Molecular Microbiology (2004) **53**(1), 283–295 doi:10.1111/j.1365-2958.2004.04125.x

Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach

Hanjing Yang, Erika Wolff, Mandy Kim, Amy Diep and Jeffrey H. Miller*

Department of Microbiology, Immunology and Molecular Genetics, and the Molecular Biology Institute, 1602 Molecular Sciences Building, 405 Hilgard Avenue, University of California, Los Angeles, CA 90095, USA.

Summary

We searched for genes that create mutator phenotypes when put on to a multicopy plasmid in *Escherichia coli*. In many cases, this will result in overexpression of the gene in question. We constructed a random shotgun library with *E. coli* genomic fragments between 3 and 5 kbp in length on a multicopy plasmid vector that was transformed into *E. coli* to screen for frameshift mutators. We identified a total of 115 independent genomic fragments that covered 17 regions on the *E. coli* chromosome. Further studies identified 12 genes not previously known as causing mutator phenotypes when overproduced. A striking finding is that overproduction of the multidrug resistance transcription regulator, EmrR, results in a large increase in frameshift and base substitution mutagenesis. This suggests a link between multidrug resistance and mutagenesis. Other identified genes include those encoding DNA helicases (UvrD, RecG, RecQ), truncated forms of the DNA mismatch repair protein (MutS) and a primosomal component (DnaT), a negative modulator of initiation of replication/GATC-binding protein (SeqA), a stationary phase regulator AppY, a transcriptional regulator PaaX and three putative open reading frames, *ycgW*, *yfjY* and *yjiD*, encoding hypothetical proteins. In addition, we found three genes encoding proteins that were previously known to cause mutator effects under overexpression conditions: error-prone polymerase IV (DinB), DNA methylase (Dam) and sigma S factor (RpoS). This genomic strategy offers an approach to identify novel mutator effects resulting from the multicopy cloning (MCC) of specific genes and therefore complementing the conventional gene inactivation approach to finding mutators.

Accepted 11 March, 2004. *For correspondence. E-mail jhmiller@mbi.ucla.edu; Tel. (+1) 310 825 8460; Fax (+1) 310 206 3088.

Introduction

DNA instability results in a mutator phenotype, an elevated spontaneous mutation rate that can often be detrimental to survival. For example, in a bacterial mutator population generated by inactivation of the DNA mismatch repair pathway, deleterious mutations accumulate at a high rate leading to loss of fitness (Gibson *et al*., 1970; Funchain *et al*., 2000; Giraud *et al*., 2001). In humans, mutator cells can be pathogenic leading to tumorigenesis (reviewed by Heinen *et al*., 2002; Loeb *et al*., 2003). On the other hand, elevated mutation rates were found in several natural isolates of pathogenic bacteria in association with possible advantageous growth under severe environmental selection (Gross and Siegel, 1981; LeClerc *et al*., 1996; Matic *et al*., 1997; Bucci *et al*., 1999; Oliver *et al*., 2000; Giraud *et al*., 2001; Richardson and Stojiljkovic, 2001; Denamur *et al*., 2002). Therefore, defining mutator genes and mutational pathways is an important step towards preventing loss of genome integrity as well as understanding their roles in survival under adverse conditions.

Gene inactivation has commonly been used to search for mutator genes and mutational pathways. In bacteria, over 15 spontaneous mutator genes and their corresponding mutational pathways have been identified by means of mutagen-mediated point mutation inactivation or transposon-mediated insertion inactivation (reviewed by Miller, 1996). For example, genes involved in mismatch repair (*dam*, *mutS*, *mutL*, *mutH* and *uvrD*; Marinus and Morris, 1975; Nevers and Spatz, 1975; Glickman and Radman, 1980), repair and avoidance of oxidative damage (*mutY*, *mutM* and *mutT*; Treffers *et al*., 1954; Cabrera *et al*., 1988; Nghiem *et al*., 1988; Maki and Sekiguchi, 1992) and DNA proofreading (*mutD*; Echols *et al*., 1983) result in a mutator phenotype when inactivated. Other genes encode nucleoside diphosphate kinase (*ndk*; Lu *et al*., 1995; Miller *et al*., 2002), the positive regulator of oxidative damage gene (*oxyR*; Imlay and Linn, 1987), superoxide dismutase (*sodA*, *sodB*; Farr *et al*., 1986) and even tRNAs (*mutA*, *mutC*; Michaels *et al*., 1990).

Some studies have also shown that certain genes cause elevated levels of mutation when overexpressed, such as *dinB* that encodes the error-prone DNA polymerase IV (Kim *et al*., 1997; Wagner and Nohmi, 2000). In *Escherichia coli*, the expression of *dinB* is elevated during the damage-induced SOS response. So far,

however, a genome-wide study of overexpression to search for mutator genes and new mutational pathways has not been conducted. In this study, using *E. coli* as a model system, we applied multicopy cloning (MCC) and the knowledge of the complete sequence of the *E. coli* genome to search for genes that create mutator phenotypes when they are overexpressed. Not every gene will be overexpressed when present in multiple copies, as some are subject to regulation by repressors or activators. However, multiple copies of operators can titrate out repressors, in cases where the repressor is not also overproduced. Thus, we expect to detect most of the genes that generate a mutator phenotype when overexpressed.

Results

Screening for mutator DNA fragments by MCC

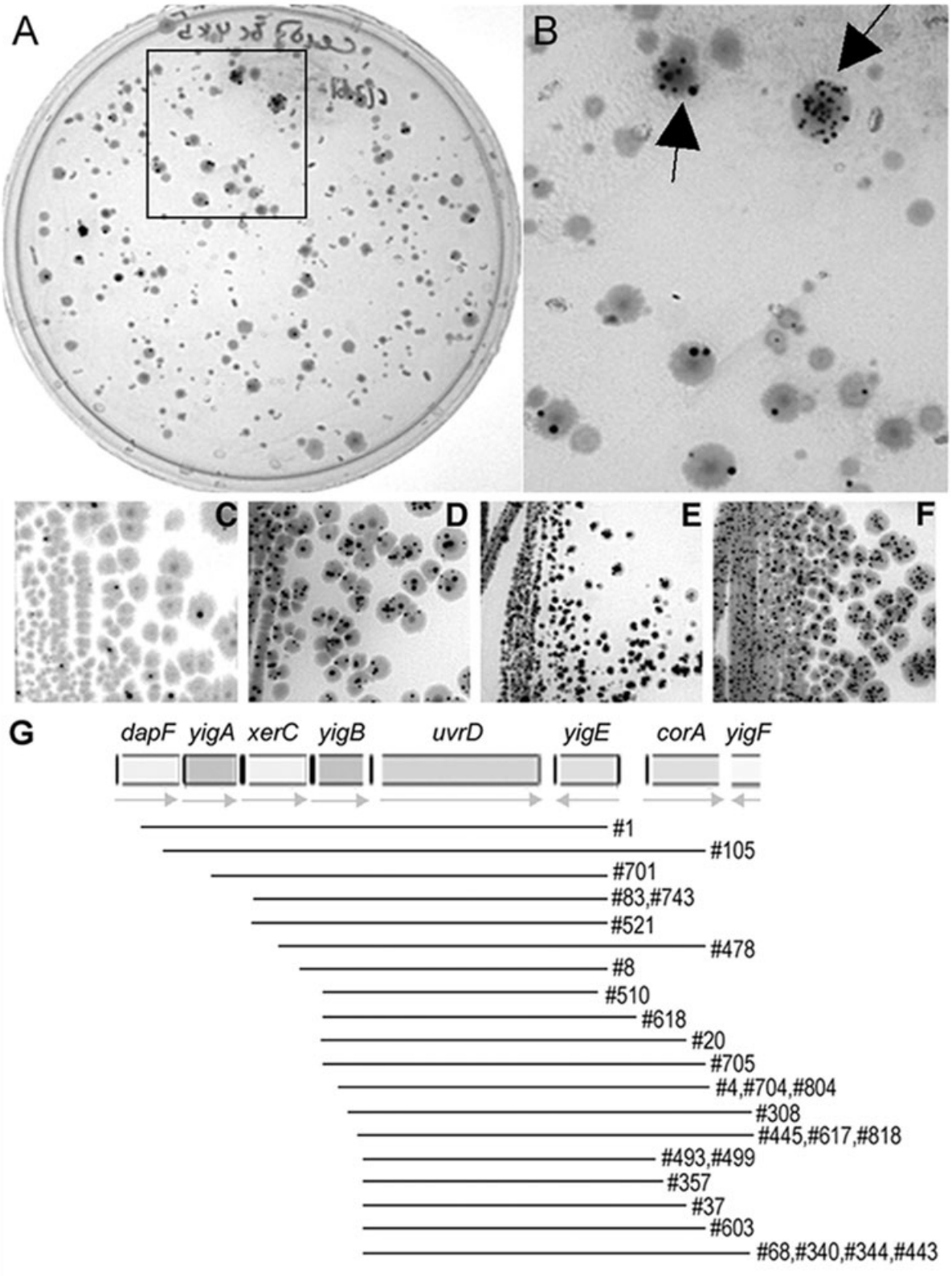
We looked for genes that, when put on to a multicopy plasmid, generate a mutator phenotype in a wild-type *E. coli* background. *E. coli* genomic fragments of 3–5 kbp in length generated by partial *Sau*3AI restriction enzyme digestion were used to create a random cloned library on a vector pCR2.1-TOPOCam derived from a multicopy pCR2.1-TOPO plasmid that encodes chloramphenicol resistance. We used an *E. coli* frameshift tester strain, CC107, as a host for the mutator screen. This strain specifically detects +1 frameshifts at a monotonous run of six Gs (six Cs on the opposite DNA strand) in *lacZ* (Cupples *et al*., 1990). The shotgun library of *E. coli* genomic DNA was electroporated into CC107, and the transformants were screened directly for frameshift mutators by looking for colonies with elevated levels of blue papillae (blue microcolonies) growing out of the surface of the colonies on glucose minimal plates supplemented with phenyl-β-D-galactoside (Pgal), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal; Nghiem *et al*., 1988) and chloramphenicol (Fig. 1A and B).

