Evolution of the RECQ Family of Helicases: A Drosophila Homolog, *Dmblm*, Is Similar to the Human Bloom Syndrome Gene

Kohji Kusano,* Mark E. Berres† and William R. Engels*

**Laboratory of Genetics and †Laboratory of Zoology, University of Wisconsin, Madison, Wisconsin 53706*

Manuscript received June 17, 1998
Accepted for publication November 10, 1998

ABSTRACT

Several eukaryotic homologs of the *Escherichia coli* RecQ DNA helicase have been found. These include the human *BLM* gene, whose mutation results in Bloom syndrome, and the human *WRN* gene, whose mutation leads to Werner syndrome resembling premature aging. We cloned a *Drosophila melanogaster* homolog of the RECQ helicase family, *Dmblm* (***D****rosophila* ***m****elanogaster* ***Bl****oo****m***), which encodes a putative 1487-amino-acid protein. Phylogenetic and dot plot analyses for the RECQ family, including 10 eukaryotic and 3 prokaryotic genes, indicate *Dmblm* is most closely related to the *Homo sapiens BLM* gene, suggesting functional similarity. Also, we found that *Dmblm* cDNA partially rescued the sensitivity to methyl methanesulfonate of *Saccharomyces cerevisiae sgs1* mutant, demonstrating the presence of a functional similarity between *Dmblm* and *SGS1*. Our analyses identify four possible subfamilies in the RECQ family: (1) the BLM subgroup (*H. sapiens Bloom*, *D. melanogaster Dmblm*, and *Caenorhabditis elegans* T04A11.6); (2) the yeast RECQ subgroup (*S. cerevisiae SGS1* and *Schizosaccharomyces pombe rqh1/rad12*); (3) the RECQL/Q1 subgroup (*H. sapiens RECQL/Q1* and *C. elegans* K02F3.1); and (4) the WRN subgroup (*H. sapiens Werner* and *C. elegans* F18C5.2). This result may indicate that metazoans hold at least three RECQ genes, each of which may have a different function, and that multiple RECQ genes diverged with the generation of multicellular organisms. We propose that invertebrates such as nematodes and insects are useful as model systems of human genetic diseases.

DNA/RNA helicase protein families have been found with seven consensus motifs (Gorbalenya *et al.* 1989). These enzymes play important roles in cellular processes such as genome replication, recombination and repair, transcription, and mRNA translation. Members within each helicase family also share consensus sequences between motifs. The RECQ helicase family includes proteins with extensive amino acid sequence similarity to the *Escherichia coli* helicase, RecQ.

The *E. coli RecQ* gene encodes a DNA helicase (Umezu *et al.* 1990) involved in homologous recombination (Nakayama *et al.* 1984). The RecQ helicase appears to interact with RecJ exonuclease because *recQ* null mutations suppress *recJ* defects such as decreased crossing over (Kusano *et al.* 1994), increased sensitivity to DNA-damaging agents (Kusano *et al.* 1994; Lovett and Sutera 1995), and decreased conjugational recombination (Lovett and Sutera 1995) in a $recB^{-}C^{-}$ $sbcA^{-}$ background of *E. coli*. A *recQ* null mutation increases illegitimate recombination in a wild-type background of *E. coli* (Hanada *et al.* 1997). The *Saccharomyces cerevisiae SGS1* gene encodes similar helicase motifs (Gangloff *et al.* 1994; Watt *et al.* 1995) and possesses a 3′-5′ DNA helicase activity (Lu *et al.* 1996; Bennett *et al.* 1998). An *sgs1* mutation suppresses slow growth in a *top3* mutant (Gangloff *et al.* 1994; Lu *et al.* 1996) and causes missegregation of chromosomes in meiosis and mitosis (Watt *et al.* 1995). The SGS1 protein physically interacts with topoisomerase III (Gangloff *et al.* 1994) and topoisomerase II (Watt *et al.* 1995). The mutations of another RECQ family member, *Schizosaccharomyces pombe rqh1/rad12*, cause increased mitotic recombination with hydroxyurea treatment (Stewart *et al.* 1997). *S. pombe rqh1/rad12* appears to function in a checkpoint dependent DNA damage response during S phase (Murray *et al.* 1997; Stewart *et al.* 1997; Davey *et al.* 1998). Three homologs of RECQ have been found in humans. One is the *BLM* gene whose defects cause immunodeficiencies and a highly increased rate of cancer, called Bloom syndrome (Ellis *et al.* 1995). Somatic cells from Bloom patients are susceptible to several DNA-damaging agents (Krepinsky *et al.* 1979; Ishizaki *et al.* 1981; Kurihara *et al.* 1987) and exhibit increased interchanges between homologous chromosomes (German 1993) and increased sister chromatid exchanges (Heartlein *et al.* 1987; Kurihara *et al.* 1987). The Bloom gene product has a 3′-5′ DNA helicase activity (Karow *et al.* 1997). The second homolog is the *WRN* gene whose defects result in clinical symptoms resembling premature aging, called Werner syndrome (Yu *et al.* 1996). Cells from Werner patients show various chromosome rearrangements (Salk *et al.* 1981; Scappaticci *et al.* 1982). More recently, *WRN* cells were shown to be hypersensitive to 4-nitro-quinoline-1-oxide (Ogburn *et al.* 1997). The Werner protein has DNA helicase activ-

Corresponding author: William R. Engels, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706.
E-mail: wrengels@facstaff.wisc.edu

ity (Gray *et al.* 1997; Suzuki *et al.* 1997) with 3′-5′ directionality (Shen *et al.* 1998). The third human RecQ homolog is RECQL/Q1, which has also been demonstrated to possess a 3′-5′ DNA helicase activity (Puranam and Blackshear 1994; Seki *et al.* 1994). The symptoms of RECQL/Q1 deficiency have not been determined.

Phylogenetic analysis of a protein family often reveals the presence of subfamilies that are likely to have functional similarities. For example, an analysis of MutS-related proteins shows that there are several distinct subgroups (Fishel *et al.* 1993), each corresponding to a different function: mitochondrial mismatch repair, DNA sequence stability, and meiotic crossing over. The family of SNF2-related proteins provides another example. Phylogenetic analysis of this family suggests that there are eight distinct subgroups (Eisen *et al.* 1995), each corresponding to different functions, including transcription activation, transcription-coupled repair, recombinational repair, and removal of TATA binding protein from DNA. These two phylogenetic analyses indicate that the gene families functionally diverged with the generation of unicellular eukaryotes such as yeast.

In this work, we cloned a cDNA of the *Drosophila melanogaster* RECQ homolog, *Dmblm* (***D****rosophila* ***m****elanogaster* ***Bl****oo****m***). It encodes a putative 1487 amino acid protein (Figure 1) and exhibits significant similarity to other RECQ family members. Our phylogenetic analysis of the RECQ family including *Dmblm* leads us to propose that the currently recognized RECQ family contains four subfamilies and that they functionally diverged with the generation of multicellular organisms. We suggest that *Dmblm* is included in the same group with Human *BLM* and *Caenorhabditis elegans* T04A11.6, each of which appears to be related to *S. pombe rqh1/rad12* and *S. cerevisiae SGS1*. Also, we found that *Dmblm* partially rescued the sensitivity to methyl methanesulfonate (MMS) of a *S. cerevisiae sgs1* mutant.

MATERIALS AND METHODS

Yeast strains and media: Yeast strain MRQ966 (from T. Enomoto) is an *sgs1::URA3* derivative of strain MR966 (*MAT***a** ura3-52 leu2-3, 112 trp-289 his1-7). The following media were used for yeast growth: YEPD, which contains 1% yeast extract, 2% peptone, and 2% glucose, and SD, which is the selective medium for strains with a plasmid carrying a TRP1 marker and contains 2% glucose, 0.67% Bacto-yeast nitrogen base, and supplements of uracil and all amino acids except tryptophan.

Plasmids: pYCp1305 (from T. Enomoto) contains the entire *SGS1* gene of *S. cerevisiae*, a derivative of pRS314 that includes a centromere element, and an ARS and TRP1 marker (Sikorski and Hieter 1989). Two plasmids were used for the expression of *Dmblm* cDNA in yeast: (1) The plasmid pAS2-1 (CLONTECH, Palo Alto, CA) contains a GAL4-DNA-binding domain under an ADH1 promoter, 2-μm origin, and a TRP1 marker. (2) The plasmid pGBT9-bd (from A. Shimamoto), a derivative of pGBT9 (CLONTECH), lacks a GAL4-DNA-binding domain but contains an ADH1 promoter, 2-μm origin, and a TRP1 marker. The deletion of the domain was made by replacing the region from the *Hin*dIII site (nucleotide 409) to the *Eco*RI site (nucleotide 878) of pGBT9 (GenBank accession number U07646) with *Eco*RI oligonucleotides, GGAATTCC (New England Biolabs, Beverly, MA). The plasmid pBSDMBLM1 contains a PCR product that spans the entire region of *Dmblm* cDNA, nucleotides 1–4725 (GenBank accession number U92536), at the *Eco*RV site of Bluescript SK(+). The next two plasmids contain the fragment that spans the region from the *Dra*I site (nucleotide 111), at 15 bp before the putative start codon, to nucleotide 4716 on *Dmblm* cDNA. The fragment was prepared by partially cutting pBSDMBLM DNA with *Dra*I restriction enzyme, ligating *Bam*HI oligonucleotides CCGGA TCCGG (New England Biolabs) to the resultant *Dra*I end and completely cutting with *Bam*HI restriction enzyme. The plasmids pYEDMBLM1 and pYEDMBLM2 possess this fragment at the *Bam*HI site downstream of the GAL4-DNA-binding domain under the ADH1 promoter on pAS2-1, and at the *Bam*HI site downstream of the ADH1 promoter on pGBT9-bd, respectively.

