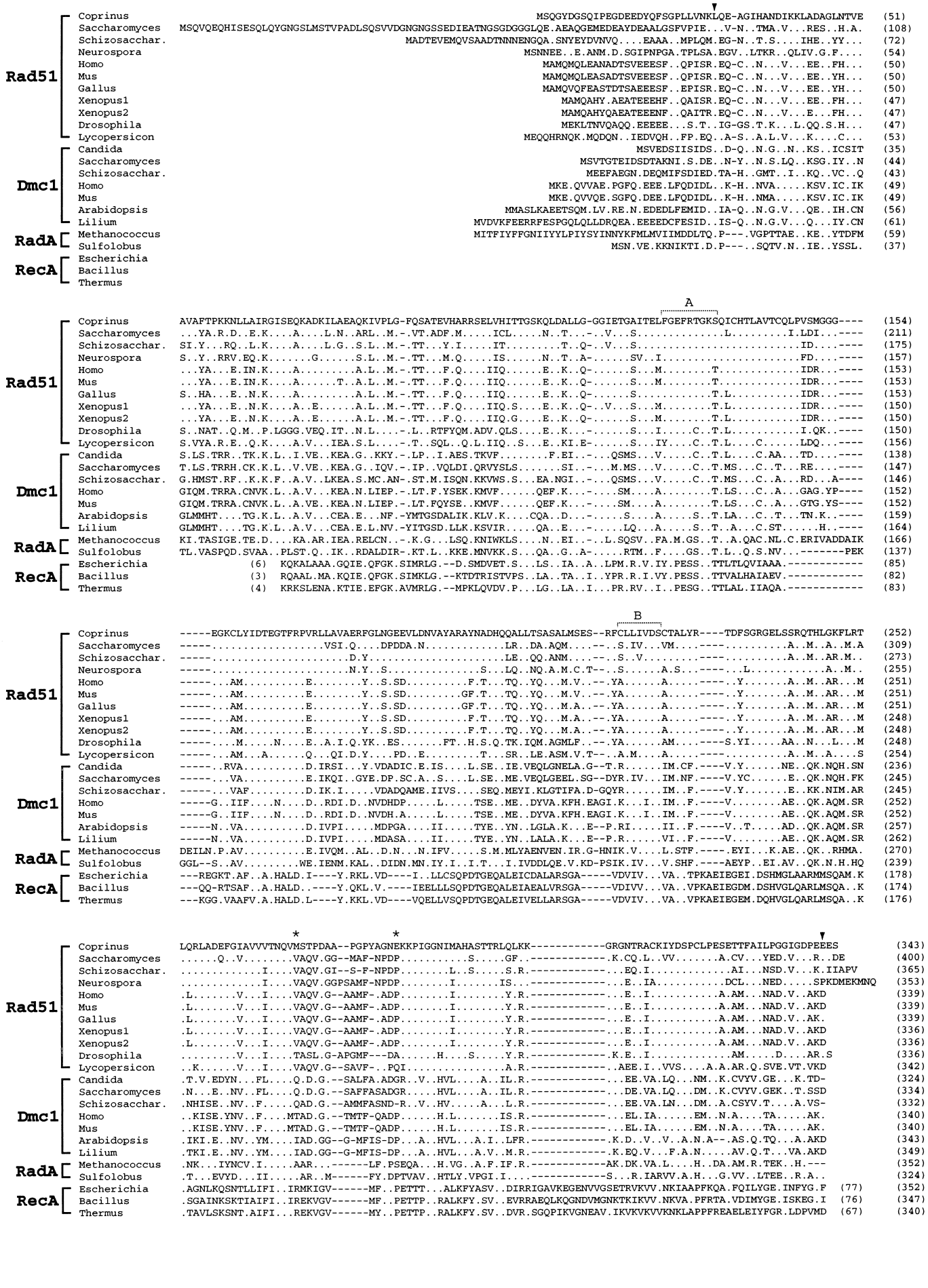
To isolate a *recA* homolog from tomato (*L. esculentum*), a cDNA library was constructed using RNA isolated from anthers. This library was screened (as described in Materials and methods) at lowered stringency using a *Saccharomyces DMC1* probe (Bishop et al. 1992). Analysis of the cDNA clone chosen for sequencing revealed an open reading frame of 1026 nucleotides (Fig. 5, GenBank accession number U22441), which, according to our sequence alignments (Fig. 6) and phylogenetic analyses (see below), encodes the *Lycopersicon* ortholog of *rad51*. Northern analysis showed that this gene is expressed in both leaf and anther tissue (data not shown).

Phylogenetic analysis of eukaryotic *recA* homologs

A number of genes with amino-acid-sequence similarity to eubacterial RecA have recently been identified in eukaryotes, including the two new sequences reported here. To elucidate their evolutionary history, especially that of gene duplication, we carried out molecular phylogenetic analyses of proteins encoded by members of this gene family. In *Saccharomyces*, four RecA-homologs have been identified, Rad51, Rad55, Rad57 and Dmc1. Although Rad51 and Dmc1 share considerable sequence identity (49%, Fig. 7), Rad55 and Rad57 are much more divergent [both show only about 31% and 17% identity to Rad51 and Dmc1, respectively, according to the alignment of Lovett (1994)]. Because of the low level of sequence similarity of Rad55 and Rad57 to the other eukaryotic RecA homologs, as well as their length differences relative to these other proteins, they are difficult to align with confidence and were therefore excluded from our alignments and phylogenetic analyses. All other RecA-like proteins from eukaryotes have sequence identities at least comparable to that observed between *Saccharomyces* Rad51 and Dmc1 (Fig. 7), and can be aligned with confidence over most of their length (Fig. 6). Therefore, these proteins and their genes were chosen as the basis of our phylogenetic study of eukaryotic RecA homologs.