The mutator phenotypes of the identified colonies were confirmed by the following steps. First, the papillated colonies were restreaked to single colonies, and the plasmid from each isolate was purified. Then, the insert size on each plasmid was estimated by agarose gel electrophoresis using a supercoiled DNA ladder as a standard. Plasmids containing the inserts were transformed back into the CC107 host and retested for blue papillae formation (Fig. 1C–F).

After the mutator phenotype was confirmed, both 3′ and 5′ ends of each insert were sequenced, and the identity of the fragment was determined by BLAST analysis against the annotated genome sequence of *E. coli* K-12 (Blattner *et al*., 1997). The inserts that covered the overlapping region on the *E. coli* genome were compiled together to generate a composite for that particular region. One extreme example is shown in Fig. 1G, in which 28 inserts, including at least 19 independent inserts, overlapped the region encoding the DNA helicase gene *uvrD*. The levels of blue papillae formation of each insert within the same region were also compared to observe whether they confer similar mutator phenotypes.

We screened $\approx 1.5–2 \times 10^5$ transformants and identified 132 inserts, among which at least 115 were independent. The inserts ranged in size from 1.6 kbp to 6.8 kbp. Overall, 90% of the identified inserts were 3–5 kbp. They were derived from 17 regions on the *E. coli* genome, 14 regions of which are represented by multiple independent isolates (Fig. 2A). The mutagenic potency of each region was scored by measuring the Lac^+ revertant frequency in liquid culture (Fig. 2B), in addition to scoring the levels of blue papillation. We could classify the 17 regions into three groups according to their mutagenic potency. Group I (seven regions: 1, 3, 6, 8, 9, 10 and 12) exhibits high levels of mutagenesis with the Lac^+ revertant frequency increasing 35- to 9700-fold higher than that of the control (CC107 containing the vector without any insert). Group II (seven regions: 4, 5, 11, 14, 15, 16 and 17) exhibits moderate levels of mutagenesis with the Lac^+ revertant frequency increasing two- to sixfold. Group III (three regions: 2, 7 and 13) exhibits low levels of mutagenesis with the Lac^+ revertant frequency close to the control. It is worth noting that, despite the low levels of mutagenesis in group III scored by the Lac^+ revertant frequency, each of the three regions in group III had multiple independent inserts detected by the papillation method (Fig. 2A).

Fig. 1. Screening *E. coli* mutator DNA fragments.
A. A random shotgun library of *E. coli* genomic DNA was electroporated into CC107 strain, and electrotransformants grew on glucose minimal agar plates supplemented with chloramphenicol, Pgal and Xgal (Nghiem *et al*., 1988). A photograph of the transformation plate is shown.
B. A magnification of part of the transformation plate showing two mutator colonies that had elevated levels of blue papillae formation, as indicated by black arrows.
C–F. Confirmation of mutator phenotypes by blue papillation test. Purified plasmid DNA from the candidate mutator colonies was transformed into CC107. The transformants were selected on LB plus chloramphenicol agar plates and streaked on glucose minimal agar plates supplemented with chloramphenicol, Pgal and Xgal.
C. CC107 with vector pCR2.1-TOPOCam alone without insert.
D–F. CC107 with vector pCR2.1-TOPOCam plus different *E. coli* genomic fragments.
G. A composite for the region that codes for DNA helicase UvrD, demonstrating the location and relative size of 19 independent genomic fragments from 28 isolates. The grey arrow indicates the annotated transcription direction for each ORF (Blattner *et al*., 1997).


A
B
C
D
E
F
G
dapF
yigA
xerC
yigB
uvrD
yigE
corA
yigF
#1
#105
#701
#83,#743
#521
#478
#8
#510
#618
#20
#705
#4,#704,#804
#308
#445,#617,#818
#493,#499
#357
#37
#603
#68,#340,#344,#443

A

Isolated Regions	Sections on *E. coli* genome	Independent Clones	Independent Inserts
1	AE000131-132	10	9
2	AE000161	2	2
3	AE000172	15	13
4	AE000213-215*	4	4
5	AE000226	4	4
6	AE000236-237	1	1
7	AE000256-257	9	9
8	AE000349-350	1	1
9	AE000352-353	10	10
10	AE000357	12	11
11	AE000357-358	4	4
12	AE000414	5	5
13	AE000442-443	18	16
14	AE000457	28	19
15	AE000457-458	2	2
16	AE000503	6	4
17	AE000507	1	1
Total 17		132	115

* Different from Genbank sequence.