PCR cloning: Genomic DNA for PCR was prepared from the Canton-S strain of *D. melanogaster*. Two degenerate oligonucleotides, sense, ATGCCNACYGGHGGHGGHAA, and antisense, GGRCGGAARTCRTGDCCCCA, were used for the PCR and cloning of a fragment of the *D. melanogaster* RECQ homolog. Two gene specific primers based on the nucleotide sequence of the obtained fragment were used for 5′ RACE and 3′ RACE: GGTGTCACGTAGAGCAATTTGACCATAGG and GCCTGCCATTCTGACCGAGGGAGTG, respectively. A cDNA library was constructed using total RNA extracted by the RNAzol B method (Chomczynski and Sacchi 1987) from Schneider II cells for the degenerate PCR, 5′ RACE and 3′ RACE. A CLONTECH Marathon cDNA amplification kit was used for constructing the cDNA library and for 5′ and 3′ RACE. The 5′ and 3′ end primers used for amplifying an entire *Dmblm* cDNA were ATTGTTTCCATTGGCATATTGCG and CTGAGCACATTGCTCATACAG, respectively. A mixture of Stratagene (La Jolla, CA) Taq DNA polymerase and Stratagene Extender was used for the amplification of the 5′ RACE, 3′ RACEs, and the entire cDNA. PCR products were always cloned into Bluescript SK(+).

Nucleotide sequencing: The M13-20 primer (GTAAAAC GACGGCCAGT) or the T3 primer (AATTAACCCTCACTA AAGGG) was used for two types of sequencing reactions: the dideoxy method and the Dye deoxy termination method (Applied Biosystems International, Foster City, CA). The Promega (Madison, WI) Erase-A-base kit was used for preparing plasmid DNA with nested deletions in the insert region.

Additional sequences: Sequences other than *D. melanogaster Dmblm* (D.m.BLM), the abbreviations for their helicase domain, and their accession numbers are as follows: *Homo sapiens Bloom* (H.s.BLM), U39817 (Ellis *et al.* 1995); *H. sapiens RECQL/Q1* (H.s.RECQL/Q1), L36140 (Puranam and Blackshear 1994); *H. sapiens Werner* (H.s.WRN), L76937 (Yu *et al.* 1996); *C. elegans* E03A3.2 (C.e.E03A3.2), Z38112; *C. elegans* K02F3.1 (C.e.K02F3.1), U00052 (Wilson *et al.* 1994); *C. elegans* F18C5.2 (C.e.F18C5.2), U29097 (Wilson *et al.* 1994); *C. elegans* T04A11.6 (C.e.T04A11.6), Z83123; *S. pombe rqh1/rad12* (S.p.rqh1/rad12), Z54354 (Murray *et al.* 1997; Stewart *et al.* 1997); *S. cerevisiae SGS1* (S.c.SGS1), L07870; *E. coli RecQ* (E.c.RecQ), M30198 (Irino *et al.* 1986); *Haemophilus influenzae RECQ* (H.i.RECQ), U32756 (Fleischmann *et al.* 1995); *Bacillus subtilus RECQ* (B.s.RECQ), L47648 (Roels *et al.* 1992; Sorokin *et al.* 1993, 1996).

Intron boundaries: We noted that the putative amino acid sequence in U00052 (C.e.K02F3.1) did not contain the region around helicase motif VI and that the putative amino acid sequence in U29097 (C.e.F18C5.2) did not contain the region

around helicase motifs V and VI. We checked each amino acid sequence of the two published nucleotide sequences in three different reading frames and found that motif V and VI of C.e.F18C5.2 were present behind putative introns in U29097 and that motif VI of C.e.K02F3.1 was present in another reading frame in U00052. Therefore, we adjusted the exon-intron boundaries in C.e.K02F3.1 and C.e.F18C5.2 to include motifs V and VI. The putative introns start at GT and stop at AG. The nucleotide sequence of C.e.F18C5.2 from U29097 contains the region, nucleotide numbers 10192–10497, which was regarded as an intron, and the region, 10498–10580, which was regarded as an exon. The region 10192–10580 was modified as follows: 10192–10239 (intron), 10240–10427 (exon), 10428–10475 (intron), and 10476–10580 (exon). According to these modifications, amino acid positions 302–398 in C.e.F18C5.2, with reference to the coordinates of Figure 2, were used instead of QKEEVVENLT, which was part of the amino acid sequence in U29097. Also, the sequence of C.e.K02F3.1, derived from U00052, contains the regions 34759–35328 (exon), 35329–35532 (intron), 35533–35701 (exon), and 35702 through the end of this clone as part of an intron. These regions were modified as follows: 34759–35221 (exon), 35222–35532 (intron), 35533–35779 (exon), and 35780 through the end (intron). According to these modifications, amino acid positions 339 through the last site of C.e.K02F3.1 in Figure 2 were used instead of the last 33 amino acid residues in U00052.

Sequence alignment: An alignment of the amino acid sequence data in the helicase domains of the above proteins was constructed using clustering algorithms incorporated into the program CLUSTALW version 1.73 (Higgins *et al.* 1992). Parameters were set as follows: gap penalty is 10.0; gap extension penalty is 0.05; and protein weight matrix is BLOSUM30 (Henikoff and Henikoff 1994). Further visual editing of the algorithmically aligned amino acid sequences was not attempted. We used the protein alignment as a guide to exclude highly variable regions that exhibited many insertion and deletion events from the final alignment to eliminate ambiguous residue positioning (Figure 2). Corresponding codon positions in the nucleotide sequence data were deleted and the resulting in-frame nucleotide alignment was used for all subsequent analyses.

Phylogenetic analysis: The skewness of a random sample of all tree-length distributions based on the g_1 statistic (Sokal and Rohlf 1981) as described by Hillis (1991), Hillis and Huelsenbeck (1992), and Huelsenbeck (1991) was estimated using PAUP version 3.1.1 (Swofford 1993). Initial tests for significant variance in nucleotide and codon substitution rates based on maximum likelihood estimates were performed according to Muse and Gaut (1994) and Muse and Weir (1992). PAUP version 3.1.1 (Swofford 1993) was used for an initial heuristic tree search applying the TBR option for branch swapping. The F84 substitution model (Thorne *et al.* 1992) incorporated into PAML version 1.3 (Yang 1996b) was used to estimate substitution parameters and the maximum likelihood of all tree topologies.

Other analyses: Dot plot analysis for protein was performed using MegAlign software (DNASTAR, Madison, WI). Cytological mapping of the *Dmblm* locus was carried out by *in situ* hybridization to polytene chromosomes (Engels *et al.* 1986).

RESULTS

Cloning a Drosophila homolog of RECQ: To obtain a fragment of a Drosophila RECQ homolog, we carried out PCR on genomic DNA of *D. melanogaster* using degenerate primers. The primers (see materials and methods) correspond to the amino acid residues MPTGGGK, located within the conserved helicase motif I, and WGHDFRP, located between motifs II and III (Figure 1), which is conserved within the RECQ helicase family (Figure 2). Resulting amplicons were cloned and sequenced. One clone was 426 bp in length and included a putative 63-bp intron whose 5′- and 3′-splice sites were ATgtgagtt and ttacttttttaacagAT, respectively. These are in general agreement with the splice donor and acceptor consensus sequences of Drosophila short introns, AGgt(a/g)agt(a/t) and ttttt(c/t)(c/t)(c/t)(c/t)tncag(A/G)T (Mount *et al.* 1992). The predicted amino acid sequence from the 363-bp region without the intron sequence indicated good homology with the corresponding region of other RECQ genes. Also, PCR on cDNA isolated from Schneider II cells using the same primer pair produced an ~360-bp fragment. This was judged as one related to the clone from genomic DNA based on digestion patterns by *Hae*III and *Mbo*I restriction enzymes.

Using the RACE cloning method with the two internal primers (materials and methods), designed on the basis of the nucleotide sequence of the 363-bp region, a 2.6-kb 5′ RACE clone and a 2.4-kb 3′ RACE clone (the largest fragments) were cloned.

Cytological localization: Each of the DNA fragments hybridized within the cytological map region 86F1-8. Also, PCR of a P1 clone carrying the region 86F1-4 (BDGP, Hartl *et al.* 1994) using the above internal primer pair produced the expected PCR fragment size.

Predicted 1487-amino-acid protein is included in RECQ helicase family: Sequencing of the two overlapping RACE fragments revealed an open reading frame of 1487 amino acids in length. The putative 5′- and 3′-untranslated regions contain two and six stop codons in the same frame, respectively. The predicted protein sequence is shown in Figure 1. The central region contains seven motifs commonly found in DNA/RNA helicases, including a putative ATP-binding site (motif I) and a DExH box (motif II; Gorbalenya *et al.* 1989). An alignment of this protein to other members of the RECQ family shows significant similarity among them, especially in the residues of the seven helicase motifs (Figure 2). Therefore, we conclude that this Drosophila gene is a member of the RECQ family.