All analyses also included two of the three recently determined RecA-like sequences from archaebacteria (Sandler et al. 1996). In keeping with relationships observed for virtually all aspects of the genetic apparatus (Brown and Doolittle 1995; Baldauf et al. 1996; Bult et al. 1996), these archaebacterial proteins are substantially more similar to those of eukaryotes (42–46% identical to *Saccharomyces* Rad51 and Dmc1; Fig. 7) than are those of eubacteria (22–26%; Fig. 7). Finally, in one set of analyses, three of the 65+ available eubacterial RecA sequences (Eisen 1995) were chosen to serve as an outgroup for our analyses of eukaryotic and archaebacterial RecA phylogeny. Despite the great divergence of the eubacterial sequences from all eukaryotic and archaebacterial sequences (Fig. 7), their mutual alignment (Sandler et al. 1996; Fig. 6) is rel-

Fig. 6 Amino-acid sequence alignment of selected (see text) RecA and RecA-like sequences. Amino-acid residues identical to the top sequence are indicated by "·", whereas gaps are indicated by "–". The cumulative length of each sequence is shown to the right in *parentheses.* The nucleotide-binding sites, *motifs A and B*, are marked above the alignment. The *arrowheads* above positions 29 and 341 in the *Coprinus* sequence circumscribe the portion of the alignment (excepting the short segment between the two *asterisks* in the last quadrant of the figure, which was excluded because of alignment difficulties) used for calculating percent identity (Fig. 7) and for phylogenetic analysis (Fig. 8). The numbers of terminal residues not presented for the three eubacterial sequences are given in *parentheses.* Full organism names and sources of sequences are given in Materials and methods


Rad51
Dmc1
RadA
RecA
Coprinus
Saccharomyces
Schizosacchar.
Neurospora
Homo
Mus
Gallus
Xenopus1
Xenopus2
Drosophila
Lycopersicon
Candida
Arabidopsis
Lilium
Methanococcus
Sulfolobus
Escherichia
Bacillus
Thermus
A
B

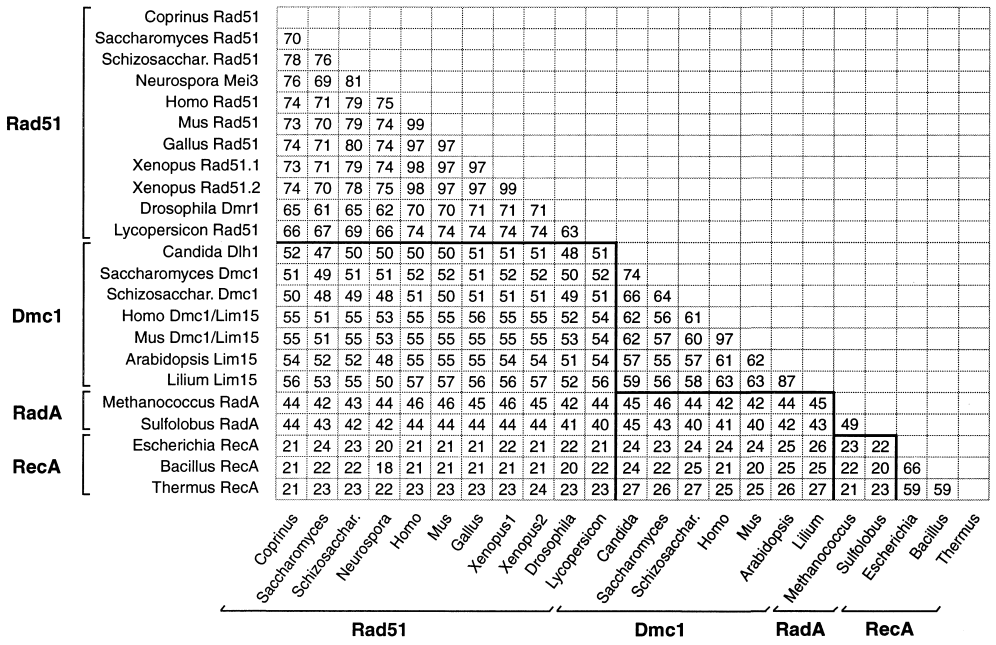


Fig. 7 Percent amino-acid identity of the 23 RecA and RecA-like proteins from Fig. 6. Values were calculated based on the alignable regions extending from amino acid 29 to 341 of the *Coprinus* sequence (Fig. 6), excluding the asterisked region in Fig. 6 (see legend) and all gaps in any pairwise comparison. The four *bracketed* protein names are those that we propose using (based on the phylogenetic results shown in Fig. 8) in place of the idiosyncratic, sequence-specific protein names (shown to the right of the column of organism names) used in the original reports on these sequences

atively unambigous compared to that of Rad55 and Rad57, which, as described above, were therefore excluded.

Phylogenetic relationships among these 23 RecA-like sequences were determined using three different tree-building methods: maximum parsimony, neighbor-joining, and maximum likelihood (see Swofford et al. 1996 for a detailed review of the principles and relative merits of these approaches). Parsimony and neighbor-joining were applied to the amino-acid data set shown in Fig. 6, while all three methods were applied to the first and second positions of a nucleotide alignment corresponding to this amino-acid alignment (maximum-likelihood analysis of amino-acid sequences is not yet computationally feasible for this many sequences). The strength of internal support for each branch on these trees was evaluated using the bootstrap re-sampling technique (Felsenstein 1985); the number on each internal branch is the number of bootstrap trees (out of 100 total) that support this grouping.

