ELSEVIER

Gene 293 (2002) 133–140

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

# Damage-repair error-prone polymerases of eubacteria: association with mobile genome elements

Elizaveta A. Permina*, Andrey A. Mironov[1], Mikhail S. Gelfand[1]

*State Scientific Center GosNIIGenetika, 1 Dorozhny pr., Moscow 113545, Russia*

Received 6 February 2002; received in revised form 12 April 2002; accepted 14 May 2002
Received by E. Sverdlov

**Abstract**

It is known that *umuDC*-like operons encoding DNA polymerase V are often found in plasmids of gamma-proteobacteria. Here we demonstrate that homologous operons are associated with mobile genomic elements in Gram-positive bacteria as well. Using the comparative analysis of transcriptional regulatory signals, we suggest that genes encoding homologs of UmuC in prophages of *Bacillus subtilis* and transposons of *Enterococus faecalis* are regulated by DinR (SOS repressor). We also predict a functional link between the polymerase V-like proteins of *B. subtilis* and the protein family containing YolD, YozL and YqjX. © 2002 Published by Elsevier Science B.V.

*Keywords*: SOS response; Computational analysis; Error-prone DNA polymerase; Prophage; Transposone

## 1. Introduction

The SOS response is the bacterial reaction to UV-light and DNA-damaging chemicals. It can be described as the rapid mobilization of the DNA reparation systems. The best studied SOS-response system is that of *Escherichia coli* (Walker, 1996; Lewin et al., 1994; de Henestrosa et al., 2000). It is regulated mainly by proteins RecA and LexA. RecA is activated by single-strand DNA and cleaves LexA, which is the repressor of the SOS regulon. The cleavage leads to the release of LexA from its binding sites and expression of the SOS-response genes. The consensus of the LexA binding sites (SOS box) is a palindrome TACTGTATATATATACAGTA with strongly conserved shoulders and an AT-rich spacer.

An important member of the SOS regulon in gamma-proteobacteria is the highly efficient error prone DNA polymerase (polymerase V) encoded by the operon *umuDC*. The polymerase V complex is $UmuD_2'UmuC$, where the UmuD′ subunit is formed by the cleavage of UmuD by RecA (Tang et al., 1999). Genomes of gamma-proteobacteria contain one copy of the *umuDC* operon; in addition there are many plasmids encoding homologous operons. For most of these operons, the regulation by LexA was shown either in experiment or simply by observation of candidate SOS boxes in upstream regions (Elledge and Walker, 1983a,b; Lodwick et al., 1990; Walker, 1996). The C-terminal domain of UmuD is homologous to the C-terminal domain of LexA, whereas the N-terminal domain of UmuC is homologous to the N-terminal domain of another *E. coli* protein, DinP, which is called DNA-polymerase IV. DinP participates in the untargeted mutagenesis (Kim et al., 1997).

SOS-regulons of Gram-positive bacteria are much less studied. The repressor DinR is homologous to LexA, but the similarity in the DNA binding helix-turn-helix domain is very weak. Accordingly, the recognition signal, called the Cheo box, does not resemble the SOS box, although it is also a palindrome with an AT-rich spacer CGAACATATGTTCG (Cheo et al., 1991; Yasbin et al., 1991). No functional analogs of the DNA polymerases IV and V are known.

In a preliminary study we identified genes encoding members of the UmuC family in the genome of *Bacillus subtilis* (Gelfand and Mironov, 1999). The aim of this work was to analyse in detail the distribution of the polymerase V genes in Gram-negative and Gram-positive bacteria and to search systematically for candidate regulatory sites. It turned out that genes homologous to *umuC* belong to the SOS-regulons not only in *E. coli* and its relatives, but also in Gram-positive bacteria, and in many genomes they are associated with mobile elements (plasmids, prophages and transposons).

* Corresponding author. Tel.: +7-95-135-2041; fax: +7-95-132-6080.
*E-mail address:* epermina@mail.ru (E.A. Permina).
[1] Present address: Integrated Genomics – Moscow, P.O. Box 348, Moscow 117333, Russia

0378-1119/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0378-1119(02)00701-1

## 2. Data and methods

### 2.1. Methods

Comparative approach to analysis of transcriptional regulation in bacterial genomes is based on the assumption that sets of genes regulated by orthologous transcription factors are conserved in related genomes (Gelfand, 1999). Thus the candidate sites occurring upstream of orthologous genes are true, whereas false positives are scattered at random. Since we are interested in the distribution of specific genes subject to horizontal transfer, we have modified the basic approach, accepting candidate sites upstream of all homologs of the *umuD* and *umuC* genes. For the same reason, the orthology relationships were established using construction of trees instead of the simple symmetric best hit criterion.