Sequence alignment: The schematic diagram in Figure 3 represents the position of each helicase domain in the 13 RECQ family proteins. The algorithmically aligned protein sequences of the 13 helicase domains formed a consensus length of 419 sites of which 70 sites could not be unambiguously aligned (Figure 2). Thorne *et al.* have shown that the true alignment between a set of sequences tends to exhibit fewer gaps than an algorithmically generated alignment (Thorne *et al.* 1991, 1992; Thorne and Kishino 1992). As a result, assignment of any residue to a particular position in an area of introduced gaps would be arbitrary because of

```
MSKKPVAQRKQLTLSSFIGLDGNSQSQPKSRAASVRSKPPAVYNPIFLDA   50
SSSDDETTEISSQSNNGTIATKKSSRDPRTAKLKKHTYLDLSVSPLAELS  100
AKKYARDSPPKPTSLDLSVSPLAELPAKKSDRDPPPKPVQNENSYTYRGL  150
SESPVENKSIGDTLRKPPPKERKTSIVWLSDSPEKKVTQNERKILDSPLQ  200
RFSFEDFPNKENGNRHHLLTLSDSPPPPQPVKKPEKTMWQNETKTIQDKD  250
SPANPLVSNNLASISTLLDSSRAPNTYKGSSRNLFEDSPEKSGSGEQGNK  300
LGSAKENEIPTKPATASLERNSVTSSPSPAAPLKPRYSVAFDNSLADYLK  350
DLAQNDNFSIDPNKQNTETLKSTLGFFRNTYVELMEKYCSLIDQIPAMHF  400
NEIAGFQPNTFLKLKVVRQKFKARTQLVQNSLDKKESQLKAEQEALEKEE  450
IEMQAEQARQTVLSSSSPEKCRPIMPLPKVQEIKDEKIPNRNQLIPDLCG  500
EPDNFSPPSSPRDTQLIPKRQQLINDLCGEPDDFSPPSKQNDPHLLRKCE  550
ELVHDLCEEPDDYLAQSMMLDGDLEEEQLNGPTQGTTTSGMDDGEDDLEG  600
LLAEIEDEHQKMQARRSEFNGYSYKELEAVKVKEKHKETPINISLDDDGF  650
PEYDEAMFEQMHSQAAANKSRVSSAGPSTSKSVVPTKQTSALHSQKLSGN  700
FHANVHNDGITGEFDGQKFEHSTRLMHGLSYSFGLKSFRPNQLQVINATL  750
LGNDCFVLMPTGGGKSLCYQLPAILTEGVTIVISPLKSLIFDQINKLASL  800
          I                    Ia
DICAESLSGEQKMADVMAIYRDLESQPPMVKLLYVTPEKISSSARFQDTL  850
DTLNSNNYISRFVIDEAHCVSQWGHDFRPDYKKLGVLKKRFPNVPTIALT  900
              II
ATATPRVRLDILAQLNLKNCKWFLSSFNRSNLRYRVLPKKGVSTLDDISR  950
III
YIRSKPQHFSGIIYCLSRKECDETSKKMCKDGVRAVSYHAGLTDTDRESR 1000
                  IV
QKDWLTGKMRVICATVAFGMGIDKPDVRFVLHYSLPKSIEGYYQEAGRAG 1050
          V                          VI
RDGDVADCILYYNYSDMLRIKKMLDSDKALQYNVKKIHVDNLYRIVGYCE 1100
NLTDCRRAQQLDYFGGHFTSEQCLENRETACDNCINKRAYKAVDALEHAR 1150
KAARAVKDLCSGRSRFTLLHIADVLKGSKIKKIIDFNHHKTPHHGVLKDW 1200
DKNDVHRLLRKMVIDGFLREDLIFTNDFPQAYLYLGNNISKLMEGTPNFE 1250
FAVTKNAKEAKAAVGSVSDGATSSTADGQSGMREIHERCYTDLLDLCRTI 1300
ASQRNVTMASIMNIQALKSMAETLPITEKDMCSIPHVTKANFDKYGAKLL 1350
EITSNYASEKLLMQAVLDEEEEQAAAKQRPSTSGWNNESVDWDMAVASQG 1400
NANTSGASGFNSFRAGKRKKIYKSGASKRYKTSTTSPAARKTTSARGRGG 1450
RAGAKRAESSASSASGWKSKKTGNSFGFDLMPLPGSK 1487
```

Figure 1.—Predicted protein sequence of the *D. melanogaster* homolog of RECQ gene, *Dmblm* (Accession number U92536). Helicase motifs I–VI are underlined by dashes, and their locations are based on previously described helicase domain alignment (Gorbalenya *et al.* 1989). The repeat near the N-terminal end is underlined. The positions of amino acid substitution in the repeat are double-underlined.

a lack of knowledge concerning prior insertion and deletion events. Therefore, areas with multiple floating gaps were omitted following a rule that usable sites must be within anchored regions of two unambiguously aligned sites. Using these criteria, the final amino acid alignment consisted of 349 aligned residues, of which 61 were constant and 288 were variable. Deletion of the corresponding ambiguously aligned codon positions in the nucleotide sequence data produced an alignment containing a total of 1047 nucleotide sites, of which 164 were constant and 883 were variable.

Skewness: On the basis of a statistical measure of skewness called the g_1 statistic, Hillis (1991) devised a method for testing whether or not a particular data set contains phylogenetic information. One prediction of the model is that for data sets containing phylogenetic information, the set of all possible topologies (or a subsample thereof) arranged in order of increasing length will have a definite skew to the right. This skewing is attributed to character correlation among sequences with a common phylogenetic history (Hillis 1991). The tree-length distribution analysis of 1×10^8 random to-

pologies from a possible 8.64×10^8 unrooted topologies of our nucleotide sequence data was highly skewed ($g_1 = -0.5598$, $P \ll 0.01$).

Evolutionary rates: In most molecular sequences, substitution rates across sites are not equally distributed. Uzzell and Corbin (1971) have shown that rate differences may be best fitted to a gamma distribution. The shape parameter, α, of the gamma distribution, is now used as a measure of substitution heterogeneity across sites (Yang 1996a). The importance of considering substitution heterogeneity across sites in phylogenetic reconstructions is established (Nei 1987, 1991; Yang 1994b). To test the nucleotide sequence data for significant substitution heterogeneity under a particular evolutionary model, we employed a likelihood ratio test (Huelsenbeck and Rannala 1997). Our null hypothesis was that substitution homogeneity existed across each codon position. Our alternative hypothesis was that substitution heterogeneity existed across each codon position. The likelihood ratio statistic is two times the difference in log likelihood of the model and is approximately χ^2 distributed with one degree of freedom. Substitution heterogeneity was accommodated under the F84 model (Thorne *et al.* 1992) by applying the discrete gamma model (F84-dΓ; Yang 1994b), which approximates the gamma distribution by allowing different rate categories for each nucleotide position, each with equal probability of occurrence. Our analysis was performed with four rate categories, which provides a good approximation of the gamma distribution (Yang 1994b). It is known that substitution heterogeneity is overestimated if a grossly incorrect tree topology (*e.g.*, a completely multifurcating star tree) is used (Sullivan *et al.* 1996). Therefore, we used the single most parsimonious tree derived from the nucleotide sequence data (MP1, Figure 4A). When the F84-dΓ model was applied to our data, substitution rates were gamma distributed across the three codon positions with α-values of 0.73, 0.61, and 0.65, corresponding to the first, second, and third codon positions, respectively. The likelihood ratio test indicates clearly that accounting for substitution heterogeneity across codon sites improves significantly the likelihood of the maximum parsimony topology (Table 1).

Significant differences in substitution rates across lineages are known to impair tree estimation methods (Felsenstein 1978; Huelsenbeck and Hillis 1993). We initially applied relative rate tests (Wu and Li 1985; Muse and Weir 1992; Muse and Gaut 1994) to determine whether both nucleotides (transition and transversion substitutions) and codons (synonymous and nonsynonymous substitutions) evolve in a clocklike fashion among the helicase sequences examined. This test consisted of a pairwise comparison of all 13 helicase domains using the RECQ sequence from *E. coli* as an outgroup to obtain an overall estimation of rate variances. Of 78 individual tests, 32 nucleotide and 24 codon sequence comparisons exhibited a significant deviation ($P < 0.05$) from the expectation that 2 test sequences should show an equal amount of divergence compared to the reference sequence. To test for this effect under the F84 substitution model, we again used a likelihood ratio test. Our null hypothesis was that rate homogeneity existed across all lineages. The alternative was that rate heterogeneity existed across one or more lineages. In this case, the likelihood ratio statistic is two times the difference in log likelihood of the model and is approximately χ^2 distributed with two degrees of freedom. This test, when applied to our data, easily rejects the presence of a molecular clock with high significance (Table 1).

Phylogenetic analysis: In maximum likelihood methods, probabilities for possible topologies are computed using all nucleotide sites, each of which is considered separately. The evaluation of the likelihood that a particular topology exhibits is conditional on the given evolutionary model and how well the model can explain the observed data. The topology with the highest likelihood is thus considered the best representation of relationship. Our estimation of the best topology was guided by the objective of accounting for parameters that have a significant bearing on tree reconstruction efficiency, such as rate heterogeneity and transition and transversion biases. We used the maximum parsimony method to construct an initial topology for the purposes of relevant parameter estimation. It is known that maximum parsimony often fails when significant rate variation exists across lineages (Felsenstein 1978; Nei 1991). However, maximum parsimony has the advantage of requiring no explicit evolutionary model. Under these assumptions, maximum parsimony recovers a single most parsimonious tree (MP1, Table 1; Figure 4A), with length 4157 and a consistency index of 0.475. MP1 was then used as the starting topology for a maximum likelihood tree search assuming the F84-dΓ model of sequence evolution, which allows for rate heterogeneity across sites, unequal base composition, and different transition and transversion rates (Yang 1994a). This tree search resulted in a best maximum likelihood tree of -14311.24 with transition/transversion rate ratio parameters of 0.58, 0.22, and 10.52 and α-parameters of 0.71, 0.61, and 0.63 for each codon position respectively (Figure 4B).

Similarities outside the helicase domain: A different kind of analysis was needed to determine whether the gene groupings indicated above are also apparent in sequences flanking the helicase domain. In these areas, the similarities were too weak to use the phylogenetic methods employed for the helicase domain. Therefore, we examined dot plots of all 78 pairwise comparisons in the C- and N-terminal regions and ranked each pair according to total length of similarity. The highest-ranking comparisons are shown in Figure 5. For the C-terminal region, 5 of the 6 within-group comparisons had clear similarity (Figure 5, B–E and I) as opposed to only 4 of the 72 between-group comparisons (Figure 5, A and F–H). In the N-terminal regions, only 1 of the 78 comparisons had notable similarity (Figure 5J). This