B

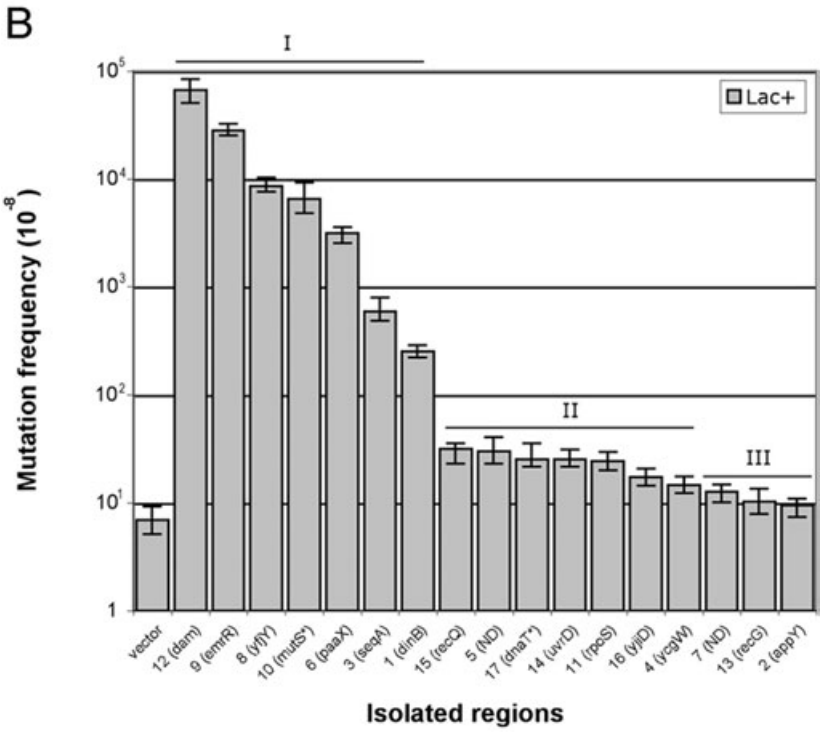


Fig. 2. Summary of isolated mutator DNA fragments.
A. Identified 17 regions on *E. coli* genome described in relation to the sections of annotated *E. coli* genome (Blattner *et al*., 1997), the number of total isolates and independent isolates.
B. The mutagenic potency of each region scored by measuring Lac$^+$ revertant frequency in liquid culture (Miller *et al*., 2002) leading to a classification into three groups as indicated (groups I, II and III). Vector alone without insert is included as a negative control. The error bars represent a 95% confidence range (Dixon and Massey, 1969). The mutator genes identified in the following section are indicated in parenthesis after their corresponding region number. Note: region 8 (*yfjY*) contains two nonsense codons in a downstream ORF *yfjZ* at positions K39/F40. Both mutS* and dnaT* are truncated. The mutator genes in regions 5 and 7 are not determined (ND). The specific isolate used from each region is listed in *Experimental procedures*.

Identification of genes responsible for the mutator effects

Although individual inserts typically covered three to four open reading frames (ORFs), multiple hits in the same region resulted in the elimination of several ORFs and left only one or two common ORFs as candidates for mutator genes. In frame nonsense codon mutations were introduced into the candidate ORFs on the plasmid (see

Experimental procedures) to generate truncated mutations, and the mutated plasmids were then transformed back to CC107 and tested for blue papillae formation. If the truncated mutations, which usually result in loss of activity, affected the mutator gene, the level of blue papillae formation would be reduced. As an example, the composite of region no. 9 from group I is shown in Fig. 3A. From five representative inserts, we identified two ORFs as candidates for mutator genes: *ygaH* (codes for a hypothetical protein 111 amino acids in length; Blattner *et al*., 1997) and *emrR* (codes for a transcriptional regulator that controls the multidrug-resistant pumps *emrA* and *emrB*, 176 amino acid residuals in length; Lomovskaya *et al*., 1995). The original isolate no. 47 was chosen for further study. In the isolate no. 47, two consecutive in frame nonsense codon mutations in either *ygaH* at positions corresponding to F70/V71 (no. 47 *ygaH8*) or *emrR* at positions corresponding to Q67/E68 (no. 47 *emrR8*) were introduced via polymerase chain reaction (PCR) and confirmed by DNA sequencing (Fig. 3A). The mutated plasmids, together with the original isolate no. 47 and the control vector, were transformed into CC107 and tested for blue papillae formation. Similar to all the other inserts in this region, the original insert no. 47 resulted in greatly elevated levels of blue papillae formation (Fig. 3A), which was correlated with a 2000-fold increase in the Lac$^+$ revertant frequency (Fig. 2B). The mutated plasmid, no. 47 *ygaH8*, generated about the same increase in the level of blue papillae formation (Fig. 3A). However, the mutated no. 47 with the in frame nonsense codons in the *emrR* gene, no. 47 *emrR8*, resulted in the elimination of the increase in papillae formation (Fig. 3A). Similar experiments were performed on another isolate no. 49 to reduce potential false positives resulting from DNA sequence errors introduced by PCR, and the results were essentially the same (data not shown). Therefore, we concluded that *emrR* is the mutator gene in region no. 9 of group I. This mutagenic activity is novel for *emrR*, a protein previously known for regulating multidrug resistance and microcins B17 and C7 production (del Castillo *et al*., 1990; Lomovskaya *et al*., 1995).

We used the introduction of nonsense mutations to identify the mutator genes in the other six regions of group I that had strong mutagenic effects (Fig. 3B). DNA polymerase IV (*dinB*) was the mutator gene in region no.1, as expected from previously published results (Kim *et al*., 1997; Wagner and Nohmi, 2000). In addition, we found two genes that are components of the *E. coli* DNA methyl-directed mismatch repair system, partially deleted *mutS* genes that generate N-terminal truncated forms of the methyl-directed mismatch repair protein MutS (region no.10) and the DNA adenine methylase gene, *dam*, which was also identified previously (region no.12; Herman and Modrich, 1981). In addition to *emrR*, we found three more genes that were novel for their mutagenic activities. *SeqA* (region no. 3) encodes a negative modulator of initiation of replication and a GATC-binding protein (Hiraga *et al*., 1998; Shakibai *et al*., 1998; Guarne *et al*., 2002); *b1399* (region no. 6) encodes a transcriptional regulator PaaX for phenylacetic acid degradation (Ferrandez *et al*., 2000); and *yfjY* (region no. 8) encodes a hypothetical protein (Blattner *et al*., 1997). It is interesting to note that the level of mutagenesis conferred by *yfjY* (region no. 8) was enhanced about 200-fold when one of the two downstream genes on the plasmid DNA insert, *yfjZ* and *ypjF*, was knocked out by in frame nonsense codon mutations. The mutator phenotype of no. 421 *yfjZ8* is shown in Fig. 3B.

The detailed descriptions of group II and III, which had moderate to weak mutagenic effects, are shown in Fig. 4. Our findings included three DNA helicase genes: *uvrD* (region no. 14, group II), *recQ* (region no.15, group II) and *recG* (region no. 13, group III); a C-terminal truncated form of primosomal component gene *dnaT* (region no. 17, group II), two stationary phase regulator genes: *rpoS* (region no. 11, group II; Tsui *et al*., 1997) and *appY* (region no. 2, group III); and two hypothetical ORFs: *yjiD* (region no. 16, group II) and *ycgW* (region no. 4, group II). Mutator genes in region no. 5 (group II) and region no. 7 (group III) were not identified.

Mutagenic effects in a recA*-deficient background*

We wanted to determine whether the mutagenic effects observed in this study resulted from the induction of the SOS system, a system that generates a low to moderate level of mutations on undamaged DNA through the induction of the error-prone DNA polymerases IV and V (Tang *et al*., 1999; Wagner *et al*., 1999). We therefore measured the Lac$^+$ revertant frequency of each region in a *recA*-deficient background (CC107 *recA*) that is deficient in both induction of SOS and homologous recombination (reviewed by Lusetti and Cox, 2002). The results are shown in Table 1. In two cases (regions 3 and 7), the *recA* derivatives displayed sharply reduced growth rates and could not be assayed. The mutator effect is essentially retained in all other members of group I. Only two members of group II (regions 4 and 11) and two from group III (regions 2 and 13) are even candidates to have their effects attributed to SOS induction.

Specificity of mutator genes

All 17 regions were identified by screening for frameshift mutators. We wanted to determine which of the identified mutator genes are specific for frameshift mutations and which have general mutator phenotypes such as causing both frameshift and base substitution mutations.