The recognition profiles were constructed using the samples of experimentally verified sites from (Walker, 1996; Gelfand and Mironov, 1999). The positional nucleotide weights in profiles were defined as:

$$W(b,k) = \log[N(b,k) + 0.5] - 0.25 \sum_{i=\mathrm{A,C,G,T}} \log[N(i,k) + 0.5]$$

where $N(b,k)$ is the count of nucleotide $b$ at position $k$. The score of a candidate site was calculated as the sum of the respective positional nucleotide weights:

$$Z(b_1 \dots b_k) = \sum_{\kappa=1\dots\kappa} W(b_k, k)$$

where $k$ is a the length of the site. The base of the logarithm was chosen such that the $Z$-score. The $Z$-score can be used to assess the significance of an individual site. The profiles are given in Table 1.

Table 1
Profiles for the SOS box (a) and the Cheo box (b)

| A | C | G | T |
|---|---|---|---|
| (a) | | | |
| −0.01 | 0.03 | −0.15 | 0.12 |
| 0.23 | −0.05 | −0.01 | −0.17 |
| −0.15 | 0.46 | −0.15 | −0.15 |
| −0.15 | −0.15 | −0.15 | 0.46 |
| −0.15 | −0.15 | 0.46 | −0.15 |
| −0.13 | −0.13 | −0.05 | 0.30 |
| 0.28 | −0.07 | −0.14 | −0.07 |
| −0.22 | 0.07 | −0.22 | 0.37 |
| 0.23 | −0.31 | 0.10 | −0.03 |
| 0.08 | −0.07 | −0.19 | 0.18 |
| 0.15 | −0.14 | −0.05 | 0.04 |
| 0.01 | −0.01 | −0.11 | 0.11 |
| 0.23 | −0.05 | −0.17 | −0.01 |
| 0.05 | 0.03 | −0.21 | 0.13 |
| 0.22 | 0.15 | −0.31 | −0.07 |
| −0.15 | 0.46 | −0.15 | −0.15 |
| 0.46 | −0.15 | −0.15 | −0.15 |
| −0.15 | −0.15 | 0.46 | −0.15 |
| −0.11 | 0.01 | −0.11 | 0.21 |
| 0.12 | −0.15 | 0.01 | 0.01 |
| (b) | | | |
| −0.04 | 0.38 | −0.30 | −0.04 |
| 0.06 | −0.32 | 0.32 | −0.06 |
| 0.55 | −0.18 | −0.18 | −0.18 |
| 0.55 | −0.18 | −0.18 | −0.18 |
| −0.24 | 0.46 | −0.24 | 0.02 |
| 0.23 | −0.12 | 0.00 | −0.12 |
| −0.27 | 0.12 | −0.27 | 0.41 |
| 0.41 | −0.27 | 0.12 | −0.27 |
| −0.12 | 0.00 | −0.12 | 0.23 |
| 0.02 | −0.24 | 0.46 | −0.24 |
| −0.18 | −0.18 | −0.18 | 0.55 |
| −0.18 | −0.18 | −0.18 | 0.55 |
| −0.06 | 0.32 | −0.32 | 0.06 |
| −0.04 | −0.30 | 0.38 | −0.04 |

### 2.2. Data

Sequences of complete genomes of *Haemophilus influenzae* (Fleischmann et al., 1995), *Escherichia coli* (Blattner et al., 1997), *Vibrio cholerae* (Heidelberg et al., 2000), *Pseudomonas aeruginosa* (Stover et al., 2000), *Yersinia pestis* (Parkhill et al., 2001), *Xylella fastidiosa* (Simpson et al., 2000) and *Bacillus subtilis* (Kunst et al., 1997), plasmids and individual genes were extracted from GenBank (http://www.ncbi.nlm.nih.gov/genbank). Sequences of partially sequenced genomes of *Salmonella typhi* and *Klebsiella pneumoniae* were obtained from the TIGR web site (http://www.tigr.org).