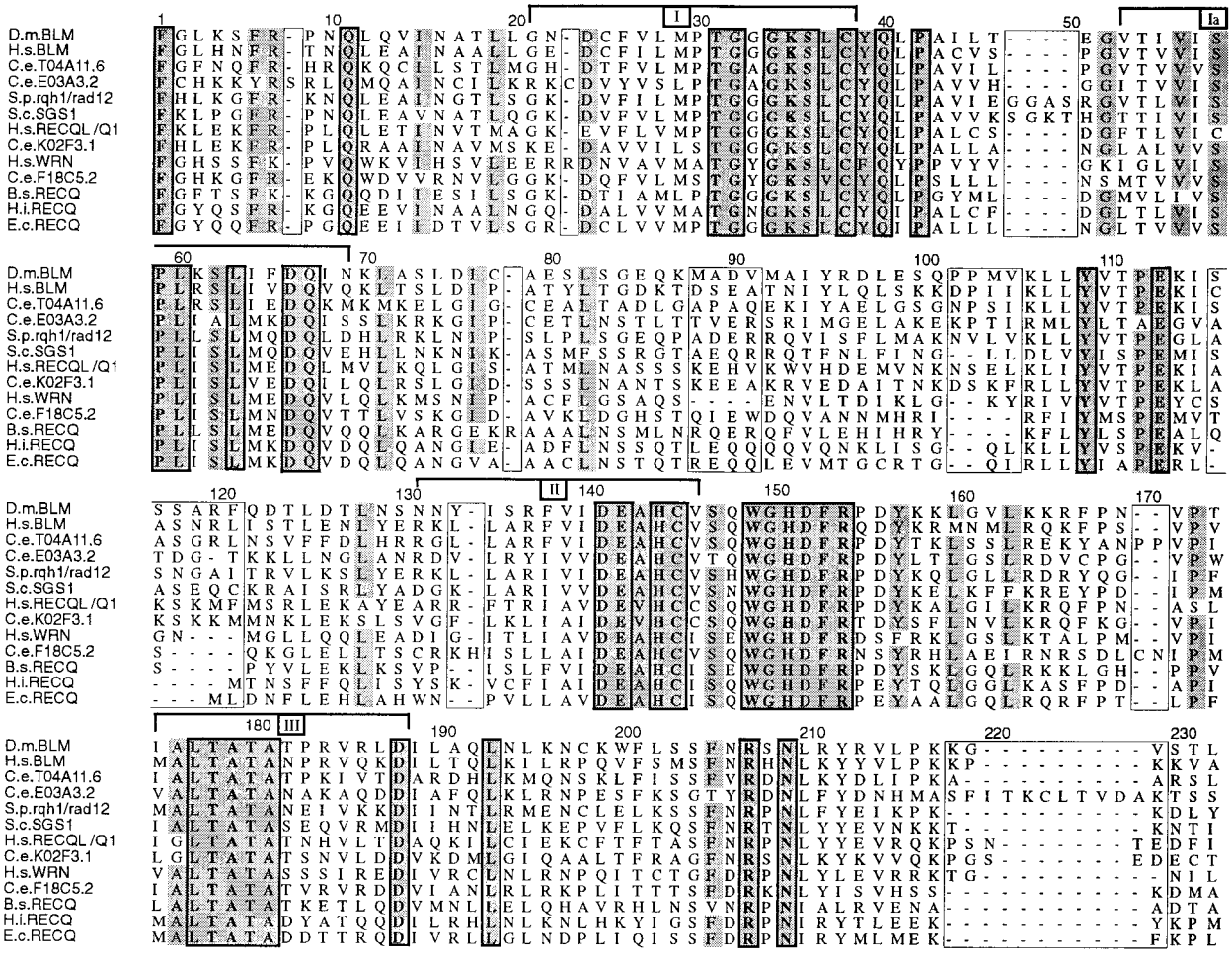


Figure 2.—Aligned amino acid residues of the 13 helicase domains. The total consensus length was 419 sites of which 349 were retained for analysis after deletion of ambiguously aligned residues. Boxed and bold regions indicate invariant sites. Shaded residue positions indicate 80% or higher identity for that particular site. Boxed regions without shading indicate areas that were not unambiguously aligned and excluded from the phylogenetic analysis. Gaps introduced by the alignment algorithm are indicated by a dash. For a key to helicase domain abbreviations, see materials and methods. For the initial and final residue numbers of each helicase domain, see Figure 3.

comparison was also a within-group pair. We conclude that sequence similarities extend beyond the helicase domain for all 6 pairwise comparisons within the proposed groups and that there is relatively little such similarity between groups.

Domain structure comparison of the RECQ family: The 13 RECQ family proteins were arranged in Figure 3 on the basis of the subgrouping postulated above. Recently, the possibility was suggested that RECQ family members can be separated into distinct functional subgroups on the basis of the presence or absence of extensive N-terminal domains (Watt and Hickson 1996). Our analyses provide evidence against this possibility. For example, the N-terminal segment of *H. sapiens RECQL/Q1* is short as are the bacterial RECQs. However, in the helicase domain, *H. sapiens RECQL/Q1* is more closely related to *C. elegans K02F3.1*, which has a long N-terminal segment (Figure 3C). Furthermore, the short N-terminal segment of *H. sapiens RECQL/Q1* shares limited similarity with the region adjacent to the helicase domain of *C. elegans K02F3.1* (Figures 3C and 5J), suggesting that these short segments are important for a common function between *H. sapiens RECQL/Q1* and *C. elegans K02F3.1*. The N-terminal domain of *H. sapiens WRN* is as long as the N-terminal domains of *Dmblm*, *H. sapiens BLM*, *C. elegans T04A11.6*, *S. cerevisiae SGS1*, *S. pombe rqh1/rad12*, and *C. elegans K02F3.1*, all of which are members of other subgroups. Overall, it is clear from Figures 3 and 4 that the length of the N-terminal regions is not well correlated with the relatedness of the helicase domain sequences.

Presence of direct sequence repeats in the RECQ family: By examining dot plots, we detected several direct sequence repeats within the genes (Figure 3). The *H. sapiens WRN* gene was found to have a perfect repeat of 27 amino acids near its helicase domain as shown