Because of the great divergence of the eubacterial sequences (Fig. 7), we carried out one set of analyses without them, in which two archaebacterial sequences were used as an outgroup to root relationships among the eukaryotic sequences (Fig. 8A). All five analytical approaches employed gave similar topologies with this data set (but see below); therefore, one representative tree (from the maximum-likelihood analysis of nucleotides) is shown in Fig. 8A. [Essentially identical phylogenetic results were also obtained in the few analyses which included all three available complete *recA* sequences from archaebacteria (Sandler et al. 1996).] Most importantly, all five approaches divided the eukaryotic sequences into two groups: 11 of these sequences, including *Saccharomyces RAD51* and the sequences newly reported here from *Coprinus* and *Lycopersicon*, are grouped into a very strongly supported clade (100% bootstrap support with all approaches), which we hereby recognize as the *rad51* (Rad51) clade. The other seven eukaryotic sequences, including *Saccharomyces DMC1*, form a group with moderate to strong support (73–98%), which we recognize as the *dmc1* (Dmc1) clade. The topology of Fig. 8A implies that a *recA* gene duplication occurred early in eukaryotic evolution, minimally in a common ancester of animals, fungi, and plants, giving rise to the *rad51* and *dmc1* genes found jointly in each of these three groups. In molecular evolutionary parlance (Li and Graur 1991), the 11 *rad51* genes are orthologous to one another, i.e., descended from a common ancestral gene by speciation, but paralogous (descended by gene duplication) to the seven *dmc1* orthologs.

It is important to realize, however, that by virtue of excluding eubacteria and rooting on archaebacteria, the analysis in Fig. 8A precludes testing the possibility that the multiple *recA* genes found in eukaryotes result instead from one or more gene duplications that actually preceded the divergence of eukaryotes and archaebacteria from a common ancestor. To address this critical question, we therefore performed analyses (Fig. 8B) that included three selected, but very divergent, eubacterial sequences as outgroups. All five analytical approaches again gave very strong bootstrap support (100%) for a clade of 11 *rad51* genes and somewhat less strong support (55–93%) for a clade of seven *dmc1* genes. Importantly, the five approaches gave moderate to strong support (71–98%) for monophyly of all 18 eukaryotic *recA* sequences relative to all prokaryotic sequences, both eubacterial and archaebacterial. Thus, among the sequences analyzed, there indeed seems to have been but a single gene duplication, leading to the *rad51* and *dmc1* clades and occurring in a common

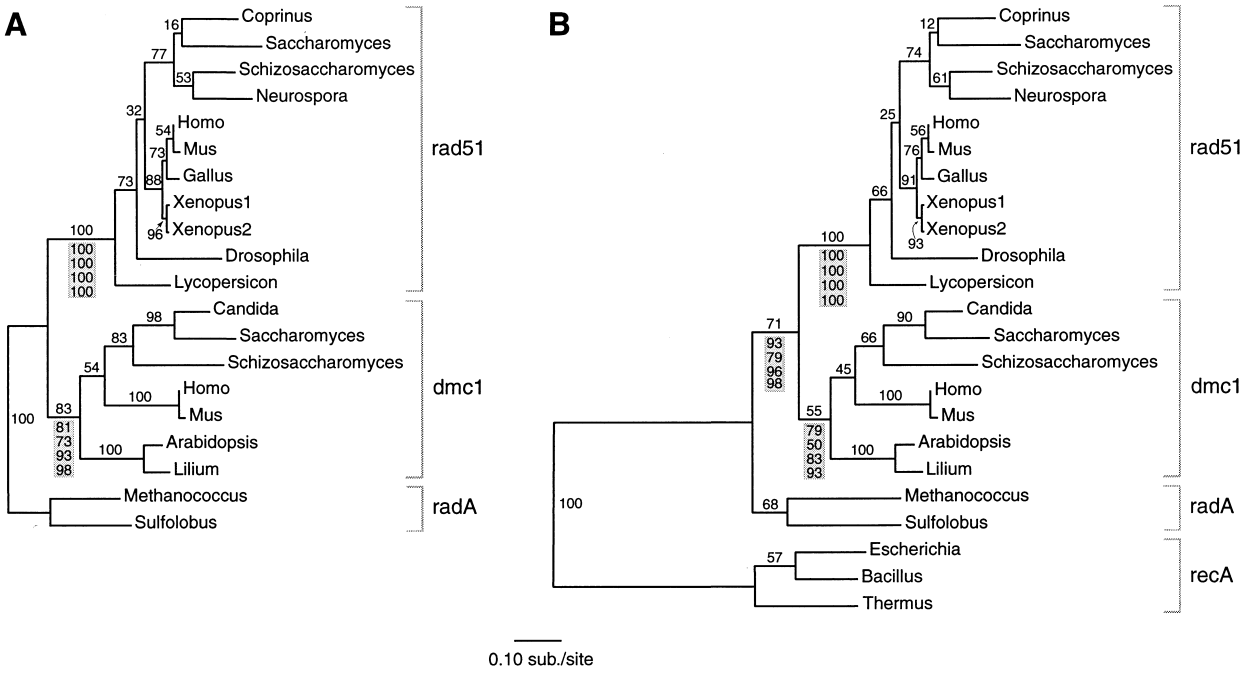


Fig. 8A, B Phylogenetic relationships of *recA* and *recA*-like sequences from maximum-likelihood analyses of first- and second-position nucleotide sequences. Shown are the highest likelihood trees from analyses that either excluded (**A**) or included (**B**) the three eubacterial sequences in the alignment of Fig. 6. In (**A**) the two archaebacterial sequences were designated as the outgroup, while in (**B**) the eubacterial sequences were the outgroup. The *numbers* above the tree branches indicate the percentage of times that the branch was recovered in 100 bootstrap samples. The *shaded numbers* below selected branches are the bootstrap values for the indicated group in, from top-to-bottom, a parsimony analysis of nucleotide characters (first and second codon positions), parsimony analysis of amino acids, neighbor-joining analysis of nucleotides, and neighbor-joining analysis of amino acids. Branch lengths are proportional to the number of inferred nucleotide substitutions per site (see scale bar)

ancestor of all eukaryotes examined thus far. It should be emphasized, however, that at present there are no *recA* sequences available from protists, which represent much of eukaryotic molecular diversity (Sogin 1991).

Finally, while the extensive sequence and length divergence of the *Saccharomyces* Rad55 and Rad57 proteins precluded their inclusion in these phylogenetic analyses, the mere existence of these two proteins constitutes *prima facie* evidence for two additional, potentially ancient, duplications in the *recA* gene family of eukaryotes. Whether these proteins are truly ancient or are simply extremely rapidly changing awaits characterization of orthologs from other organisms, especially fungi that are closely related to *Saccharomyces*.