### 2.3. Programs and databases

Genomic analyses (genomic scale similarity searches, site searches using profiles etc.) were made using GenomeExplorer (Mironov et al., 2000). ClustalX 1.5 (Thompson et al., 1997) and Phylip (Felsenstein, 1989) with default parameters were used for multiple alignment and for construction of phylogenetic trees. The latter were plotted using GeneMaster (Mironov, A.A., unpublished). Database similarity searches were made using BLAST (Altschul et al., 1997) at the NCBI web site (http://www.ncbi.nlm.nih.gov/BLAST).

## 3. Results and discussion

### 3.1. Phylogeny of DNA polymerases IV and V

The UmuC protein family is related to several families of proteins. One such a family is DinP/DinB found in Archaea, Eubacteria and Eukaryotae (Fig. 1, eukaryotic members are not shown in the tree). This family contains UvrX from *B. subtilis*, its paralogs, and homologs from other Gram-positive bacteria. Fig. 1 features the tree of proteins from the DinP and UvrX sub-families from Gram-positive and Gram-negative bacteria, and archaebacterium *Sulfolobus solfatar-*

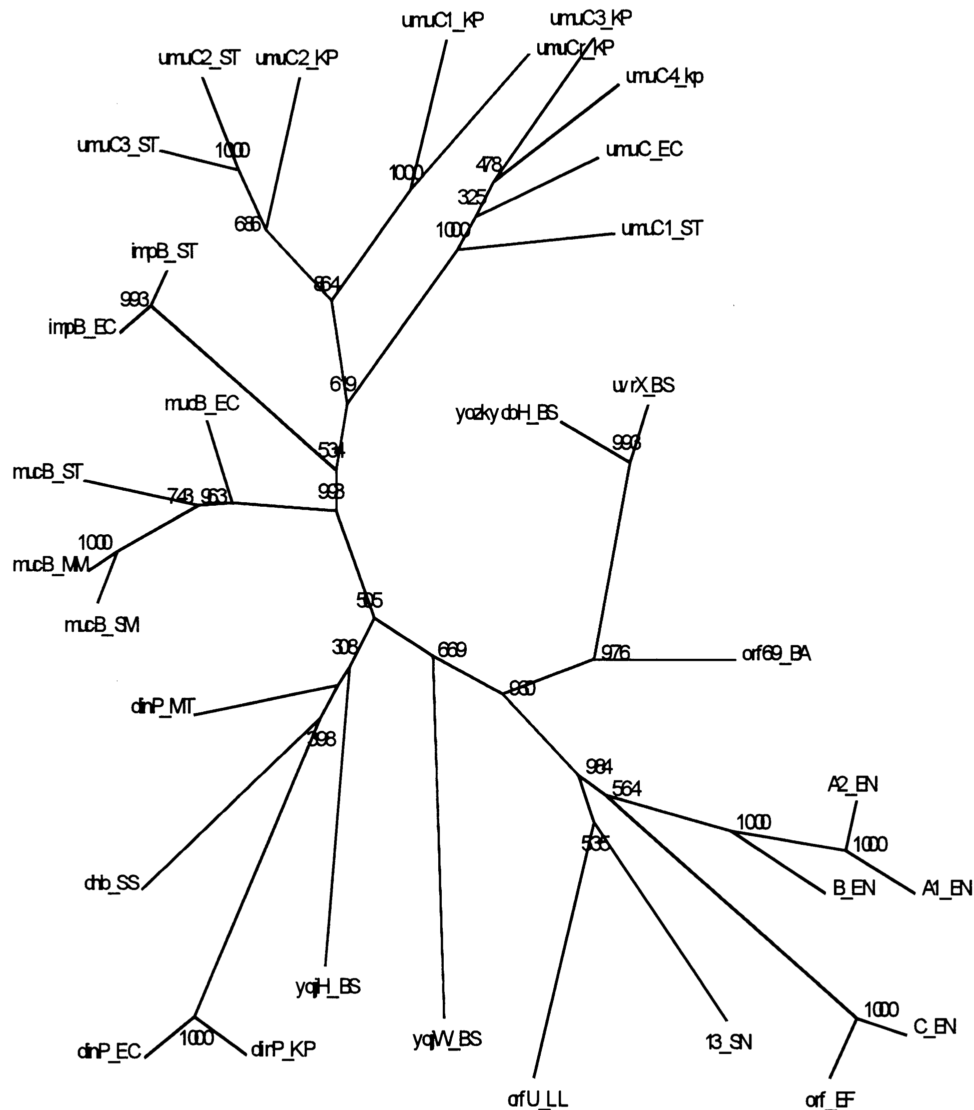


Fig. 1. Phylogenetic tree of UmuC homologs in different groups of organisms (Eubacteria: Firmicutes and gamma subdivision of Proteobacteria; Archaea). EC, *Escherichia coli*; KP, *Klebsiella pneumoniae*; ST, *Salmonella typhi*; MM, *Morganella morganii*; SM, *Serratia marcescens*; IJ, *Incj* plasmid; BS, *Bacillus subtilis*; BA, *Bacillus anthracis*; EN, *Enterococcus faecalis*; EF, *Enterococcus faecium*; SN, *Streptococcus pneumoniae*; LL, *Lactococcus lactis*; SS, *Sulfolobus solfataricus*; MT, *Mycobacterium tuberculosis*. Bootstrap values are shown on the nodes of the tree.