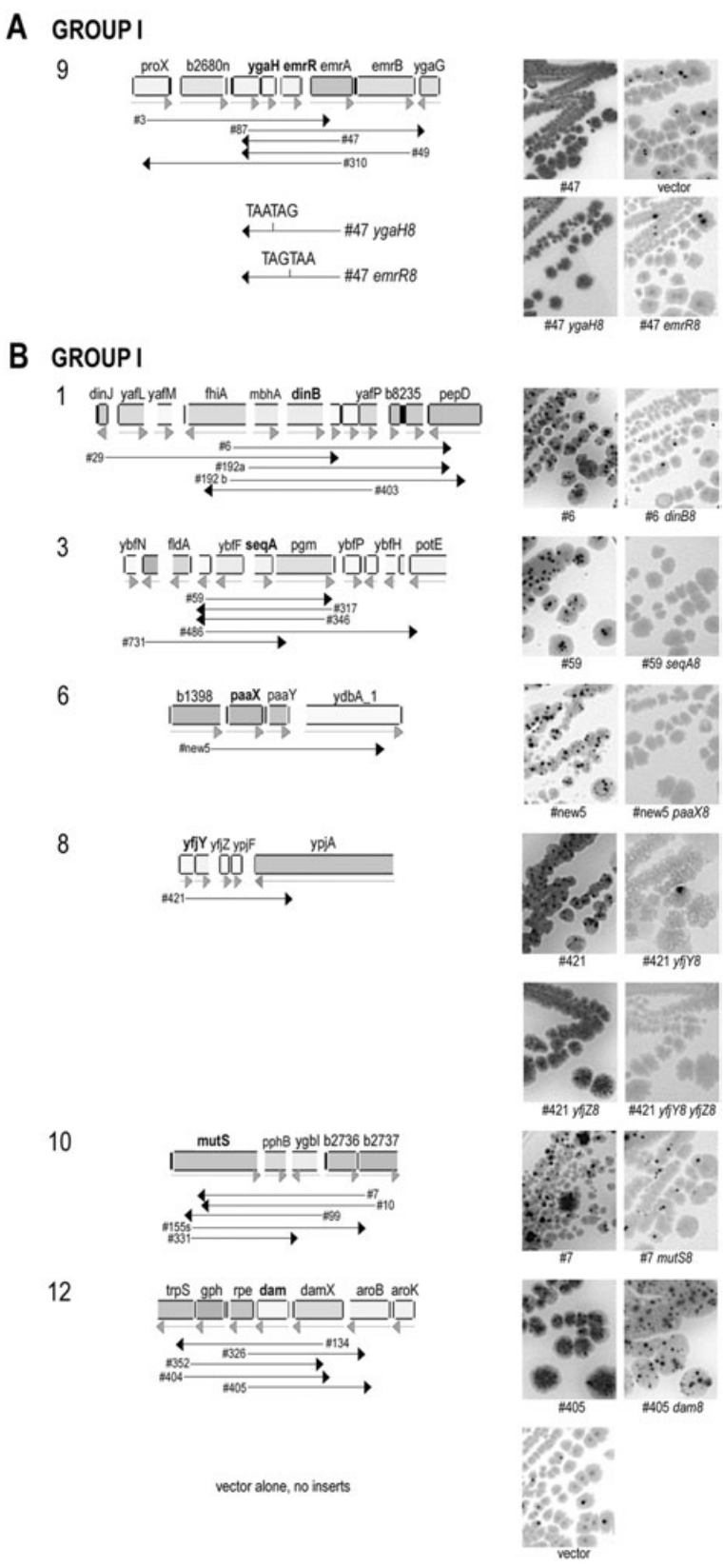


Fig. 3. Identification of mutator genes in group I.

A. A schematic diagram showing a composite of region no. 9 with the location, relative size and the insert orientation of five representatives; the positions for the two in frame stop codons in *ygaH* (*ygaH8*) and *emrR* (*emrR8*); blue papillation phenotypes with the original isolate no. 47, the control vector and two plasmids carrying mutations in *ygaH* (no. 47 *ygaH8*) and *emrR* (no. 47 *emrR8*). The grey arrow indicates the annotated transcription direction for each ORF (Blattner *et al*., 1997). The candidate mutator genes are bold labelled.

B. Identification of mutator genes in regions 1, 3, 6, 8, 10 and 12. A composite showing the location, relative size and the insert orientation of up to five representatives in each region of group I. Blue papillation phenotypes are shown for the identified mutator genes. The grey arrow indicates the annotated transcription direction for each ORF (Blattner *et al*., 1997). The mutator genes are bold labelled. The protein length and the positions for two in frame stop codons are as follows: DinB (351 aa, Q124/E125), SeqA (181 aa, K19/H20), PaaX (316 aa, Q104/T105), YfjY (160 aa, Q71/L72), MutS (N-terminal truncated forms: #7–756 aa, #10–633 aa, #99–827 aa, #155 s and #331–788 aa, E560/R561) and Dam (278 aa, E96/E97).