To reduce possible biases introduced by inclusion of the highly divergent eubacterial sequences, the following discussion on relationships *within* the *rad51* and *dmc1* clades will focus entirely on results obtained using archaebacteria as the outgroup (e.g., Fig. 8A). Relationships among the 11 *rad51* sequences are relatively poorly resolved, especially in comparison to the well-established phylogeny of their respective organisms (e.g., Baldauf and Palmer 1993; Gargas et al. 1995). Only two groupings, both involving the five vertebrate sequences, are consistently well supported with all five analytical approaches used. As expected, these five sequences themselves form a well-supported monophyletic group (88% bootstrap with likelihood, 100% with the other four approaches). Also, the two *Xenopus* sequences consistently affiliate (78–98%), which implies a recent duplication of *rad51* in a *Xenopus*-specific lineage. However, *rad51* relationships among the four vertebrates, whose relationships as organisms are unequivocal, are not well resolved. Bootstrap support for monophyly of the two mammalian sequences is relatively low (e.g., Fig. 8A), and although all three nucleotide analyses recover the expected grouping of these sequences with *Gallus rad51*, both amino-acid analyses instead group the *Xenopus* and mammalian sequences to the exclusion of *Gallus*. These results are almost certainly the consequence of too few differences, expecially at the amino-acid level, among these very closely related sequences (Figs. 7 and 8).

The only other group of *rad51* sequences recovered with all five phylogenetic approaches consists of the four fungal sequences. However, this group has relatively low bootstrap support (44–82%) considering that fungi are quite clearly a monophyletic group based on many criteria, both morphological and molecular (e.g., fungal monophyly receives 91–100% bootstrap with the four different protein genes examined by Baldauf and Palmer 1993). Furthermore, relationships among the four fungi are in essence completely unresolved, varying substantially between analyses and with bootstrap support for these variable groupings always low (e.g., Fig. 8A). Although Ascomycetes (*Schizosaccharomyces*, *Saccharomyces*, *Neuro-*

spora) are clearly a monophyletic group relative to Basidiomycetes (*Coprinus*), and *Schizosacharomyces* is probably the sister group to the other two Ascomycetes (e.g., Baldauf and Palmer 1993; Gargas et al. 1995), this expected topology was never recovered.

Relationships among *Drosophila rad51*, *Lycopersicon rad51*, and the two clades consisting of the five vertebrate and four fungal *rad51* sequences are also completely unresolved, again varying among the different trees and with low bootstrap support for these variable groupings. Here again, this is despite a strong *a priori* expectation of where these sequences should group: *Drosophila* is of course an animal, yet in only two of five analyses does it even weakly (25–51% bootstrap) affiliate with the vertebrate sequences. Similarly, animals and fungi are now clearly established as sister-groups relative to plants (Baldauf and Palmer 1993; Wainwright et al. 1993; Nikoh et al. 1994), yet in only three of five analyses is this relationship recovered, with *Lycopersicon* the deepest *rad51* sequences (34–73%).

Why is the *rad51* phylogeny so poorly resolved for these ten organisms, when their relationships are so well established both on traditional and molecular criteria? Assuming taxonomic authenticity of the sequences, there are two general classes of answers: either the sequences are not entirely orthologous, owing to one or more events of lateral transfer (xenology) or gene duplication (paralogy), and thus the *rad51* gene tree should *not* mirror the organismal tree, or they are orthologous but peculiarities in their tempo and mode of sequence evolution make recovery of the expected organismal phylogeny difficult. For three reasons, we favor the latter explanation. First, evolution by xenology or paralogy will, in most cases, produce a gene tree whose topology is strongly and consistently resolved, but which differs from that of the organisms in question. This is not the case here; instead the various analyses produce conflicting but weakly supported gene trees, such that a strict consensus of these trees would show little resolution at all. Second, there is no indication that duplication has occurred (except within the *Xenopus* lineage, as already mentioned), or expectation that lateral transfer might have occurred. Lateral transfer, while moderately common among bacteria (Mazodier and Davies 1991; Syvanen 1994; Delwiche and Palmer 1996), is exceedingly rare among eukaryotes; in fact, aside from the special cases of mobile elements such as transposons and group-I and -II introns (Kidwell 1993; Lambowitz and Belfort 1993), we are unaware of any well documented cases of lateral gene transfer between and within plants, fungi, and animals. As for gene duplication, no one has yet recovered two diverse, yet still *rad51*-like, genes from any of the eukaryotes in question. If duplication has occurred, then either one gene copy has: (1) been repeatedly and differentially lost (this would have to be the case for the completely sequenced *Saccharomyces* genome), (2) diverged so much as to be no longer recognizable as a *"rad51"* gene (this could be the case for the extremely divergent *RAD55* and *RAD57* genes of *Saccharomyces*), or (3) yet to be recovered, despite intense efforts for many of these genomes. Third, as discussed in some detail in the next section, there is clear evidence for what is probably the major source of artefact in reconstructing deep phylogeny, namely, lineage-specific inequities in rates of sequence evolution (Felsenstein 1978; Kuhner and Felsenstein 1994; Palmer and Delwiche 1996; Swofford et al. 1996). Unequal-rate effects normally produce what is known as "long-branch-attraction" – the artefactual clustering of long branches – and often this is manifest by the placement of a rapidly changing gene lineage artefactually deeply in a tree. This is precisely what is seen in most of the analyses for the two sequences, from *Saccharomyces* and *Drosophila*, which most clearly show evidence of rapid evolution (see next section).

In summary, then, we believe that the poorly resolved, somewhat anomalous *rad51* phylogeny is probably the result of major inequities in rates of sequence evolution rather than events of gene duplication or lateral evolution. Whether the extent of rate heterogeneity in *rad51* evolution is sufficiently great as to seriously undermine its general utility for phylogeny reconstruction awaits more comprehensive sequencing of the gene across eukaryotic diversity.