*icus*, the only one that has a homolog of the error-prone polymerase. The tree was constructed using the conserved portion of the protein alignment, cut at the end of *orf6* of *E. faecium* (GenBank AF076604) which is the shortest protein. High bootstrap values show that trees are stable and reliable. The overall length of the alignment is 222 positions. The tree is formed by three branches: UmuC-like proteins from Gram-negatives, UvrX-like proteins from Gram-positives (including a plasmid protein from *Bacillus anthracis*, GenBank AF188935, and a *Lactococcus lactis* protein from plasmid pNP40, GenBank LLU36837), and the branch of DinP proteins from eubacteria and *S. solfataricus* (the eukaryotic proteins also belong to this branch, not shown).

Two phylogenetic trees constructed for the members of the *umuDC* operon of gamma-proteobacteria are very similar (Fig. 2). Clustering of the UmuC and UmuD homologs is the same, with the only exception being the *rumAB* operon of the IncJ plasmid: RumA belongs to the branch of chromosome-encoded UmuD proteins from *E. coli* and *S. typhi*, whereas RumB clusters with UmuC from *K. pneumoniae* and ImpB from *S. typhi*.

### 3.2. Gamma-proteobacteria

All considered bacteria have standard SOS-response elements, the genes *recA* and *lexA*. Operons highly similar to *umuDC* were found in the genomes of *S. typhi* and *K. pneumoniae*. Both these genomes contain two copies of the *umuDC* homologs, each copy having an upstream candidate SOS boxes (Table 2), and several isolated copies of *umuC*