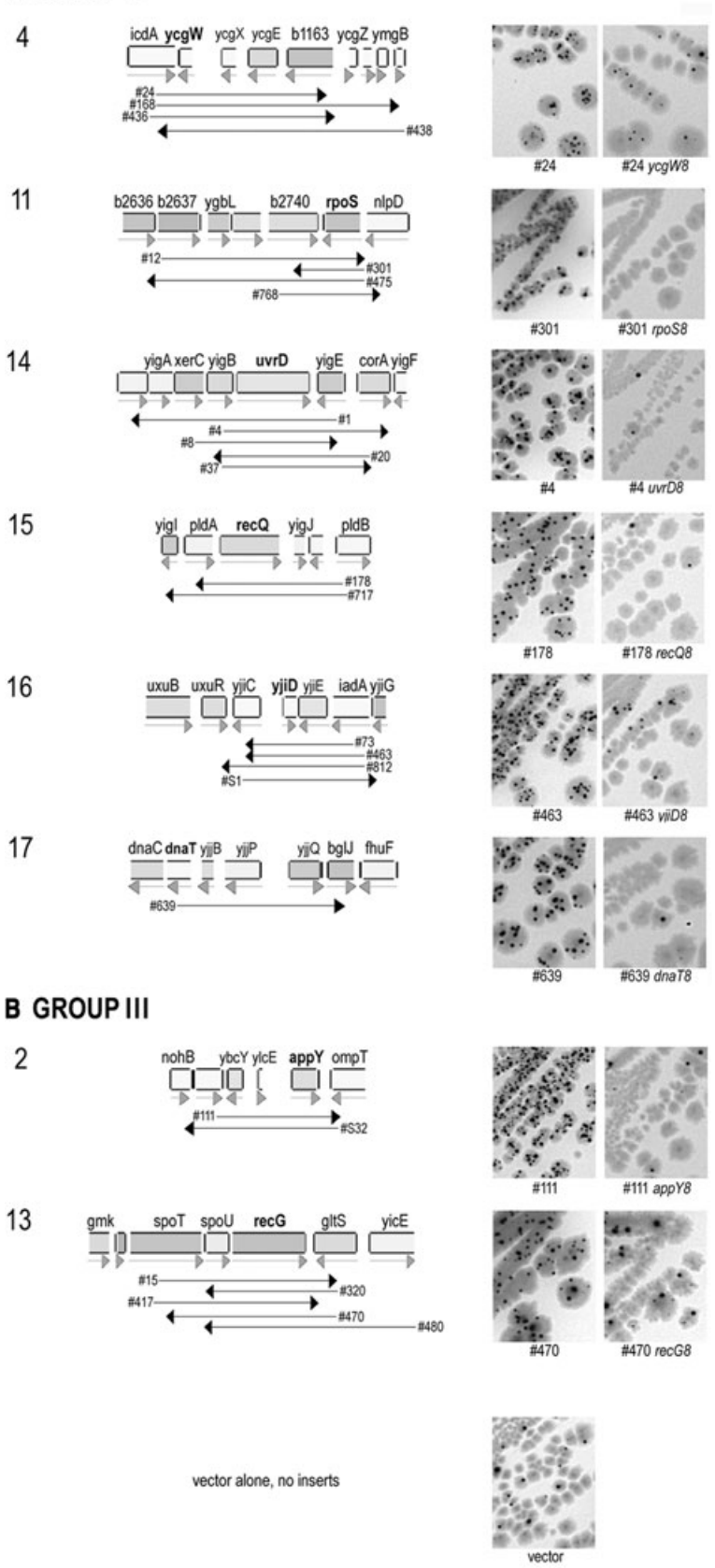


Fig. 4. Identification of mutator genes in group II (A) and group III (B). A composite showing the location, relative size and the clone orientation of up to five representatives in each region of groups II and III. Blue papillation phenotypes are shown for the identified mutator genes. The grey arrow indicates the annotated transcription direction for each ORF (Blattner *et al*., 1997). The mutator genes are bold labelled. The protein length and the positions for two in frame stop codons are as follows: YcgW (107 aa, I42/F43), RpoS (330 aa, E19/N20), UvrD (720 aa, K13/Q14), RecQ (610 aa, E355/E356), YjiD (133 aa, W54/Q55), DnaT (C-terminal truncated form: 70 aa, H20/Q21), AppY (249 aa, K121/K122) and RecG (693 aa, E180/E181).

Table 1. Mutagenic effects in a *recA*-deficient background.

	ORF[a]	Wild type[b]	Range[c]	*recA*[b]	Range[c]
Vector		6.9	5.1–9.3	18.9	13.1–23.3
Group I					
12 (#134)	*dam*	67 100	50 600–84 200	44 500	39 500–72 600
9 (#47)	*emrR*	28 200	25 200–32 600	47 900	36 100–59 100
8 (#421 *yfjZ8*)	*yfjY*	8 620	7600–10 230	40 963	37 589–47 510
10 (#7)	*mutS*[d]	6 500	4820–9270	27 560	21 921–37 429
6 (#new5)	*paaX*	3 120	2550–3570	1 480	967–1720
3 (#59)	*seqA*	591	485–797	–	–
1 (#6)	*dinB*	250	221–286	348	261–453
Group II					
15 (#178)	*recQ*	31.4	23.1–35.4	60	43.4–74.5
5 (#454)	ND[e]	29.8	23.0–40.3	70.6	50.8–117
17 (#639)	*dnaT*[d]	25.3	21.5–35.5	72.2	59.9–90.9
14 (#4)	*uvrD*	25.2	21.6–30.9	258	174–339
11 (#301)	*rpoS*	24.1	20.0–29.7	26.4	14.7–44.1
16 (#73)	*yjiD*	17.3	14.4–20.6	44.3	35.8–55.2
4 (#168)	*ycgW*	14.6	12.3–17.6	19.7	14.3–25.7
Group III					
7 (#349)	ND[e]	12.6	10.1–14.8	–	–
13 (#470)	*recG*	10.3	7.8–13.5	17.5	12.7–31.2
2 (#111)	*appY*	9.47	7.4–11.0	19.5	17.5–24.4

a. Identified mutator gene.
b. Lac$^+$ revertant frequency ($\times 10^{-8}$) was determined in liquid culture (see *Experimental procedures*).
c. The 95% confidence limits were calculated according to the method of Dixon and Massey (1969).
d. Truncated form.
e. Not determined.

To examine the mutagenic effects on base substitutions in each region, we measured the frequency of rifampicin resistance, Rif^r, as a result of base substitution mutations in the *rpoB* gene of *E. coli* that encodes for the beta subunit of RNA polymerase. CC107 containing the vector alone without any insert was used as a negative control. Our results show that groups II and III have essentially no detectable effect on the frequency of Rif^r mutants. In group I, however, five out of seven mutator genes have significantly increased frequencies of Rif^r mutants. These are *emrR*, *dam*, *seqA*, *dinB* and the partially truncated forms of *mutS* (Fig. 5). The Rif^r frequency was found to be five- to 125-fold higher. Two mutator genes in group I, *paaX* and *yfjY*, had limited effects on base substitution mutations (Fig. 5), although they had quite significant effects on frameshift mutations (Fig. 2B).

We determined the specificity of the base substitution mutations on the *rpoB* gene in the overproduction of *emrR* background. Our results show that it had a prominent hot-spot of an AT → GC transition at position 1547 accounting for 90% Rif^r (Fig. 6B). This hot-spot has been seen in the previously determined spectra for strains lacking either the functional nucleoside diphosphate kinase (*ndk*) (Fig. 6C; Miller *et al.*, 2002) or mismatch repair (*mutS*) (Fig. 6D; Miller *et al.*, 2002). Moreover, similar to the *ndk*-deficient strain, the overproduction of *emrR* causes much diminished satellite peaks of AT → GC transitions at positions

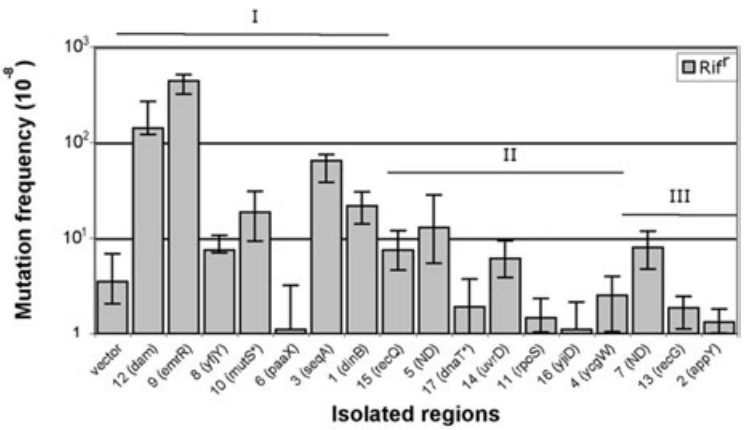


Fig. 5. Frequency of rifampicin (Rif)-resistant revertants. The experiments were done with the same group of cells used in Fig. 2B. Vector alone without insert is included as a control. The error bars represent a 95% confidence range (Dixon and Massey, 1969). Identified mutator genes in each region are indicated in parenthesis after their corresponding region numbers. Note: region 8 (*yfjY*) contains two nonsense codons in a downstream ORF *yfjZ* at positions K39/F40. Both mutS* and dnaT* are truncated. The mutator genes in regions 5 and 7 are not determined (ND). The specific isolate used from each region is listed in *Experimental procedures*.