The phylogeny of the seven *dmc1* sequences is somewhat better resolved than for the 11 *rad51* sequences. The same *dmc1* topology shown in Fig. 8A was recovered with the other four approaches used, and it non-controversially clusters (1) each of the two angiosperms, the two mammals, and the three fungi, (2) the budding yeasts *Saccharomycyces* and *Candida*, and (3) animals and fungi.

Rates of RecA evolution

At the deep divergences considered here, rates of protein-gene evolution are best evaluated at the level of amino-acid sequences (Fig. 7). These data provide abundant evidence of major inequities in RecA evolution across life's panoply. Eubacteria, which at >3.5 billion years are probably the oldest of life's three "domains", show, for the three diverse representatives included here (see Fig. 3 of Eisen 1995), relatively high sequence conservation (59–66% amino-acid identity; Fig. 7). This equals the extremes of divergence seen for Rad51 and Dmc1, which represent eukaryotic lineages thought to be no more than a third as old as eubacteria. RecA is also more conserved among eubacteria than archaebacteria (49% identity for representatives of the two archaebacterial kingdoms). The huge gulf of RecA sequence divergence [numerous major alignment gaps (Fig. 6) and on average only 22% sequence identity (Fig. 7)] between eubacteria and archaebacteria plus eukaryotes implies a period(s) of probably rapid and extensive remolding of this protein in the lineage(s) leading to one or both groups.

Within eukaryotes, it is clear that Dmc1 is evolving somewhat more rapidly than Rad51 in all cases (e.g., Dmc1 and Rad51 identities average 62% and 74%, respectively, for plants vs vertebrates, 60% and 74% for fungi vs vertebrates, 64% and 76% for *Saccharomyces* vs *Schizosaccharomyces*, and 97% and 99% for *Homo* vs *Mus*). Yet within the limits of their taxonomic overlap, the two genes show

similar patterns of within-gene rate variation: in both cases, fungi show rapid rates of change and vertebrates low rates. Rapid evolution in fungi is evident in the long branch lengths leading to fungi in Fig. 8 and by the fact that fungi are equally or less similar (70–80% for Rad51, 56–62% for Dmc1) to the slowly-evolving vertebrates than are plants (74% and 61–63%), a genealogically more-distant group (Baldauf and Palmer 1993; Wainwright et al. 1993; Nikoh et al. 1994). Within the fungi, *Saccharomyces* is notably rapidly changing; for both genes it is the most divergent fungal sequence relative to an animal outgroup sequence (Fig. 7). A slow rate of evolution in vertebrates is especially evident for Rad51, which is 97–98% identical between *Xenopus* and either mammals or *Gallus*, despite their approximately 400 million year divergence.

Comparisons that include both the rapidly evolving Rad51 of *Saccharomyces* and the slowly evolving Rad51s of vertebrates are particularly striking; here *Saccharomyces* is actually less similar to its fellow Ascomycetes *Schizosaccharomyces* (76%) and *Neurospora* (69%) than they are to vertebrates (78–80% and 75–75%)! Importantly, the same pattern, where *Schizosaccharomyces* is equally or more similar in amino-acid sequence to vertebrates than it is to *Saccharomyces*, has been seen for several other proteins (Radford and Dix 1988; Sipiczki 1989; Loppes et al. 1991; Moreno et al. 1991; Jannatipour and Rokeach 1995), an extreme example being γ-tubulin (Keeling and Logsdon 1996). This pattern, together with considerations of some uniquely derived features of *Saccharomyces* cell-cycle control, heat-shock response, and splicing, has provoked the erroneous conclusion that fission yeast is equally or more closely related to humans than it is to budding yeast, as well as the largely inappropriate conclusion that it should therefore serve as a better model organism than *Saccharomyces* (e.g., Sipiczki 1989; Fosburg and Nurse 1991; Moreno et al. 1991; see Taylor et al. 1993 for a critical appraisal of these issues), when it merely reflects a consistent pattern of evolutionary rate variation in multiple genes.

The other unusually divergent sequence is *Drosophila* Rad51. Here we are faced with a similar divergence paradox as described in the preceding paragraph: Rad51 from *Drosophila*, indisputably an animal, is actually less similar to vertebrate Rad51s (70–71% identity), than are the Rad51s from fungi (73–80%; excepting the most rapidly diverging sequence, from *Saccharomyces*) and even the genealogically more distant organism, *Lycopersicon* (74%). Since there is only 2–3% Rad51 divergence within 400 million years of vertebrate divergence, most of the 29–30% divergence that has accumulated in the approximately 600 million years of *Drosophila*/vertebrate divergence must have occurred specifically in the lineage leading to *Drosophila*. Is the exceptionally divergent *rad51* gene from *Drosophila* actually orthologous to the vertebrate *rad51* genes? We think it probably is, because this is the only *rad51*-like gene that has been isolated from *Drosophila* by two independent groups (Akaboshi et al. 1994; McKee et al. 1996). However, the presence of an unrecovered *rad51* ortholog cannot be ruled out given the Southern hybridization results of McKee et al. (1996) indicating the presence of distantly related sequences to this gene in the *Drosophila* genome.

Although evolutionary rate variation is widespread in the *recA* gene family, it is not clear that this gene is significantly more afflicted by lineage-specific rate heterogeneity than other genes. For there is gathering evidence that evolutionary rates vary substantially for most if not all genes, encoding both protein and rRNA, and in unpredictable ways (see Palmer and Delwiche 1996 and references therein). For instance, while fungi also show rapid rates of evolution for both α- and β-tubulin, *Drosophila* and vertebrates do not (Baldauf and Palmer 1993). Conversely, rRNA evolution is slower in fungi than in vertebrates and, especially, *Drosophila* (Cavalier-Smith 1993; Carmean and Crespi 1995). All of this makes us deeply skeptical of the recent study by Doolittle et al. (1996; also see Martin 1996; Mooers and Redfield 1996; Morrell 1996), in which they estimated divergence times for all of life from molecular-clock-analyses of protein-sequence data.