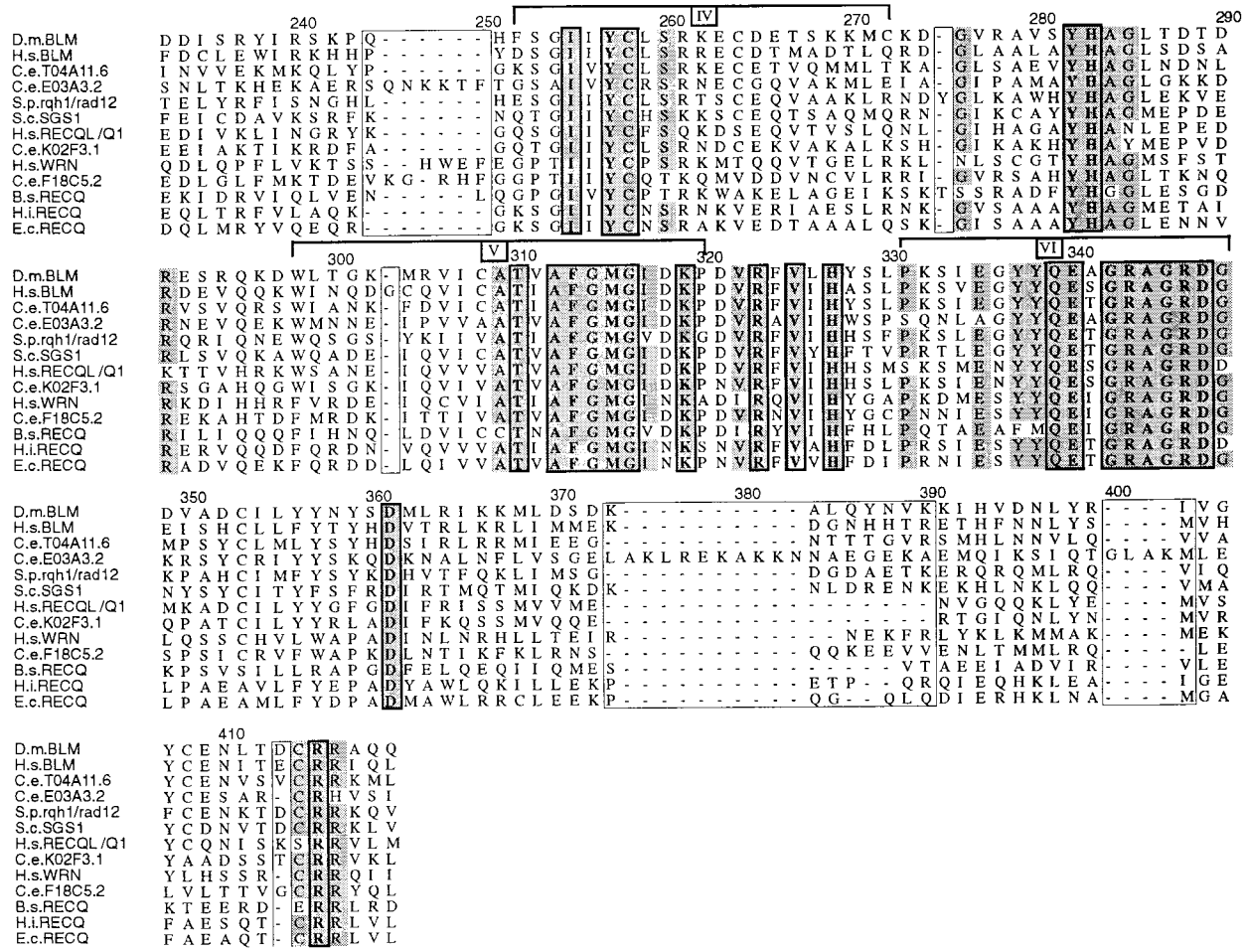


Figure 2.—*Continued.*

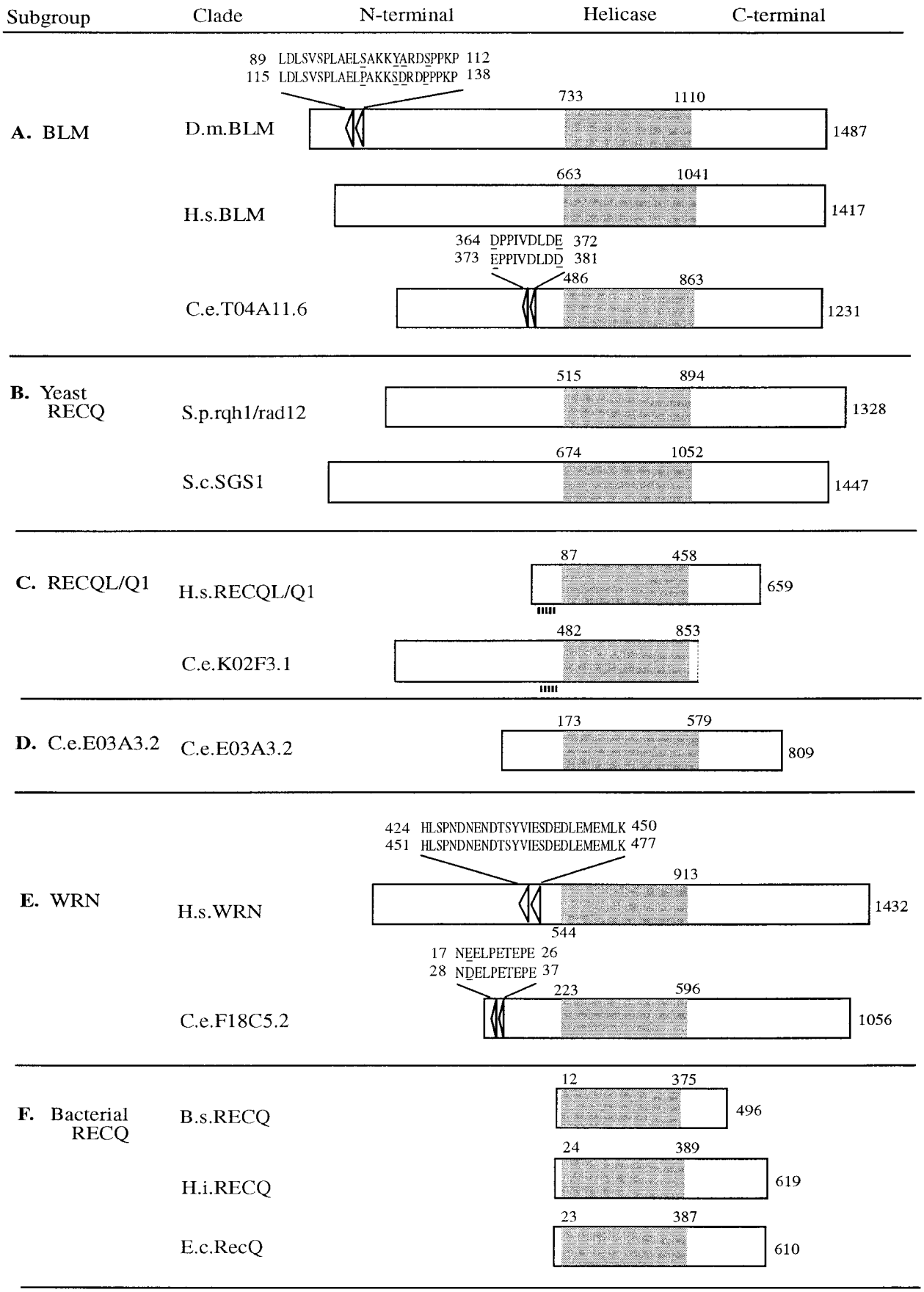
previously (Yu *et al.* 1996; Figure 3E). There is also a repeat of 20/24 amino acids near the N-terminal end of *Dmblm* (Figures 1 and 3A), a repeat of 7/9 amino acids in the N-terminal domain of the *C. elegans* T04-A11.6 gene (Figure 3A), and a repeat of 9/10 amino acids near the N-terminal end of the *C. elegans* F18C5.2 gene (Figure 3E). These repeat structures might be correlated with the apparent evolutionary flexibility of the length of the N-terminal sequences, as discussed below.

Rescue of the *S. cerevisiae sgs1* mutant by the *Dmblm* cDNA: The structural homology between the BLM and yeast RECQ subgroups (Figures 4 and 5) suggests the possibility that *Dmblm* functions in a manner similar to *SGS1*. First we tested sensitivity of the *sgs1* mutant to MMS. The *sgs1::URA3* mutant showed hypersensitivity to MMS (Figure 6A). The *SGS1* gene cloned in the yeast centromere plasmid complemented this MMS hypersensitivity (Figure 6A). We examined whether *Dmblm* can functionally substitute for *SGS1* in *S. cerevisiae*. The *Dmblm* cDNA placed downstream of the ADH1 promoter in the yeast 2-μm plasmids conferred a 12-fold increase in the survival fraction of the *sgs1* mutant cells against 0.02% MMS (Figure 6B), although it did not reach that of the wild-type strain carrying the vector.

DISCUSSION

***D. melanogaster Bloom*:** We have identified a new *D. melanogaster* gene and include it in the RECQ helicase family. On the basis of our phylogenetic analysis, dot plot analysis, and comparative domain structure analysis, this new gene, designated *Dmblm*, is most similar to the *BLM* gene of *H. sapiens*, strongly suggesting functional similarity.

The subgroups in the RECQ family: Maximum likelihood estimations, assuming the F84-dΓ model (Yang 1994b), were applied to the helicase domain present in all 13 RECQs. Dot plot analyses were executed for all domains of all 13 RECQs. On the basis of these results, we suggest that there are four possible subgroups in the RECQ family: (i) the BLM subgroup (*H. sapiens Bloom*, *D. melanogaster Dmblm*, and *C. elegans* T04A11.6); (ii) the yeast RECQ subgroup (*S. cerevisiae SGS1* and *S. pombe rqh1/rad12*); (iii) the RECQL/Q1 subgroup (*H. sapiens RECQL/Q1* and *C. elegans* K02F3.1); and (iv) the WRN


Subgroup
Clade
N-terminal
Helicase
C-terminal
89 LDLSVSPLAELSAKKYARDSPPKP 112
115 LDLSVSPLAELPAKKSDRDPPPKP 138
A. BLM
D.m.BLM
733
1110
1487
663
1041
H.s.BLM
1417
364 DPPIVDLDE 372
373 EPPIVDLDD 381
486
863
C.e.T04A11.6
1231
B. Yeast RECQ
S.p.rqh1/rad12
515
894
1328
674
1052
S.c.SGS1
1447
C. RECQL/Q1
H.s.RECQL/Q1
87
458
659
482
853
C.e.K02F3.1
D. C.e.E03A3.2
C.e.E03A3.2
173
579
809
424 HLSPNDNENDTSYVIESDEDLEMEMLK 450
451 HLSPNDNENDTSYVIESDEDLEMEMLK 477
E. WRN
H.s.WRN
913
544
1432
17 NEELPETEPE 26
28 NDELPETEPE 37
223
596
C.e.F18C5.2
1056
F. Bacterial RECQ
B.s.RECQ
12
375
496
H.i.RECQ
24
389
619
E.c.RecQ
23
387
610

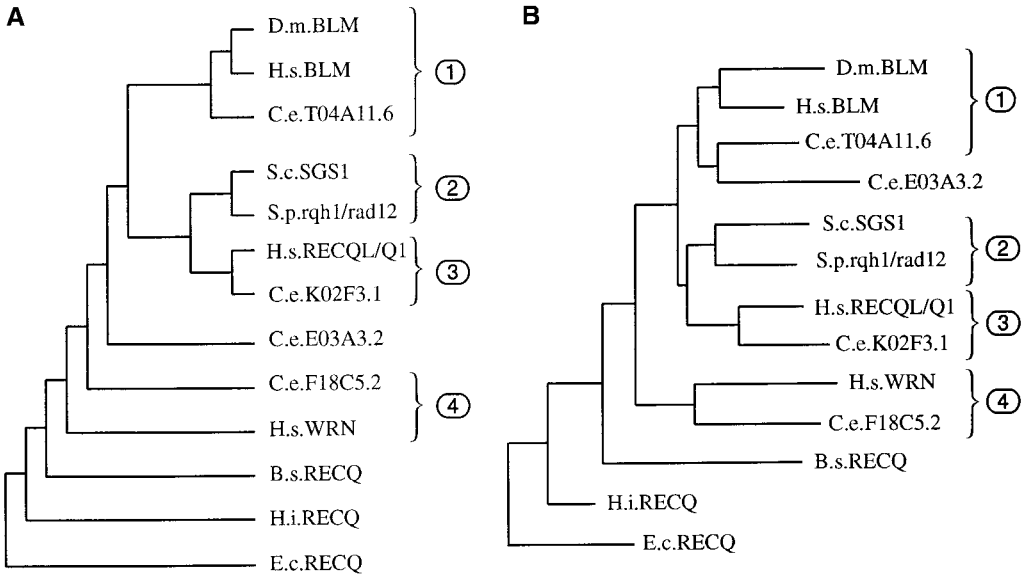


Figure 4.—Topology of the 13 helicase domains depicted with branch lengths proportional to the expected substitutions. The four groups were inferred from the helicase domains as well as from the flanking regions (see Figure 5). (A) MP1. Maximum parsimony tree. (B) Tree 2. Maximum likelihood topology of the best tree found by global branch swapping of the initial maximum parsimony tree (MP1) using the F84-dΓ substitution model.

subgroup (*H. sapiens Werner* and *C. elegans* F18C5.2). Our designation of subgroups provides reasonable insights for functional analysis of RECQ family genes.

N-terminal length variation in the RECQ family: The presence of four direct repeats in the N-terminal domains allows us to suggest a correlation with the apparent evolutionary variability of the length of the N-terminal sequences. Gene elongation during evolution seems to have depended largely on domain duplications (Barker *et al.* 1978). The multiple duplication of small units would partially contribute to elongation of N-terminal ends of an ancestor gene that might be similar to bacterial RECQs. The small duplications shown in Figure 3 in *Dmblm*, *C. elegans T04A11.6*, *C. elegans F18C5.2*, and *H. sapiens WRN* would likely have occurred recently in evolutionary history. The *WRN* homologues of two vertebrates, *Mus musculus* (Imamura *et al.* 1997) and *Xenopus laevis* (Yan *et al.* 1998), have only one such sequence. The length difference of their N-terminal domains might depend on the length of the original repeat unit and the frequency of duplication events. However, the first half of the N-terminal segment of the Werner protein includes a nuclease domain homologous to bacterial RNaseD and the 3′-5′ exonuclease domain of DNA polymerase I (Mian 1997; Mushegian *et al.* 1997). This indicates that other mechanisms such as exon shuffling contributed to the N-terminal length variation.

Features and perspectives in the functional subgroups: The relationships among members of the RECQ family do not follow the distinction between vertebrates and invertebrates, as seen in the three subgroups that contain both human and invertebrate RECQ genes (Figure 4). These features may indicate that gene duplication of ancestral RECQ sequences occurred several times. Over evolutionary time each RECQ gene would diverge and acquire a new function stabilized by natural selection or would be lost, as proposed recently by Fryxell (1996).