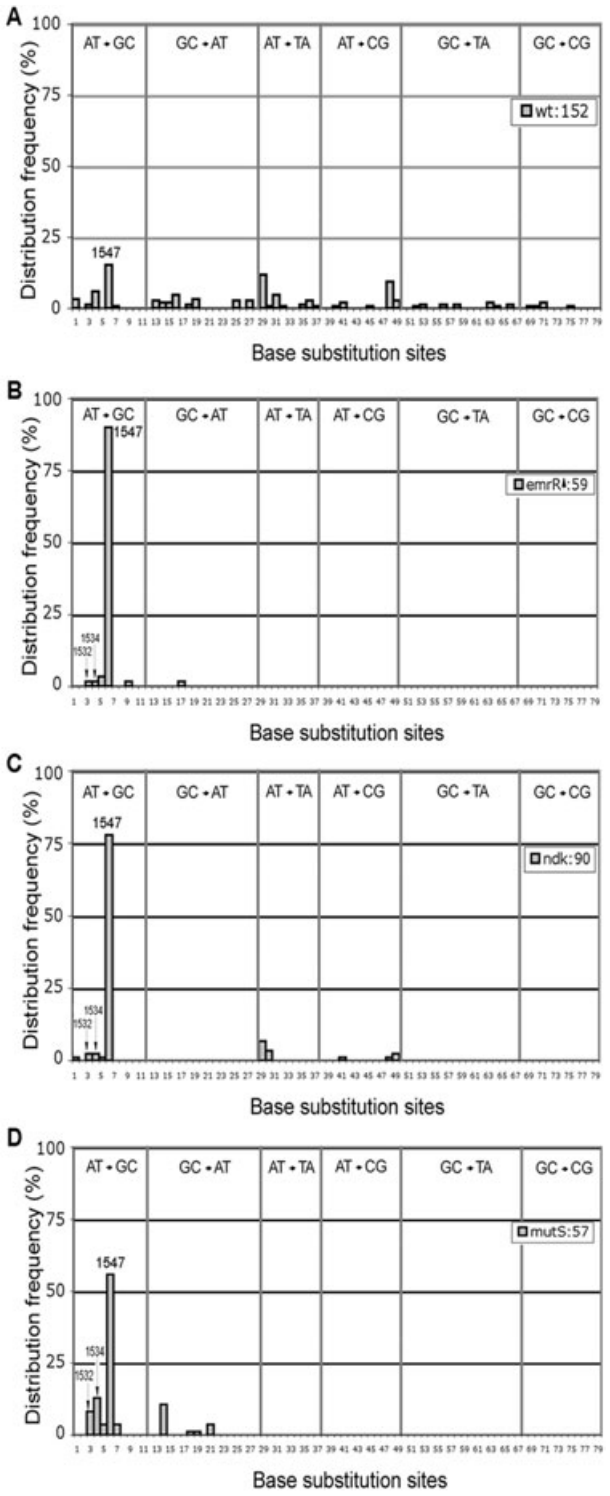


Fig. 6. Distribution of base substitution mutations in *rpoB*.
A. Wild type (Miller *et al*., 2002).
B. No. 47 (*emrR*↑).
C. *ndk* (Miller *et al*., 2002).
D. *mutS* (Miller *et al*., 2002).
The base substitution sites and their corresponding mutation positions on the *rpoB* gene are listed in *Experimental procedures*.

1532 and 1534 that were clearly the satellite peaks in the *mutS* background (Fig. 6B and D; Miller *et al*., 2002).

Discussion

How many *E. coli* genes cause elevated levels of mutations when overexpressed? In this study, with the aid of a frameshift tester strain, we used a multicopy cloning (MCC) approach to conduct a genome-wide search and identified a total of 115 independent genomic fragments that covered 17 regions on the *E. coli* chromosome. Further studies identified 15 mutator genes, among which 12 genes were not previously identified as causing mutator phenotypes when put on to a multicopy plasmid (Table 2). The annotation of the *E. coli* genome includes 4288 protein-coding genes, among which 38% are hypothetical (Blattner *et al*., 1997). However, the genome-wide MCC approach reported here allows us to identify a small subset of the genes, among which virtually all the genes with known function are involved in replication, recombination or repair. Therefore, this approach does not appear to generate artifacts and thus complements the conventional gene inactivation approach to finding mutators. Using *E. coli* as a model system, our results indicate that this method offers a way of finding DNA replication-, recombination- or repair-related components, such as error-prone polymerase (DinB), enzymes that are involved in marking the parental DNA strand (Dam and SeqA), DNA helicases (UvrD, RecQ and RecG) and a truncated mismatch repair component (MutS). The novel mutator genes (*paaX*, *emrR*, *appY*, *yfjY*, *yjiD*, *ycgW*) detected here are candidates for the structural or regulatory elements in DNA replication, recombination or repair.

We can speculate about the possible mutational mechanisms of several mutator genes identified in this study. One possible mutational mechanism is the ineffective methyl-directed DNA mismatch repair when *E. coli* overexpresses SeqA or truncated MutS. SeqA is a GATC-binding protein. It has been proposed that both Dam and SeqA control the methylation status of the cell, which is an important factor in forming and/or maintaining chromosome structure (Lobner-Olesen *et al*., 2003). As the methylation status of the cell is also critical for DNA mismatch repair, the results reported here indicate that the SeqA protein acts counter to Dam. Overexpression of SeqA may cause hypomethylation of DNA, leading to ineffective DNA mismatch repair. In the case of the mismatch repair protein MutS, all four N-terminal truncated forms of MutS lack the intact mismatch recognition domain (Obmolova *et al*., 2000). Therefore, overexpression of these truncated forms of MutS negatively complements MutS, perhaps by competing with wild-type MutS.

Can we define new mutational mechanisms by assaying for the novel mutator genes? One interesting finding is that

Table 2. List of identified mutator genes and their functions.

Region no.	ORF	Product	+1G frameshift	Base substitution
Genes previously identified to be mutagenic after overexpression				
1	*dinB*	Error-prone DNA polymerase DinB	Strong	Weak
11	*rpoS*	RNA polymerase, sigma S (sigma38) factor	Moderate	–
12	*dam*	DNA adenine methylase	Strong	Strong
Additional genes identified in this study				
Helicases				
13	*recG*	DNA helicase, resolution of Holliday junctions, branch migration	Weak	–
14	*uvrD*	DNA-dependent ATPase I and helicase II	Moderate	–
15	*recQ*	ATP-dependent DNA helicase	Moderate	–
Mismatch repair component				
10	*mutS*[a]	Methyl-directed mismatch repair protein	Strong	Moderate
DNA replication				
3	*seqA*	Negative modulator of initiation of replication, GATC-binding protein	Strong	Moderate
17	*dnaT*[a]	Primosomal component for DNA replication fork assembly and homologous recombination	Moderate	–
Stationary phase regulator				
2	*appY*	Regulatory protein	Weak	–
Phenylacetic acid degradation regulator				
6	*paaX*	Transcriptional repressor	Strong	–
Multidrug-resistant pump regulator				
9	*emrR*	Transcriptional repressor of *emrAB* operon	Strong	Strong
Hypothetical proteins				
4	*ycgW*	Hypothetical protein	Moderate	–
8	*yfjY*	Putative DNA repair protein	Strong	–
16	*yjiD*	Hypothetical protein	Moderate	–

a. Truncated form.

the overproduction of the multidrug resistance transcription regulator, EmrR, results in a large increase in frameshift and base substitution mutagenesis. The sequence analysis of the base substitution mutations on the *rpoB* gene shows that overproduction of *emrR* induces a prominent hot-spot of AT → GC transition mutations at position 1547, which is also the most prominent hot-spot in a *mutS* strain and in an *ndk*-deficient strain (Miller *et al*., 2002). Interestingly, our preliminary results on the global gene expression patterns after overexpression of *emrR* using DNA microarray technology show that the level of *ndk* expression is downregulated (unpublished data). The expression of EmrR is naturally inducible in the presence of certain drugs (Lomovskaya *et al*., 1995). The possible link between multidrug resistance and increased mutagenesis is tantalizing and will be the object of continuing studies.

It is also worth noting that one of the hypothetical proteins identified in this study, YfjY, is a putative DNA repair protein homologue of the RadC protein family. As a result of the recent finding that the original *E. coli radC* allele *radC102* is actually a *recG* allele (Lombardo and Rosenberg, 2000), the evidence for RadC being a DNA repair protein is minimal. Further study of the molecular mechanism of the mutator effect caused by YfjY may elucidate DNA repair functions of YfjY, RadC and two other *E. coli* putative RadC homologues, YeeS and YkfG.