Conclusions

We have isolated two new *recA*-homologous genes, one from a fungus and one from a plant, both of which are orthologous to previously characterized fungal and animal *rad51* genes. Like *Saccharomyces rad51*, the *Coprinus* gene is highly induced by gamma irradiation and during meiosis. Based on phylogenetic analyses, it is clear that *recA* genes have a complex evolutionary history in eukaryotes, one characterized by a number of gene duplications and by highly unequal evolutionary rates among animal and fungal genes. A duplication early in eukaryotic evolution has given rise to a group of *rad51* orthologs on the one hand and a group of *dmc1* orthologs on the other. Two other gene duplications, of less certain vintage, are clearly implied by the existence of two highly divergent *recA*-like genes, *RAD55* and *RAD57*, in *Saccharomyces*. *dmc1* genes consistently change faster than *rad51* genes, while fungi (especially *Saccharomyces*) evolve considerably faster than vertebrates for both types of genes. Finally, the *rad51* gene has diverged to a remarkable degree in *Drosophila* compared to vertebrates. Additional sequences are needed to further clarify evolutionary relationships, the timing of gene duplication events, and rates of molecular evolution of these key genes of DNA recombination and repair.

Acknowledgements We thank Lisa Seitz and Keliang Tang for excellent technical assistance, Chuck Delwiche for valuable advice and help with the phylogenetic analysis, John Clark and Steve Sandler for providing the archaebacterial *radA* sequences in advance of their publication, Hirokazu Inoue for providing the complete *mei-3* sequence from *Neurospora crassa* in advance of publication, Akira Shinohara and Akiyo Yamazaki for providing their unpublished *Schizosaccharomyces pombe dmc1* sequence, and Doug Bishop for providing a *DMC1* clone from *Saccharomyces cerevisiae*. This work was supported by NIH grant GM43930 to M.E.Z., NIH grant GM35087 to J.D.P., a Graduate Assistance in Areas of National Need Fellowship, USDE, P200A20417 to N.Y.S., and a fellowship from the Indiana University College of Arts and Sciences to J.M.L. H.H.O. was supported by LNV stimulans grant EPS1C11.

References

Aboussekhra A, Chanet R, Adjiri A, Fabre F (1992) Semidominant suppressors of Srs2 helicase mutations of *Saccharomyces cerevisiae* map in the *RAD51* gene, whose sequence predicts a protein with similarities to procaryotic RecA proteins. Mol Cell Biol 12:3224–3234

Akaboshi E, Inoue Y, Ryo H (1994) Cloning of the cDNA and genomic DNA that correspond to the *recA*-like gene of *Drosophila melanogaster*. Jpn J Genet 69:663–670

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Aviv H, Leder P (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid cellulose. Proc Natl Acad Sci USA 69:1408–1412

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci USA 90:11558–11562

Baldauf SL, Palmer JD, Doolittle WF (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc Natl Acad Sci USA 93:7749–7754

Basile G, Aker M, Mortimer RK (1992) Nucleotide sequence and transcriptional regulation of the yeast recombinational repair gene *RAD51*. Mol Cell Biol 12:3235–3246

Bezzubova O, Shinohara A, Mueller RG, Ogawa H, Buerstedde J-M (1993) A chicken RAD51 homologue is expressed at high levels in lymphoid and reproductive organs. Nucleic Acids Res 21:1577–1580

Binninger DM, Skrzynia C, Pukkila PJ, Casselton LA (1987) DNA-mediated transformation of the basidiomycete *Coprinus cinereus*. EMBO J 6:835–840

Bishop DK (1994) RecA homologs Dmc1 and rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. Cell 79:1081–1092

Bishop DK, Park D, Xu L, Kleckner N (1992) DMC1: A meiosis-specific yeast homolog of *E. coli recA* required for recombination, synaptonemal complex formation, and cell-cycle progression. Cell 69:439–456

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GC, Blake JA, Fitzgerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Petterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk HP, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058–1073

Camerini-Otero RD, Hsieh P (1995) Homologous recombination proteins in prokaryotes and eukaryotes. Annu Rev Genet 29:509–552

Carmean D, Crespi BJ (1995) Do long branches attract flies? Nature 373:666

Cathala G, Savouret JF, Mendez B, West BL, Karin M, Martial JA, Baxter JD (1983) A method for isolation of intact translationally active ribonucleic acid. DNA 2:329–335

Cavalier-Smith T (1993) Kingdom Protozoa and its 18 phyla. Microbiol Rev 57:953–994

Cheng R, Baker TI, Cords CE, Radloff RJ (1993) *mei-3*, a recombination and repair gene of *Neurospora crassa*, encodes a RecA-like protein. Mut Res DNA repair 294:223–234

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Clark AJ, Sandler SJ (1994) Homologous genetic recombination: the pieces begin to fall into place. Crit Rev Microbiol 20:125–142

Cox MM (1993) Relating biochemistry to biology: how the recombinational repair function of RecA protein is manifested in its molecular properties. BioEssays 15:617–622

Delwiche CF, Palmer JD (1996) Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol Biol Evol 13:873–882

Diener AC, Fink GR (1996) *DHL1* is a functional *Candida albicans* homologue of the meiosis-specific gene *DMC1*. Genetics 143:769–776

Doolittle RF, Feng DF, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

Edelmann SE, Staben C (1994) A statistical analysis of sequence features within genes from *Neurospora crassa*. Exp Mycol 18:70–81

Eisen JA (1995) The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16*S* rRNAs from the same species. J Mol Evol 41:1105–1123

Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. Systematic Zool 27:401–410

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Forsburg SL, Nurse P (1991) Cell-cycle regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Annu Rev Cell Biol 7:227–256

Game J (1993) DNA double-strand breaks and the *RAD50–RAD57* genes in *Saccharomyces*. Sem Cancer Biol 4:73–83