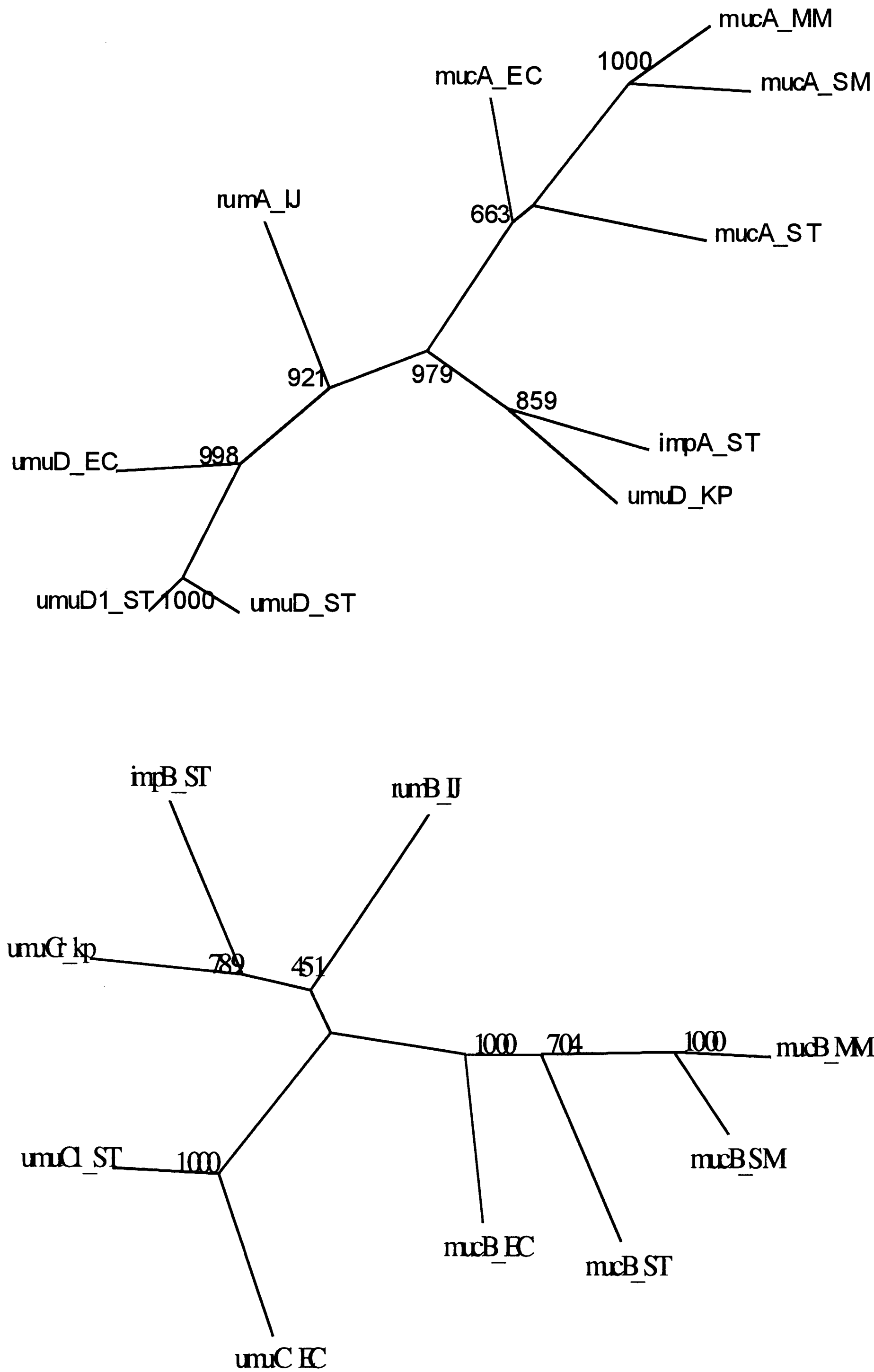


Fig. 2. Phylogenetic trees of *umuD* and *umuC* homologs from Gram-negative bacteria. For the legend see Fig. 1.

without upstream SOS boxes. *Y. pestis* and *V. cholerae* have no orthologs of *umuDC* but contain orthologs of *dinP* with upstream candidate SOS boxes. No orthologs of *umuDC* were found in the genome of *P. aeruginosa*. There are no candidate SOS boxes upstream of *dinP* of *P. aeruginosa*. Neither *umuDC*, nor *dinP* orthologs were found in the genomes of *H. influenzae* and *X. fastidiosa*.

Additional copies of *umuDC* homologs were found in plasmids of enteric bacteria and also in genomic fragments of *Morganella morganii*, *Serratia marcescens* and *Pseudomonas syringae*. Most of these genes have upstream SOS boxes (Table 2).

Thus the complete *umuDC* operons are consistently regulated by LexA, whereas the single *umuC* genes do not

Table 2
SOS boxes upstream of operons from gamma-proteobacteria and plasmids homologous to *umuDC* and *dinP*[a]