Mutations of *S. cerevisiae SGS1* increased mitotic recombination in ribosomal DNA (rDNA) repeats (Gangloff *et al.* 1994) and between direct repeated genes (Watt *et al.* 1996). We found that an *sgs1* mutant is hypersensitive to MMS. Mutations of *S. pombe rqh1*/*rad12* caused the hypersensitivity to UV, hydroxyurea or γ-ray (Murray *et al.* 1997; Stewart *et al.* 1997; Davey *et al.* 1998) and increased mitotic recombination with hydroxyurea treatment (Stewart *et al.* 1997). *S. pombe rqh1*/*rad12* appear to function on a checkpoint protein dependent DNA damage response during S phase (Murray *et al.* 1997; Stewart *et al.* 1997; Davey *et al.* 1998). Yeast RECQ proteins might inhibit inappropriate homologous recombination associated with postreplicational repair against DNA damage. Also, the *SGS1* gene was shown to be involved in cellular aging (Sinclair *et al.* 1997), which was associated with accumulation of extrachromosomal rDNA circles (Sinclair and Guarente 1997).

Figure 3.—(A–F) Schematic representation of proteins of RECQ family members. Each entire protein is exhibited by a box with its total amino acid residue number on the right end. Shaded regions indicate the helicase domains whose sequences are shown in Figure 2 with the initial and final residue numbers. The C-terminal region of *C. elegans* K02F3.1 is not yet sequenced completely and is shown by a dashed line. Each protein is aligned along the position of the initial residue of the helicase domain to compare the length of N-terminal domains between each protein. Each of the short dotted lines in (C) shows the region sharing sequence similarity to the N-terminal domains of *C. elegans* K02F3.1 and *H. sapiens RECQL*/*Q1* (see Figure 5J). A pair of triangles indicates a duplication with the indicated amino acid sequence. Substitutions within the duplications are underlined.

TABLE 1

Likelihood ratio tests for rate heterogeneity across nucleotide sites and among the RECQ homologs

	H_O (null hypothesis)			H_A (alternative hypothesis)				
Tree	Model	Clock	ln L	Model	Clock	ln L	2(ln L1 − ln L0)	*P* value
MP1[a]	F84	No	−14763.39	F84-dΓ	No	−14327.03	436.36	<0.001
MP1[a]	F84-dΓ	Yes	−14668.76	F84-dΓ	No	−14327.03	341.73	<0.001

[a] The maximum parsimony tree.

Somatic cells from patients with Bloom syndrome show increased sister chromatid exchanges (Heartlein *et al.* 1987; Kurihara *et al.* 1987) and increased interchanges between homologous chromosomes (German 1993). BLM cells are shown to be hypersensitive to ethyl methane-sulfonate (Krepinsky *et al.* 1979). *Dmblm* cDNA partially rescued the MMS sensitivity of an *sgs1* mutant. *H. sapiens BLM* has a similar rescue activity (Yamagata *et al.* 1998). These results provide evidence of the functional similarity between the BLM subgroup and the yeast RECQ subgroup. The BLM subfamily members might inhibit extra homologous recombination at/near the site of DNA damage. In *D. melanogaster*, *P*-element-induced double strand gaps are repaired by gene conversion accompanied very infrequently with crossing over (Nassif and Engels 1993). Interchromosomal re-

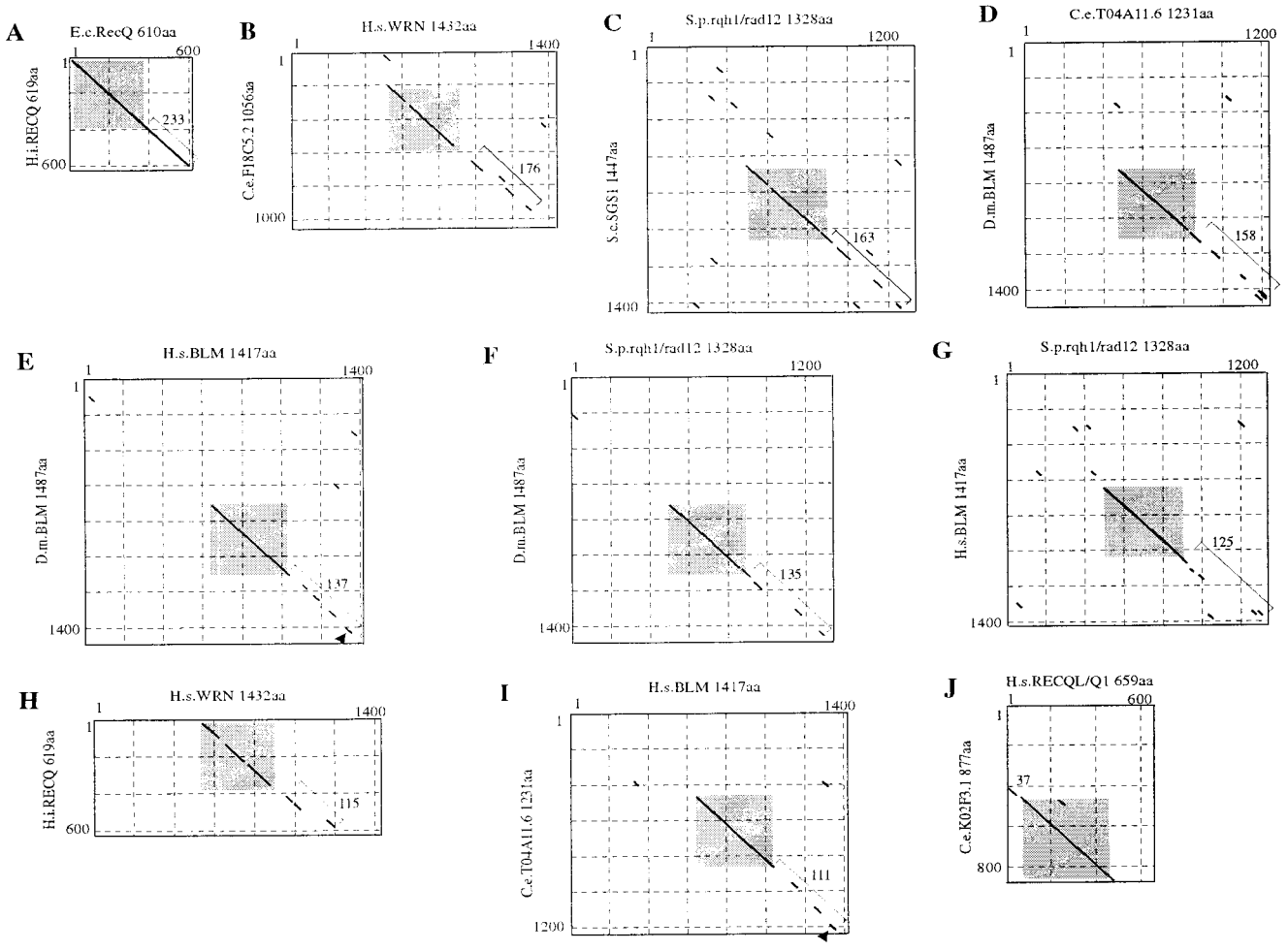


Figure 5.—Dot plot analysis between proteins of RECQ family members. The entire amino acid sequences were compared by a PAM250 scoring matrix, a window size of 30 residues, and a minimum score of 35% similarity. Shading indicates the helicase domains aligned in Figure 2. The total of length of the dots (excluding helicase domain) that lie on the diagonal line are shown. Dots with a closed triangle include the sequences similar to the nuclear translocation signal of *H. sapiens BLM* (Kaneko *et al.* 1997).

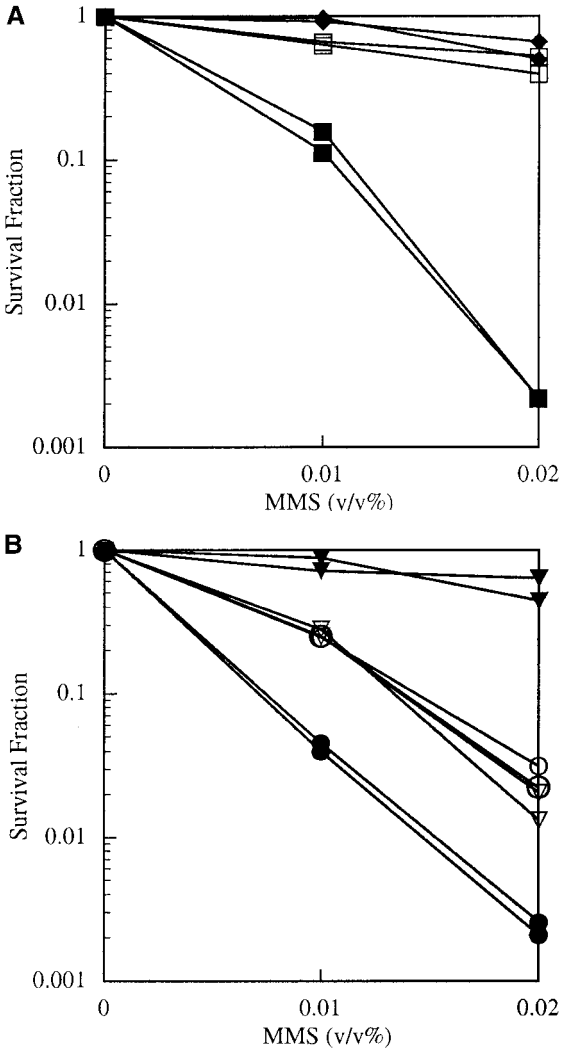


Figure 6.—Rescue of sensitivity in *S. cerevisiae* to MMS of an *sgs1* mutant by *Dmblm*. MMS sensitivity of the *sgs1* mutant and its wild-type strain carrying a centromere plasmid (A) or a 2-μm plasmid (B) with a TRP marker. Symbols in (A) are as follows: ■, MRQ966 (*sgs1::URA3*) with pRS314 (vector); ◆, MR966 (*SGS1*) with pRS314 (vector); and □, MRQ966 (*sgs1::URA3*) with YCp1305 (*SGS1* plasmid). Symbols in (B) are as follows: ●, MRQ966 (*sgs1::URA3*) with pAS2-1 (vector); ▼, MR966 (*SGS1*) with pAS2-1 (vector); ○, MRQ966 (*sgs1::URA3*) with pYEDMBLM1 (Gal4-DNA-binding domain::*Dmblm* cDNA); and ▽, MRQ966 (*sgs1::URA3*) with pYEDMBLM2 (*Dmblm* cDNA). Cells were grown in selective medium without tryptophan to maintain selection of plasmids in late-exponential phase, diluted, and plated on YEPD plates containing 0.01 and 0.02% MMS. The plates were incubated for 4 days at 30° before scoring viable colonies. Measurements were done for two independent transformants from each of the strains.

combination in mitotic cells of mammals is very low compared with that in yeast (Shulman *et al.* 1995). These reports indicate that mitotic crossing over is strictly inhibited in metazoans. *H. sapiens BLM* and *Dmblm* might be involved in preventing the loss of heterozygosity induced by DNA damage.