Gargas A, DePriest PT, Grube M, Tehler A (1995) Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science 268:1492–1495

Gurr SJ, Unkles SE, Kinghorn JR (1987) The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn JR (ed) *Gene structure in eukaryotic microbes*. IRL Press, London, England, pp 93–139

Habu T, Taki T, West A, Nishimune Y, Morita T (1996) The mouse and human homologs of *DMC1*, the yeast meiosis-specific homologus recombination gene, have a common unique form of exon-skipped transcript in meiosis. Nucleic Acids Res 24:470–477

Hatakeyama S, Ishii C, Inoue H (1995) Identification and expression of the *Neurospora crassa mei-3* gene which encodes a protein homologous to Rad51 of *Saccharomyces cerevisiae*. Mol Gen Genet 249:439–446

Heyer WD (1994) The search for the right partner: homologous pairing and DNA strand-exchange proteins in eukaryotes. Experientia 50:223–233

Holm PB, Rasmussen SW, Zickler D, Lu BC, Sage J (1981) Chromosome pairing, recombination nodules and chiasma formation in the basidiomycete *Coprinus cinereus*. Carlsberg Res Commun 46:305–346

Jang YK, Jin YH, Kim EM, Fabre F, Hong SH, Park SD (1994) Cloning and sequence analysis of *rhp51+*, a *Schizosaccharomyces pombe* homolog of the *Saccharomyces cerevisiae RAD51* gene. Gene 142:207–211

Jannatipour M, Rokeach LA (1995) The *Schizosaccharomyces pombe* homologue of the chaperone calnexin is essential for viability. J Biol Chem 270:4845–4853

Keeling PJ, Logsdon JM Jr (1996) Highly divergent *Caenorhabditis* and *Saccharomyces* tubulins evolved recently from genes encoding γ-tubulin. Trends Cell Biol 6:375

Kidwell MG (1993) Lateral transfer in natural populations of eukaryotes. Annu Rev Genet 27:235–256

Kobayashi T, Kobayashi E, Sato S, Hotta Y, Miyajima N, Tanaka A, Tabata S (1994) Characterization of cDNAs induced in meiotic prophase in lily microsporocytes. DNA Res 1:15–26

Kowalczykowski SC (1991) Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. Annu Rev Biophys Chem 20:536–575

Kuhner MK, Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–468

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622

Li WH, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland, Massachusetts

Logsdon JM Jr, Tyshenko MG, Dixon C, Jafari JD, Walker VK, Palmer JD (1995) Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. Proc Natl Acad Sci USA 92:8507–8511

Loppes R, Michels R, Decroupette I, Joris B (1991) Sequence analysis of the *ARG7* gene of *Schizosaccharomyces pombe* coding for argininosuccinate lyase. Curr Genet 19:255–260

Lovett ST (1994) Sequence of the *RAD55* of *Saccharomyces cerevisiae:* similarity of RAD55 to prokaryotic RecA and other RecA-like proteins. Gene 142:103–106

Maeshima K, Morimatsu K, Shinohara A, Horii T (1995) *Rad51* homologues in *Xenopus laevis*: two distinct genes are highly expressed in ovary and testis. Gene 160:195–200

Martin WF (1996) Is something wrong with the three of life? Bioessays 18:523–527

May G, Le Chevanton L, Pukkila PJ (1991) Molecular analysis of the *Coprinus cinereus* mating-type A factor demonstrates an unexpectedly complex structure. Genetics 128:529–538

Mazodier P, Davies J (1991) Gene transfer between distantly related bacteria. Annu Rev Genet 25:147–171

McCombie WR, Adams MD, Kelley JM, FitzGerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AR, Venter JC, Fields C (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. Nature Genet 1:124–131

McKee BD, Ren XJ, Hong CS (1996) A *recA*-like gene in *Drosophila melanogaster* that is expressed at high levels in female but not male meiotic tissues. Chromosoma 104:479–488

Mooers AO, Redfield RJ (1996) Digging up the roots of life. Nature 379:587–588

Morell V (1996) Proteins 'clock' the origins of all creatures – great and small. Science 271:448

Moreno S, Klar A, Nurse P (1991) Molecular genetic analysis of fission yeast *Schizosacharomyces pombe.* Methods Enzymol 194:795–826

Morita T, Yoshimura Y, Yamamoto A, Murata K, Mori M, Yamamoto H, Matsushiro A (1993) A mouse homolog of the *Escherichia coli recA* and *Saccharomyces cerevisiae RAD51* genes. Proc Natl Acad Sci USA 90:6577–6580

Muris DFR, Vreeken K, Carr AM, Broughton BC, Lehmann AR, Lohman PHM, Pastink A (1993) Cloning the *RAD51* homologue of *Schizosaccharomyces pombe.* Nucleic Acids Res 21: 4589–4591

Nikoh N, Hayase N, Iwabe N, Kuma K, Miyata T (1994) Phylogenetic relationships of the kingdoms Animalia, Plantae and Fungi, inferred from 23 different protein sequences. Mol Biol Evol 11:762–768

Ogawa T, Shinohara A, Ogawa H, Tomizawa JI (1992) Functional structures of the RecA protein found by chimera analysis. J Mol Biol 226:651–660

Ogawa T, Yu X, Shinohara A, Egelman E (1993a) Similarity of the yeast RAD51 filament to the bacterial RecA filament. Science 259:1896–1898

Ogawa T, Shinohara A, Nabetani A, Ikeya T, Yu X, Egelman EH, Ogawa H (1993b) RecA-like recombination proteins in eukaryotes: functions and structures of *RAD51* genes. Cold Spring Harbor Symp Quant Biol 58:567–576

Palmer JD, Delwiche CF (1996) Second-hand chloroplasts and the case of the disappearing nucleus. Proc Natl Acad Sci USA 93:7432–7435

Pukkila PJ, Lu BC (1985) Silver staining of meiotic chromosomes in the fungus, *Coprinus cinereus.* Chromosoma 91:108–112