| Genome | Operon | (Candidate) site | Score | GenBank acc. no. |
|---|---|---|---|---|
| *E. coli* | *umuDC* | TACTGTATATAaAaACAGTA | 5.46 | |
| *K. pneumoniae* | ~ *umuDC1* | TACTGTATgcATAacCAGTA | 5.11 | |
| *K. pneumoniae* | ~ *umuDC2* | TACTGTATgcATAacCAGTA | 5.11 | |
| *S.typhi* | ~ *umuDC1* | TACTGgATATtTAacCAGTA | 5.03 | |
| *S.typhi* | ~ *umuDC2* | TACTGTATATAaAaACAGTA | 5.46 | |
| *Y. pestis* | ~ *dinP* | GACTGTATActTATACAGct | 4.70 | |
| *V. cholerae* | ~ *dinP* | TACTGTtTATgTATACAGTA | 5.09 | |
| *E. coli* plasmid pKM101 | *mucAB* | TACTGTATAaATAaACAGTt | 5.35 | D90147 |
| *M. morganii* plasmid R446b | *mucAB* | TACTGTATgTATAaACAGTt | 5.32 | U13611 |
| *S. marcescens* plasmid R471a | *mucAB* | CACTGTATgTATAaACAGTt | 5.23 | U13612 |
| *S. typhimurium* plasmid R394 | *mucAB* | TACTGTATAaATgTACAGct | 4.83 | AF039836 |
| *S. typhmurium* plasmid | *umuDC* | TgCTGTATATAaAaACAGgc | 4.63 | D90202 |
| *S. typhimurium* plasmid TP110 | *impCAB* | TACTGTATATAcATACAGcA | 5.32 | X53528 |
| Plasmid IncJ | *rumAB* | TACTGTATAaAcAaACAGTA | 5.34 | U13633 |
| Phage N15 | *umuD* | TACTGTATAaAaAaACAGTt | 5.25 | U13633 |

[a] *E. coli* plasmid pB171 contains *umuC* homolog, *impB*, but does not contain homologs of *umuD* or SOS-box sites. In most plasmids the SOS boxes were annotated in GenBank entries by similarity to the SOS-box consensus, although the LexA regulation was not verified in experiment.

belong to the SOS-regulon. Phylogenetic trees of the UmuC proteins are shown in Fig. 1. The chromosome-encoded proteins UmuC encoded by the *umuDC* operons from *E. coli*, *S. typhi* and *K. pneumoniae* are separated from the plasmid proteins.

### 3.3. Bacillus subtilis

The genome of *B. subtilis* contains four UvrX-like proteins. One of them, YqjH, is a clear orthologue of DinP (Fig. 1). The gene *yqjH* has no upstream candidate Cheo boxes. The gene *yqjW* has a pair of upstream candidate Cheo boxes. The two other genes, *uvrX* and *yozK*, are likely to be the second genes in operons (Fig. 3); upstream of these operons there are also pairs of candidate Cheo boxes (Table 3). It is likely that the two open reading frames annotated as independent genes *yozK* and *yobH* in fact form one gene with a frameshift introduced either by a sequencing error or a recent mutation. Anyway, the genes around *uvrX* and *yozKyobH* are pairwise homologous (Fig. 3) and both genes were shown to reside in prophage regions (Kunst et al., 1997).

The third gene, *yqjW*, is thus the original chromosomal orthologue of UvrX. It is noteworthy that the gene situated downstream of *yqjW*, namely *yqjX*, is homologous to the genes *yolD* and *yozL* situated upstream of *uvrX* and

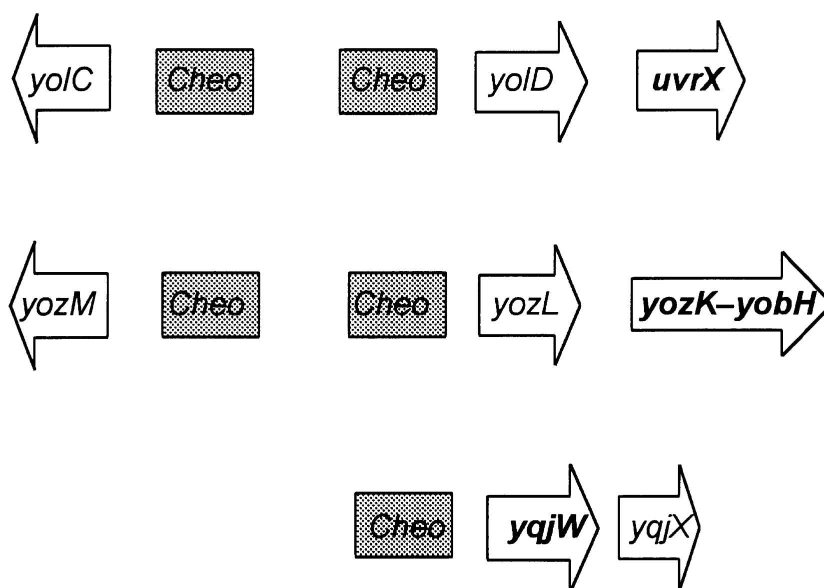


Fig. 3. Collinear gene chains containing homologs of *umuC* in *B. subtilis*. Genes homologous to *uvrX* are marked bold.