Cells from patients with Werner syndrome exhibit various chromosome rearrangements (Salk *et al.* 1981; Scappaticci *et al.* 1982), but do not show increased sister chromatid exchange (Gebhart *et al.* 1988), in contrast to *BLM* mutants. *WRN* cells exhibited a slower rate of DNA replication (Fujiwara *et al.* 1977). More recently *WRN* cells were shown to be hypersensitive to 4-nitro-quinoline-1-oxide (Ogburn *et al.* 1997). The Werner protein is likely to work with replication protein A (Shen *et al.* 1998). Yan *et al.* (1998) suggested the role of Werner protein and replication protein A in DNA replication. The WRN subgroup members might be involved in resumption of DNA replication following postreplicational repair. The human *WRN* gene and the *C. elegans* F18C5.2 share siginificant similarity in their helicase and C-terminal domains. This indicates that *C. elegans* F18C5.2 may be a useful model of Werner syndrome.

In the RECQL/Q1 subgroup, there is no genetical data. We suggest that a genetical approach using *C. elegans* K02F3.1 may be useful to address the role of RECQL/Q1 in humans. A biochemical finding suggests that the human RECQL/Q1 protein possesses DNA helicase activity and translocates in a 3′ to 5′ direction on the DNA to which it binds (Seki *et al.* 1994). The manner in which the short N-terminal segment shared by the members of this subgroup affects their DNA helicase activities may be a key for the clarification of their specific function.

In *E. coli*, homologous recombination is initiated in several ways (Clark and Low 1988; Lloyd and Low 1996). The *E. coli* RecQ helicase appears to initiate a type of homologous recombination with RecJ exonuclease (Kusano *et al.* 1994). In a *recB*$^-$*C*$^-$ *sbcA*$^-$ background, *recQ*$^-$*J*$^-$ strains are more resistant to UV-irradiation than *recJ*$^-$ strains (Kusano *et al.* 1994; Lovett and Sutera 1995). In a *recB*$^-$*C*$^-$ *sbcB*$^-$*C*$^-$ background, however, a *recQ*$^-$*J*$^-$ strain is more sensitive to UV-irradiation than a *recJ*$^-$ strain (data not shown). These two results indicate that *E. coli* RecQ contains an additional function(s) as well as one epistatic to RecJ function. Also, the *E. coli* RecQ gene appears to control nonhomologous recombination (Hanada *et al.* 1997). The functional relationships between bacterial RECQs and eukaryotic RECQs are important in order to understand how RECQ family members have diverged.

Meaning of suppression by mutations in the RECQ family members: Consider two genes A and B. A$^-$ suppresses a defect due to B$^-$. This genetic interaction is not allele-specific, because a null mutation in A, such as a deletion or insertion, causes suppression. The following two phenomena appear to correspond to this genetic interaction: the suppression of the defect of a *top3* mutation (B$^-$) by an *sgs1* null mutation (A$^-$) in *S. cerevisiae* (Gangloff *et al.* 1994; Lu *et al.* 1996) and the suppression of the defects of *recJ* mutations (B$^-$) by *recQ* null mutations (A$^-$) in *E. coli.* (Kusano *et al.* 1994; Lovett and Sutera 1995). The two phenomena could

be explained by the hypothesis that A^- activates a pathway substituted for the AB pathway so as to alleviate any severe phenotype caused by B^-. In this way, the *sgs1*$^-$ mutation might enable type I topoisomerase to substitute for Top3 in *S. cerevisiae* cells (Lu *et al.* 1996). The *recQ*$^-$ mutation might open other homologous recombination pathways for *E. coli* cells (Kusano *et al.* 1994). The RECQ family members might therefore operate on the early steps in the intracellular events that involve them.

We thank Christine Preston and Dena Johnson-Schlitz for helping in the examination of polytene chromosomes and Carlos Flores for helpful comments on the manuscript. We are grateful to Takemi Enomoto and Akira Shimamoto for providing yeast strains and plasmids. This work was supported by National Institutes of Health grant GM-30948 and Uehara Memorial Foundation. This is paper 3526 from the University of Wisconsin Laboratory of Genetics.

LITERATURE CITED

Barker, W. C., L. K. Ketcham and M. O. Dayhoff, 1978 Duplications in protein sequences, pp. 359–362 in *Atlas of Protein Sequence and Structure*, edited by M. O. Dayhoff. National Biomedical Research Foundation, Silver Spring, MD.

Bennett, R. J., J. A. Sharp and J. C. Wang, 1998 Purification and characterization of the Sgs1 DNA helicase activity of Saccharomyces cerevisiae. J. Biol. Chem. **273:** 9644–9650.

Chomczynski, P., and N. Sacchi, 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156–159.

Clark, A. J., and K. B. Low, 1988 Pathways and systems of homologous recombination in *Escherichia coli*, pp. 155–215 in *The Recombination of Genetic Material*, edited by K. B. Low. Academic Press, Inc., NY.

Davey, S., C. S. Han, S. A. Ramer, J. C. Klassen, A. Jacobson *et al.*, 1998 Fission yeast *rad12*$^+$ regulates cell cycle checkpoint control and is homologous to the Bloom's syndrome disease gene. Mol. Cell. Biol. **18:** 2721–2728.

Eisen, J. A., K. S. Sweder and P. C. Hanawalt, 1995 Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. Nucleic Acids Res. **23:** 2715–2723.

Ellis, N. A., J. Groden, T.-Z. Ye, J. Straughen, D. J. Lennon *et al.*, 1995 The Bloom's syndrome gene product is homologous to RecQ helicases. Cell **83:** 655–666.

Engels, W. R., C. R. Preston, P. Thompson and W. B. Eggleston, 1986 *In situ* hybridization to Drosophila salivary chromosomes with biotinylated DNA probes and alkaline phosphatase. Focus **8:** 6–8.

Felsenstein, J., 1978 Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27:** 401–410.

Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N. A. Jenkins *et al.*, 1993 The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. Cell **75:** 1027–1038.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness *et al.*, 1995 Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd.* Science **269:** 496–512.

Fryxell, K. J., 1996 The coevolution of gene family trees. Trends Genet. **12:** 364–369.

Fujiwara, Y., T. Higashikawa and M. Tatsumi, 1977 A retarded rate of DNA replication and normal level of DNA repair in Werner's syndrome fibroblasts in culture. J. Cell. Physiol. **92:** 365–374.

Gangloff, S., J. P. McDonald, C. Bendixen, L. Arthur and R. Rothstein, 1994 The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. Mol. Cell. Biol. **14:** 8391–8398.

Gebhart, E., R. Bauer, U. Raub, M. Schinzel, K. W. Ruprecht *et al.*, 1988 Spontaneous and induced chromosomal instability in Werner syndrome. Hum. Genet. **80:** 135–139.

German, J., 1993 Bloom syndrome: a mendelian prototype of somatic mutational disease. Medicine **72:** 393–406.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko and V. M. Blinov, 1989 Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:** 4713–4730.

Gray, M. D., J.-C. Shen, A. S. Kamath-Loeb, A. Blank, B. L. Sopher *et al.*, 1997 The Werner syndrome protein is a DNA helicase. Nat. Genet. **17:** 100–103.

Hanada, K., T. Ukita, Y. Kohno, K. Saito, J. Kato *et al.*, 1997 RecQ DNA helicase is a suppressor of illegitimate recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **94:** 3860–3865.

Hartl, D. L., D. I. Nurminsky, R. W. Jones and E. R. Lozovskaya, 1994 Genome structure and evolution in *Drosophila*: applications of the framework P1 map. Proc. Natl. Acad. Sci. USA **91:** 6824–6829.

Heartlein, M. W., H. Tsuji and S. A. Latt, 1987 5-Bromodeoxyuridine-dependent increase in sister chromatid exchange formation in Bloom's syndrome is associated with reduction in topoisomerase II activity. Exp. Cell Res. **169:** 245–254.

Henikoff, S., and J. G. Henikoff, 1994 Position-based sequence weights. J. Mol. Biol. **243:** 574–578.

Higgins, D. G., A. J. Bleasby and R. Fuchs, 1992 CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:** 189–191.

Hillis, D. M., 1991 Discrimination between phylogenetic signal and random noise in DNA sequences, pp. 278–294 in *Phylogenetic Analysis of DNA Sequences*, edited by M. M. Miyamoto and J. Cracraft. Oxford University Press, Oxford, UK.

Hillis, D. M., and J. P. Huelsenbeck, 1992 Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. **83:** 189–195.

Huelsenbeck, J. P., 1991 Tree-length distribution skewness: an indicator of phylogenetic information. Syst. Zool. **40:** 257–270.

Huelsenbeck, J. P., and D. M. Hillis, 1993 Success of phylogenetic methods in the four-taxon case. Syst. Biol. **42:** 247–264.

Huelsenbeck, J. P., and B. Rannala, 1997 Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science **276:** 227–232.

Imamura, O., K. Ichikawa, Y. Yamabe, M. Goto, M. Sugawara *et al.*, 1997 Cloning of a mouse homologue of the human Werner syndrome gene and assignment to 8A4 by fluorescence *in situ* hybridization. Genomics **41:** 298–300.

Irino, N., K. Nakayama and H. Nakayama, 1986 The *recQ* gene of *Escherichia coli* K12: primary structure and evidence for SOS regulation. Mol. Gen. Genet. **205:** 298–304.

Ishizaki, K., T. Yagi, M. Inoue, O. Nikaido and H. Takebe, 1981 DNA repair in Bloom's syndrome fibroblasts after UV irradiation or treatment with Mitomycin C. Mutat. Res. **80:** 213–219.