Pukkila PJ, Casselton LA (1991) Molecular genetics of the agaric *Coprinus cinereus.* In: Bennett JW, Lasure LL (eds) More gene manipulations in fungi. Academic Press Inc., San Diego, California, pp 126–150

Pukkila PJ, Yashar BM, Binninger DM (1984) Analysis of meiotic development in *Coprinus cinereus.* In: Evans CW, Dickinson HG (eds) Controlling events in meiosis. Society for Experimental Biology, Cambridge, England, pp 177–194

Radford A, Dix NIM (1988) Comparison of the orotidine 5′-monophosphate decarboxylase sequences of eight species. Genome 30:501–505

Raju NB (1972) Timing of meiotic events in *Coprinus.* PhD thesis, University of Guelph, Canada

Rao PS, Niederpruem DJ (1969) Carbohydrate metabolism during morphogenesis of *Coprinus lagopus* (sensu Buller). J Bacteriol 100:1222–1228

Roca AI, Cox MM (1990) The RecA protein: structure and function. Crit Rev Biochem Mol Biol 25:415–456

Rymond BC, Rosbash M (1992) Yeast pre-mRNA splicing. In: Jones EW, Pringle JR, Broach JR (eds) The molecular and cellular biology of the yeast *Saccharomyces*: gene expression, 1st edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 143–192

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Springs Harbor Laboratory, Cold Springs Harbor, New York

Sandler SJ, Satin LH, Samra HS, Clark AJ (1996) *recA*-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae.* Nucleic Acids Res 24:2152–2132

Sato S, Seki N, Hotta Y, Tabata S (1995a) Expression profiles of a human gene identified as a structural homologue of meiosis-specific *recA*-like genes. DNA Res 2:183–186

Sato S, Kobayashi T, Hotta Y, Tabata S (1995b) Characterization of a mouse *recA*-like gene specifically expressed in testis. DNA Res 2:147–150

Sato S, Hotta Y, Tabata S (1995c) Structural analysis of a *recA*-like gene in the genome of *Arabidopsis thaliana.* DNA Res 2:89–93

Seitz LC, Tang K, Cummings WJ, Zolan ME (1996) The *rad9* gene of *Coprinus cinereus* encodes a proline-rich protein required for meiotic chromosome condensation and synapsis. Genetics 142:1105–1117

Shinohara A, Ogawa H, Ogawa T (1992) Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. Cell 69:457–470

Shinohara A, Ogawa H, Matsuda Y, Ushio N, Ikeo K, Owaga T (1993) Cloning of human, mouse and fission yeast recombination genes. Nature Genet 4:239–243

Sipiczki M (1989) Taxonomy and phylogenesis. In: Nasim A, Young P, Johnson BF (eds) Molecular biology of the fission yeast. Academic Press, San Diego, pp 431–452

Skryzynia C, Binninger DM, Alspaugh JAI, Pukkila PJ (1989) Molecular characterization of *TRP1*, a gene coding for tryptophan synthetase in the basidiomycete *Coprinus cinereus.* Gene 81:73–82

Sogin ML (1991) Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1:457–463

Sung P (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science 265:1241–1243

Sung P, Robberson DL (1995) DNA strand exchange mediated by a RAD51-SSDNA nucleoprotein filament with polarity opposite to that of recA. Cell 82:453–461

Swofford DL (1996) PAUP: phylogenetic analysis using parsimony. Version 4.od47

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics, 2nd edn. Sinauer, Sunderland, Massachusetts, pp 407–514

Syvanen M (1994) Horizontal gene transfer: evidence and possible consequences. Annu Rev Genet 28:237–261

Taylor JW, Bowman BH, Berbee ML, White TJ (1993) Fungal model organisms: phylogenetics of *Saccharomyces*, *Aspergillus*, and *Neurospora.* Systematic Biol 42:440–457

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple-sequence alignment through sequence weighting, position-specific gap penalties and weight-matrix choice. Nucleic Acids Res 22:4673–4680

Tymon AM, Kues U, Richardson WV, Casselton LA (1992) A fungal mating-type protein that regulates sexual and asexual development contains a POU-related domain. EMBO J 11:1805–1813

Valentine G, Wallace YJ, Turner FR, Zolan ME (1995) Pathway analysis of radiation-sensitive meiotic mutants of *Coprinus cinereus.* Mol Gen Genet 247:169–179

Wainwright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260:340–342

West SC (1992) Enzymes and molecular mechanisms of genetic recombination. Annu Rev Biochem 61:603–640

Wu MMJ, Cassidy JR, Pukkila PJ (1983) Polymorphisms in DNA of *Coprinus cinereus.* Curr Genet 7:385–392

Yoshimura Y, Morita T, Yamamoto A, Matsushiro A (1993) Cloning and sequence of the human RecA-like gene cDNA. Nucleic Acids Res 21:1665

Zolan ME, Pukkila PJ (1986) Inheritance of DNA methylation in *Coprinus cinereus.* Mol Cell Biol 6:195–200

Zolan ME, Tremel CJ, Pukkila PJ (1988) Production and characterization of radiation-sensitive meiotic mutants of *Coprinus cinereus*. Genetics 120:379–387

Zolan ME, Crittenden JR, Heyler NK, Seitz LC (1992) Efficient isolation and mapping of *rad* genes of the fungus *Coprinus cinereus* using chromosome-specific libraries. Nucleic Acids Res 20:3993–3999

Zolan ME, Heyler NK, Ramesh MA (1993) Gene mapping using marker chromosomes in *Coprinus cinereus.* In: Baltz RH, Hegeman GD, Skatrud PL (eds) Industrial microorganisms: basic and applied molecular genetics. American Society for Microbiology, Washington, pp 31–35

Zolan ME, Yeager Stassen N, Ramesh MA, Valentine G (1995) Meiotic mutants and DNA repair genes of *Coprinus cinereus.* Can J Bot 73:S226–S233