Can this approach be used in other systems? In eukaryotes, overexpression of certain genes also causes mutator phenotypes. For example, increased expression of the mismatch repair protein MLH1 in yeast or of DNA polymerase beta in Chinese hamster ovary cells results in a mutator phenotype (Canitrot *et al*., 1998; Shcherbakova *et al*., 2001). It has also been shown that a transient excess of MYC activity causes genome instability and tumorigenesis (Karlsson *et al*., 2003). Therefore, with an appropriate mutator tester strain, it is possible to conduct similar mutator searches in eukaryotes. Given the fact that there is a close link between a mutator phenotype and cancer, one may find genes that, when overproduced, cause certain mutator phenotypes that contribute to the onset of cancer.

Experimental procedures

Bacteria strains and plasmids

The *E. coli* strain, J93, carries a deletion of the *lac* operon and appears to be wild type for all other markers. The frameshift tester strain, CC107, has been described previously (Cupples *et al*., 1990). It is *ara* Δ(*gpt-lac*)5 *thi*/F′128 *lacIZ*

proA$^+$B$^+$. It carries a frameshift mutation in *lacZ* on the F′ episome, which reverts only by insertion of a G in a run of six Gs. A *recA* derivative of CC107 was constructed by P1 transduction using a donor strain CSH126 (Miller, 1992) that contains a Tn*10* insert in *srl* with a deletion extending into *recA*. The pCR2.1-TOPOCam was modified from the vector pCR2.1-TOPO (Invitrogen) by insertion of a chloramphenicol-resistant gene at the PCR cloning site and elimination of a 1.6 kb *Pst*I fragment.

Construction of E. coli *genomic library*

Escherichia coli genomic DNA was prepared from the strain J93Δ*lac*. After partial digestion by the restriction enzyme *Sau*3AI, DNA fragments between 3 and 5 kbp in length were isolated from the agarose gel and ligated into the *Bam*HI site of the vector pCR2.1Cam using protocols described previously (Ausubel *et al*., 1987).

Detection and isolation of E. coli *mutators*

The random shotgun-generated genomic DNA library was electroporated into the CC107 strain using the protocol described previously (Ausubel *et al*., 1987). The electrotransformants were plated directly on the glucose minimal medium supplemented with phenyl-β-D-galactoside (Pgal) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal) as described previously (Nghiem *et al*., 1988). After a 4 day incubation at 37°C, the mutator candidates were identified as colonies that contained elevated levels of blue papillae.

To confirm the mutator phenotype, single colonies were obtained by streaking the candidate mutator colony on the papillae indicator plate, and these were used to prepare the plasmid. The insert size on each plasmid was estimated by agarose gel electrophoresis using a supercoiled DNA ladder as a standard (Invitrogen). Plasmids containing the inserts were transformed back into the CC107 host and retested for blue papillae formation.

After the mutator phenotype was confirmed, both 3′ and 5′ ends of each insert were sequenced using two vector primers, 5′-CAGGAAACAGCTATGAC-3′ and 5′-TGGCAGA AATTCGATGATAAGCT-3′. The identity of the insert was determined by BLAST analysis against the annotated genome sequence of *E. coli* K-12 (Blattner *et al*., 1997).

Site-directed mutagenesis

The nonsense mutations were introduced into the candidate ORFs on the plasmid using a Quick-change kit (Stratagene) with reagents and protocols supplied by the manufacturer with the exception of using DH5α as an *E. coli* host for transformation. The upper and lower mutant primers were synthesized (Invitrogen), and detailed sequence information can be obtained upon request. The mutations on the plasmid were subsequently confirmed by DNA sequencing. For each candidate mutator gene, two independent experiments were carried out to reduce the possibility of DNA polymerase errors during PCR. The plasmid carrying the desired nonsense mutations was transformed to CC107 to test for the mutator phenotype.

Determination of mutational frequencies

The mutational frequencies of Rif^r and Lac$^+$ revertants were determined as described previously (Miller *et al*., 2002). Mutational frequency was determined as the median frequency from a set of cultures (the number of cultures varied from 22 to 44). Ninety-five per cent confidence limits were determined according to the method of Dixon and Massey (1969). The specific isolate used from each region is as follows (region no.–isolate no.): 1–#6, 2–#111, 3–#59, 4–#168, 5–#454, 6–new5, 7–#349, 8–#421 *yfjZ8*, 9–#47, 10–#7, 12–#134, 13–#470, 14–#4, 15–#178, 16–#73 and 17–#639.

Determination of the specificity of the base substitution mutations on the rpoB *gene*

The base substitution mutation spectrum on the *rpoB* gene in the overproduction of *emrR* background was determined using the protocols described previously (Garibyan *et al*., 2003). The base substitution sites and their corresponding mutation positions on the *rpoB* gene are as follows (base substitution site–position on *rpoB*): AT → GC: 1–443, 2–1522, 3–1532, 4–1534, 5–1538, 6–1547,7–1552, 8–1577, 9–1598, 10–1703, 11–1715; GC → AT: 12–1520, 13–1535, 14–1546, 15–1565, 16–1576, 17–1585, 18–1586, 19–1592, 20–1595, 21–1600, 22–1601, 23–1609, 24–1610, 25–1691, 26–1708, 27–1721, 28–2060; AT → TA: 29–443, 30–1532, 31–1538, 32–1547, 33–1568, 34–1577, 35–1598, 36–1714, 37–1715; AT → CG: 38–437, 39–443, 40–1525, 41–1532, 42–1534, 43–1538, 44–1547, 45–1577, 46–1598, 47–1687, 48–1714, 49–1715; GC → TA: 50–436, 51–442, 52–444, 53–1527, 54–1535, 55–1537, 56–1546, 57–1565, 58–1576, 59–1578, 60–1586, 61–1592, 62–1595, 63–1600, 64–1601, 65–1708, 66–1721, 67–1585; GC → CG: 68–444, 69–1527, 70–1574, 71–1576, 72–1578, 73–1585, 74–1600, 75–1601, 76–1691, 77–1709, 78–1716, 79–2059.

Acknowledgements

J.H.M. was supported by a grant from the National Institutes of Health (ES0110875).

References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1987) *Current Protocols in Molecular Biology.* New York: John Wiley & Sons, pp. 5.0.5–5.11.2 and 1.8.4–1.8.5.

Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453–1474.

Bucci, C., Lavitola, A., Salvatore, P., Del Giudice, L., Massardo, D.R., Bruni, C.B., and Alifano, P. (1999). Hypermutation in pathogenic bacteria: frequent phase variation in meningococci is a phenotypic trait of a specialized mutator biotype. *Mol Cell* **3:** 435–445.

Cabrera, M., Nghiem, Y., and Miller, J.H. (1988) *mutM*, a second mutator locus in *Escherichia coli* that generates GC→TA transversions. *J Bacteriol* **170:** 5405–5407.

Canitrot, Y., Cazaux, C., Frechet, M., Bouayadi, K., Lesca, C., Salles, B., and Hoffmann, J.S. (1998). Overexpression of DNA polymerase beta in cell results in a mutator phenotype and a decreased sensitivity to anticancer drugs. *Proc Natl Acad Sci USA* **95:** 12586–12590.

del Castillo, I., Gomez, J.M., and Moreno, F. (1990) *mprA*, an *Escherichia coli* gene that reduces growth-phase-dependent synthesis of microcins B17 and C7 and blocks osmoinduction of *proU* when cloned on a high-copy-number plasmid. *J Bacteriol* **172:** 437–445.

Cupples, C.G., Cabrera, M., Cruz, C., and Miller, J.H. (1990) A set of *lacZ* mutations in. *Escherichia coli* that allow rapid detection of specific frameshift mutations. *Genetics* **125:** 275–280.

Denamur, E., Bonacorsi, S., Giraud, A., Duriez, P., Hilali, F., Amorin, C., *et al.* (2002) High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J Bacteriol* **184:** 605–609.

Dixon, W.J., and Massey, J.R.F.J. (1969) *Introduction to Statistical Analysis*. New York: McGraw-Hill.