Kaneko, H., K. Orii, E. Matsui, N. Shimozawa, T. Fukao *et al.*, 1997 BLM (the causative gene of Bloom syndrome) protein translocation into the nucleus by a nuclear localization signal. Biochem. Biophys. Res. Commun. **240:** 348–353.

Karow, J. K., R. K. Chakraverty and I. D. Hickson, 1997 The Bloom's syndrome gene product is a 3′-5′ DNA helicase. J. Biol. Chem. **272:** 30611–30614.

Krepinsky, A. B., J. A. Heddle and J. German, 1979 Sensitivity of Bloom's syndrome lymphocytes to ethyl methanesulfonate. Hum. Genet. **50:** 151–156.

Kurihara, T., M. Inoue and K. Tatsumi, 1987 Hypersensitivity of Bloom's syndrome fibroblasts to N-ethyl-N-nitrosourea. Mutat. Res. **184:** 147–151.

Kusano, K., Y. Sunohara, N. Takahashi, H. Yoshikura and I. Kobayashi, 1994 DNA double-strand break repair: genetic determinants of flanking crossing-over. Proc. Natl. Acad. Sci. USA **91:** 1173–1177.

Lloyd, R. G., and K. B. Low, 1996 Homologous recombination, pp. 2236–2255 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by F. C. Neidhardt. ASM Press, Washington, DC.

Lovett, S. T., and V. A. Sutera, Jr., 1995 Suppression of RecJ exonuclease mutants of *Escherichia coli* by alterations in DNA Helicases II (*uvrD*) and IV (*helD*). Genetics **140:** 27–45.

Lu, J., J. R. Mullen, S. J. Brill, S. Kleff, A. M. Romeo *et al.*, 1996 Human homologues of yeast helicase. Nature **383:** 678–679.

Mian, I. S., 1997 Comparative sequence analysis of ribonucleases HII, III, II, PH and D. Nucleic Acids Res. **25:** 3187–3195.

Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White *et al.*,

1992 Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res. **20:** 4255–4262.

Murray, J. M., H. D. Lindsay, C. A. Munday and A. M. Carr, 1997 Role of *Schizosaccharomyces pombe* RecQ homolog, Recombination, and checkpoint genes in UV damage tolerance. Mol. Cell. Biol. **17:** 6868–6875.

Muse, S. V., and B. S. Gaut, 1994 A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11:** 715–724.

Muse, S. V., and B. S. Weir, 1992 Testing for equality of evolutionary rates. Genetics **132:** 269–276.

Mushegian, A. R., D. E. Bassett, Jr., M. S. Boguski, P. Bork and E. V. Koonin, 1997 Positionally cloned human disease genes: patterns of evolutionary conservation and functional motifs. Proc. Natl. Acad. Sci. USA **94:** 5831–5836.

Nakayama, H., K. Nakayama, R. Nakayama, N. Irino, Y. Nakayama *et al.*, 1984 Isolation and genetic characterization of a thymineless death-resistant mutant of *Escherichia coli* K12: identification of a new mutation (*recQ1*) that blocks the RecF recombination pathway. Mol. Gen. Genet. **195:** 474–480.

Nassif, N., and W. R. Engels, 1993 DNA homology requirements for mitotic gap repair in Drosophila. Proc. Natl. Acad. Sci. USA **90:** 1262–1266.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, NY.

Nei, M., 1991 Relative efficiencies of different tree-making methods for molecular data, pp. 90–128 in *Phylogenetic Analysis of DNA Sequences*, edited by M. M. Miyamoto and J. L. Cracraft. Oxford University Press, Oxford, UK.

Ogburn, C. E., J. Oshima, M. Poot, R. Chen, K. E. Hunt *et al.*, 1997 An apoptosis-inducing genotoxin differentiates heterozygotic carriers for Werner helicase mutations from wild-type and homozygous mutants. Hum. Genet. **101:** 121–125.

Puranam, K. L., and P. J. Blackshear, 1994 Cloning and characterization of RECQL, a potential human homologue of the *Escherichia coli* DNA helicase RecQ. J. Biol. Chem. **269:** 29838–29845.

Roels, S., A. Driks and R. Losick, 1992 Characterization of *spoIVA*, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. J. Bacteriol. **174:** 575–585.

Salk, D., K. Au, H. Hoehn, M. R. Stenchever and G. M. Martin, 1981 Evidence of clonal attenuation, clonal succession, and clonal expansion in mass cultures of aging Werner's syndrome skin fibroblasts. Cytogenet. Cell Genet. **30:** 108–117.

Scappaticci, S., D. Cerimele and M. Fraccaro, 1982 Clonal structural chromosomal rearrangements in primary fibroblast cultures and in lymphocytes of patients with Werner's syndrome. Hum. Genet. **62:** 16–24.

Seki, M., H. Miyazawa, S. Tada, J. Yanagisawa, T. Yamaoka *et al.*, 1994 Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to *Escherichia coli* Rec Q helicase and localization of the gene at chromosome 12p12. Nucleic Acids Res. **22:** 4566–4573.

Seki, M., J. Yanagisawa, T. Kohda, T. Sonoyama, M. Ui *et al.*, 1994 Purification of two DNA-dependent Adenosinetriphosphatases having DNA helicase activity from Hela cells and comparison of the properties of the two enzymes. J. Biochem. **115:** 523–531.

Shen, J.-C., M. D. Gray, J. Oshima and L. A. Loeb, 1998 Characterization of Werner syndrome protein DNA helicase activity: directionality, substrate dependence and stimulation by replication protein A. Nucleic Acids Res. **26:** 2879–2885.

Shulman, M. J., C. Collins, A. Connor, L. R. Read and M. D. Baker, 1995 Interchromosomal recombination is suppressed in mammalian somatic cells. EMBO J. **16:** 4102–4107.

Sikorski, R. S., and P. Hieter, 1989 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics **122:** 19–27.

Sinclair, D. A., and L. Guarente, 1997 Extrachromosomal rDNA circles—a cause of aging in yeast. Cell **91:** 1033–1042.

Sinclair, D. A., K. Mills and L. Guarente, 1997 Accelerated aging and nucleolar fragmentation in yeast *sgs1* mutants. Science **277:** 1313–1316.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry*, Ed. 2. W. H. Freeman, San Francisco, CA.

Sorokin, A., E. Zumstein, V. Azevedo, S. D. Ehrlich and P. Serror, 1993 The organization of the *Bacillus subtilis* 168 chromosome region between the *spoVA* and *serA* genetic loci, based on sequence data. Mol. Microbiol. **10:** 385–395.

Sorokin, A., V. Azevedo, E. Zumstein, N. Galleron, S. D. Ehrlich *et al.*, 1996 Sequence analysis of the *Bacillus subtilis* chromosome region between the *serA* and *kdg* loci cloned in a yeast artificial chromosome. Microbiology **142:** 2005–2016.

Stewart, E., C. R. Chapman, F. Al-khodairy, A. M. Carr and T. Enoch, 1997 $rqh1^+$, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest. EMBO J. **16:** 2682–2692.

Sullivan, J., K. E. Holsinger and C. Simon, 1996 The effect of topology on estimates of among-site rate variation. J. Mol. Evol. **42:** 308–312.

Suzuki, N., A. Shimamoto, O. Imamura, J. Kuromitsu, S. Kitao *et al.*, 1997 DNA helicase activity in Werner's syndrome gene product synthesized in a baculovirus system. Nucleic Acids Res. **25:** 2973–2978.

Swofford, D. L., 1993 PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.

Thorne, J. L., and H. Kishino, 1992 Freeing phylogenies from artifacts of alignment. Mol. Biol. Evol. **9:** 1148–1162.

Thorne, J. L., H. Kishino and J. Felsenstein, 1991 An evolutionary model for maximum likelihood alignment of DNA sequences. J. Mol. Evol. **33:** 114–124.

Thorne, J. L., H. Kishino and J. Felsenstein, 1992 Inching toward reality: an improved likelihood model of sequence evolution. J. Mol. Evol. **34:** 3–16.

Umezu, K., K. Nakayama and H. Nakayama, 1990 *Escherichia coli* RecQ protein is a DNA helicase. Proc. Natl. Acad. Sci. USA **87:** 5363–5367.

Uzzell, T., and K. W. Corbin, 1971 Fitting discrete probability distributions to evolutionary events. Science **172:** 1089–1096.

Watt, P. M., and I. D. Hickson, 1996 Genome stability: failure to unwind causes cancer. Curr. Biology **6:** 265–267.

Watt, P. M., E. J. Louis, R. H. Borts and I. D. Hickson, 1995 Sgs1: a eukaryotic homolog of E. coli RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation. Cell **81:** 253–260.

Watt, P. M., I. D. Hickson, R. H. Borts and E. J. Louis, 1996 *SGS1*, a homologue of the Bloom's and Werner's syndrome genes, is required for maintenance of genomic stability in *Saccharomyces cerevisiae*. Genetics **144:** 935–945.

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks *et al.*, 1994 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature **368:** 32–38.

Wu, C.-I., and W.-H. Li, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

Yamagata, K., J. Kato, A. Shimamoto, M. Goto, Y. Furuichi *et al.*, 1998 Bloom's and Werner's syndrome genes suppress hyperrecombination in yeast *sgs1* mutant: implication for genomic instability in human diseases. Proc. Natl. Acad. Sci. USA **95:** 8733–8738.

Yan, H., C.-Y. Chen, R. Kobayashi and J. Newport, 1998 Replication focus-forming activity 1 and the Werner syndrome gene product. Nat. Genet. **19:** 375–378.

Yang, Z., 1994a Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39:** 105–111.

Yang, Z., 1994b Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39:** 306–314.

Yang, Z., 1996a Among site rate variation and its impact on phylogenetic analysis. Trends Ecol. Evol. **11:** 367–372.

Yang, Z., 1996b PAML: phylogenetic analysis by maximum likelihood, version 1.3. Department of Integrative Biology, University of California, Berkeley, CA.

Yu, C.-E., J. Oshima, Y.-H. Fu, E. M. Wijsman, F. Hisama *et al.*, 1996 Positional cloning of the Werner's syndrome gene. Science **272:** 258–262.

Communicating editor: R. S. Hawley