Table 3
Cheo boxes upstream of *B. subtilis*, *E. faecalis* and *E. faecium* operons containing genes homologous to *uvrX*

| Genome | Operon | (Candidate) site | Score | GenBank acc. no. |
|---|---|---|---|---|
| *B. subtilis* | yozLyozKyobH | CGAACtTtTGTTCt | 4.35 | |
| *B. subtilis* | *yolDuvrX* | CGAACtTtTGTTCt | 4.35 | |
| *B. subtilis* | yqjWyqjX | CGAACATActTTCG | 4.87 | |
| *B. anthracis* plasmid pX02 | *pX02-69* | aGAACgTAgGTTCG | 4.80 | AF188935 |
| *E. faecalis* | A1 | aGAACgTtaGTTCG | 4.12 | |
| *E. faecalis* | A2 | aGAACgTtaGTTCG | 4.12 | |
| *E. faecalis* | B | aGAACATtcGTTCG | 4.47 | |
| *E. faecalis* | C | CGAACATAcGTTtG | 5.31 | |
| *E. faecium* | orf2-3-4 | CtAACtTAcGTTtG | 4.58 | AF076604 |

*yozKyobH*, respectively (the alignment of the three genes is shown in Fig. 4). The different order of genes in the operons proves that the prophage operons are not a result of a duplication of the chromosomal operon *yqjWX* and hence suggests functional coupling of the UvrX/YqjW proteins and the YqjX/YolD proteins.

No homologs of UmuD were found in Gram-positive bacteria. Thus it is very tempting to predict that the proteins of the YqjX/YolD/YozL family are functionally equivalent to the UmuD subunit of the polymerase V from Gram-negatives. However, since no representatives of this family could be found in other Gram-positive bacteria, and thus this protein seems to be dispensable, this prediction requires strong experimental support. The regions upstream of the *yolD-uvrX* and *yozL-yozKyobH* operons are very similar (more than 90% identity). Thus it is likely that these two loci were introduced either by a recent duplication, or by independent recent integration of phage DNA.

### 3.4. Enterococus faecalis

A candidate Cheo box was observed upstream of a gene from the *E. faecalis* plasmid pAD1 (Ozawa et al., 1997). The gene, named *uvrA*, had about 20% identity to *umuC* of *E. coli*. On the other hand, the similarity of this gene to *uvrA* genes from both *E. coli* and *B. subtilis* is insignificant. Thus *E. faecalis* is one of the genomes that has a plasmid copy of umuC-like protein. The sequenced fraction of the *E. faecalis* genome contains four genes that are homologous to *umuC* and *yqjW*. Two of them (denoted A1 and A2) are very similar, whereas the two others (B and C) are more diverged. A2 and C are located in one genomic locus. All four genes contain candidate Cheo boxes in the upstream regions (Table 3) which otherwise show no significant similarity (Fig. 5). The four genes are clustered in the phylogenetic tree with a candidate gene from *Enterococcus faecium*, the closest relative of the latter gene is the gene C.

The loci around the *uvrX* family genes of *E. faecalis* are schematically shown in Fig. 6. In all cases these loci contain genes encoding proteins involved in gene mobility, namely resolvases and two types of transposases. Other genes forming the collinear chains are homologous to the pheromone binding protein gene *traB* from *E. faecalis* (GenBank U0068), the plasmid copy control protein gene *rapA* from *Lactobacillus reuteri* (GenBank AF036766), and the replication protein A gene *repA* from *E. faecalis* plasmid pAD1 (GenBank L01794). The transposases of the first type are quite common in *E. faecalis* (11 copies in the 4.5 Mb sequenced fraction of the genome). Since the collinear chains are flanked by copies of the transposase genes, it is highly likely that the described loci are transposons. On the other hand, collinearity of the genes is not perfect. Thus the complete characterization of this transposon family will be possible only after the sequencing of the *E. faecalis* genome is completed.