Echols, H., Lu, C., and Burgers, P.M. (1983) Mutator strains of *Escherichia coli*, *mutD* and *dnaQ*, with defective exonucleolytic editing by DNA polymerase III holoenzyme. *Proc Natl Acad Sci USA* **80:** 2189–2192.

Farr, S.B., D'Ari, R., and Touati, D. (1986) Oxygen-dependent mutagenesis in *Escherichia coli* lacking superoxide dismutase. *Proc Natl Acad Sci USA* **83:** 8268–8272.

Ferrandez, A., Garcia, J.L., and Diaz, E. (2000) Transcriptional regulation of the divergent *paa* catabolic operons for phenylacetic acid degradation in *Escherichia coli*. *J Biol Chem* **275:** 12214–12222.

Funchain, P., Yeung, A., Stewart, J.L., Lin, R., Slupska, M.M., and Miller, J.H. (2000) The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154:** 959–970.

Garibyan, L., Huang, T., Kim, M., Wolff, E., Nguyen, A., Nguyen, T., *et al.* (2003) Use of the rpoB gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair (Amst)* **2:** 593–608.

Gibson, T.C., Scheppe, M.L., and Cox, E.C. (1970) Fitness of an *Escherichia coli* mutator gene. *Science* **169:** 686–688.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., and Taddei, F. (2001) Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* **291:** 2606–2608.

Glickman, B.W., and Radman, M. (1980) *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction. *Proc Natl Acad Sci USA* **77:** 1063–1067.

Gross, M.D., and Siegel, E.C. (1981) Incidence of mutator strains in *Escherichia coli* and coliforms in nature. *Mutat Res* **91:** 107–110.

Guarne, A., Zhao, Q., Ghirlando, R., and Yang, W. (2002) Insights into negative modulation of *E. coli* replication initiation from the structure of SeqA-hemimethylated DNA complex. *Nature Struct Biol* **9:** 839–843.

Heinen, C.D., Schmutte, C., and Fishel, R. (2002) DNA repair and tumorigenesis: lessons from hereditary cancer syndromes. *Cancer Biol Ther* **1:** 477–485.

Herman, G.E., and Modrich, P. (1981) *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. *J Bacteriol* **145:** 644–646.

Hiraga, S., Ichinose, C., Niki, H., and Yamazoe, M. (1998) Cell cycle-dependent duplication and bidirectional migration of SeqA-associated DNA-protein complexes in *E. coli*. *Mol Cell* **1:** 381–387.

Imlay, J.A., and Linn, S. (1987) Mutagenesis and stress responses induced in *Escherichia coli* by hydrogen peroxide. *J Bacteriol* **169:** 2967–2976.

Karlsson, A., Deb-Basu, D., Cherry, A., Turner, S., Ford, J., and Felsher, D.W. (2003) Defective double-strand DNA break repair and chromosomal translocations by MYC overexpression. *Proc Natl Acad Sci USA* **100:** 9974–9979.

Kim, S.R., Maenhaut-Michel, G., Yamada, M., Yamamoto, Y., Matsui, K., Sofuni, T., *et al.* (1997) Multiple pathways for SOS-induced mutagenesis in *Escherichia coli*: an overexpression of dinB/dinP results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. *Proc Natl Acad Sci USA* **94:** 13792–13797.

LeClerc, J.E., Li, B., Payne, W.L., and Cebula, T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274:** 1208–1211.

Lobner-Olesen, A., Marinus, M.G., and Hansen, F.G. (2003) Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proc Natl Acad Sci USA* **100:** 4672–4677.

Loeb, L.A., Loeb, K.R., and Anderson, J.P. (2003) Multiple mutations and cancer. *Proc Natl Acad Sci USA* **100:** 776–781.

Lombardo, M.J., and Rosenberg, S.M. (2000) radC102 of *Escherichia coli* is an allele of recG. *J Bacteriol* **182:** 6287–6291.

Lomovskaya, O., Lewis, K., and Matin, A. (1995) EmrR is a negative regulator of the *Escherichia coli* multidrug resistance pump EmrAB. *J Bacteriol* **177:** 2328–2334.

Lu, Q., Zhang, X., Almaula, N., Mathews, C.K., and Inouye, M. (1995) The gene for nucleoside diphosphate kinase function s as a mutator gene in *Escherichia coli*. *J Mol Biol* **254:** 337–341.

Lusetti, S.L., and Cox, M.M. (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* **71:** 71–100.

Maki, H., and Sekiguchi, M. (1992) MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* **355:** 273–275.

Marinus, M.G., and Morris, N.R. (1975) Pleiotropic effects of a DNA adenine methylation mutation (dam-3) in *Escherichia coli* K12. Mutat Res 28, 15–26.

Matic, I., Radman, M., Taddei, F., Picard, B., Doit, C., Bingen, E., *et al.* (1997) Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* **277:** 1833–1834.

Michaels, M.L., Cruz, C., and Miller, J.H. (1990) *mutA* and *mutC*: two mutator loci in *Escherichia coli* that stimulate transversions. *Proc Natl Acad Sci USA* **87:** 9211–9215.

Miller, J.H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Miller, J.H. (1996) Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu Rev Microbiol* **50:** 625–643.

Miller, J.H., Funchain, P., Clendenin, W., Huang, T., Nguyen, A., Wolff, E., *et al.* (2002) *Escherichia coli* strains (*ndk*) lacking nucleoside diphosphate kinase are powerful mutators for base substitutions and frameshifts in mismatch-repair-deficient strains. *Genetics* **162:** 5–13.

Nevers, P., and Spatz, H. (1975) *Escherichia coli* mutants *uvrD uvrE* deficient in gene conversion of lambda heteroduplexes. *Mol Gen Genet* **139:** 233–243.

Nghiem, Y., Cabrera, M., Cupples, C.G., and Miller, J.H. (1988) The *mutY* gene: a mutator locus in *Escherichia coli* that generates GC→TA transversions. *Proc Natl Acad Sci USA* **85:** 2709–2713.

Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407:** 703–710.

Oliver, A., Cantón, R., Campo, P., Baquero, F., and Blázquez, J. (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288:** 1251–1253.

Richardson, A.R., and Stojiljkovic, I. (2001) Mismatch repair and the regulation of phase variation in *Neisseria meningitidis*. *Mol Microbiol* **40:** 645–655.

Shakibai, N., Ishidate, K., Reshetnyak, E., Gunji, S., Kohiyama, M., and Rothfield, L. (1998) High-affinity binding of hemimethylated oriC by *Escherichia coli* membranes is mediated by a multiprotein system that includes SeqA and a newly identified factor, SeqB. *Proc Natl Acad Sci USA* **95:** 11117–11121.

Shcherbakova, P.V., Hall, M.C., Lewis, M.S., Bennett, S.E., Martin, K.J., Bushel, P.R., *et al.* (2001) Inactivation of DNA mismatch repair by increased expression of yeast MLH1. *Mol Cell Biol* **21:** 940–951.

Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., and Goodman, M.F. (1999) UmuD′ (2)C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc Natl Acad Sci USA* **96:** 8919–8924.

Treffers, H.P., Spinelli, V., and Belser, N.O. (1954) A factor (or mutator gene) influencing mutation rates in *Escherichia coli*. *Proc Natl Acad Sci USA* **40:** 1064–1071.

Tsui, H.C., Feng, G., and Winkler, M.E. (1997) Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. *J Bacteriol* **179:** 7476–7487.

Wagner, J., and Nohmi, T. (2000) *Escherichia coli* DNA polymerase IV mutator activity: genetic requirements and mutational specificity. *J Bacteriol* **182:** 4587–4595.

Wagner, J., Gruz, P., Kim, S.R., Yamada, M., Matsui, K., Fuchs, R.P., and Nohmi, T. (1999) The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol Cell* **4:** 281–286.