### 3.5. Other Gram-positive bacteria

Homologs of *uvrX* were found in the *Bacillus anthracis* plasmid pX02, the *Lactococcus lactis* plasmid pNP40, and the transposon Tn5252 of *Streptococcus pneumoniae* (GenBank L29324). Genes from the *L. lactis* plasmid and the *S. pneumoniae* transposon lack upstream candidate Cheo boxes, whereas the *B. anthracis* gene has a candidate site (Table 3). The proteins encoded by the Cheo box-less genes cluster together on the phylogenetic tree, as they contain

```
yozL  MML-EQLIQLKQDLIDGSKVEKPSLDDKQIDEMDILVSEALEFNKELKFKLFNKGFVENVTGRVH-YINFEQQKLHVKDQNDNTVYINMNNIIRVI
yolD  MMLPEHLTQLKQDLIDVSKIEKPSLDDQQIEEMDILVSEALEFNKELQFKLFHNGFVENVTGRVH-YINFEQQKLHVKDQNDNTVYINMNNIIGVT
yqjX  MFLPEHKQSLLERKRLKQKLQKPILDPDKLEEMNQTLCAAMEFAQDITVSCFQDGEIVCCTGKICRYEEFEKAVWIKGDE-DQLYKLKLDQVLDIV
      *:*:*:   .* :      .*::** ** .:::**:  :. *:** ::: .. *:.* :   **::  * :**:       *: *:   ::::::: :
```

Fig. 4. Alignment of YqjX, YolD and YozL proteins of *B. subtilis*.

```
A1  ATTCGGAAAACTGAAATTTTGTTGTTGCAAAACAGAACGTTAGTTCGTATAATG--TAACTCGAAAGGACGGTTTTTTGCATGAAA
A2  ATTCGGAAAACTGAAATTTTGTTGTTGCAAAAAAGAACGTTAGTTCGTATAATG--TAACTCGAAAGGACGGTTTTTTGCATGAAA
C   -TTGTGTGTTTAGCTATTTCTTCTTTACAATATCGAACATACGTTTGTATAATG--TAAAAAGAAGGG---GTGTTATTCATG---
B   -TACGGAAAACTGAATTTTACTAGTTGCAAAAGAGAACATTCGTTCGTATAATGGGTTAGTAGAAAGAA-GGGATTTTGGATGAAT
     *   *      *   ***  *  ** *** *  **** *  *** ********  * *   *** *    *  ** *  ***
```

Fig. 5. Multiple alignment of the upstream regions of *umuC* homologues from *E. faecalis*.

specific inserts of 27 (in Tn 5252) and 34 amino acids (in pNP40) in the conserved N-domain.

### *3.6. Overview: regulated and non-regulated members of the umuC/uvrX family*

The error-prone DNA polymerase V is a special mechanism for rapid reaction to the large-scale DNA damage (Tang et al., 1999). It is dangerous to the cell in the normal state. Thus the polymerase V is strictly regulated. Indeed, in *E. coli*, the polymerase V activity is under dual control: the operon is transcribed only after release of the repression by LexA, and the active multisubunit complex is formed after the cleavage of UmuD by RecA. The *umuC*-like genes in *S. typhi* and *K. pneumoniae* lacking for SOS boxes constitute monocistronic operons (without *umuD*) and thus cannot produce the active polymerase V without UmuD expressed from other operons. It is noteworthy that in all cases the *umuDC*-like operons have upstream candidate SOS boxes, as does the *umuD* gene from phage N15 (see Table 2 for the site; GenBank AF064539). All members of the *uvrX* subfamily have upstream Cheo boxes. The two exclusions, the genes from the *L. lactis* plasmid and the *S. pneumoniae* transposon, encode proteins with an insert. The function of these proteins is probably altered or modified to prevent possible damage to the genome.

## 4. Conclusions

We have demonstrated that polymerase V genes are associated with mobile genetic elements not only in Gram-negative bacteria, but in Gram-positive ones. In most cases these genes are known or predicted to be under transcriptional control by repressors LexA or DinR. Exceptions are rare and can be explained by control on the protein level. The types of mobile elements are quite diverse: plasmids, phages and prophages, transposons. It could mean that there is some specific mechanism linking genome rearrangements and error-prone repair. Finally, we tentatively assign a polymerase-related function to the family of *B. subtilis* proteins containing YolD, YozL and YqjX.

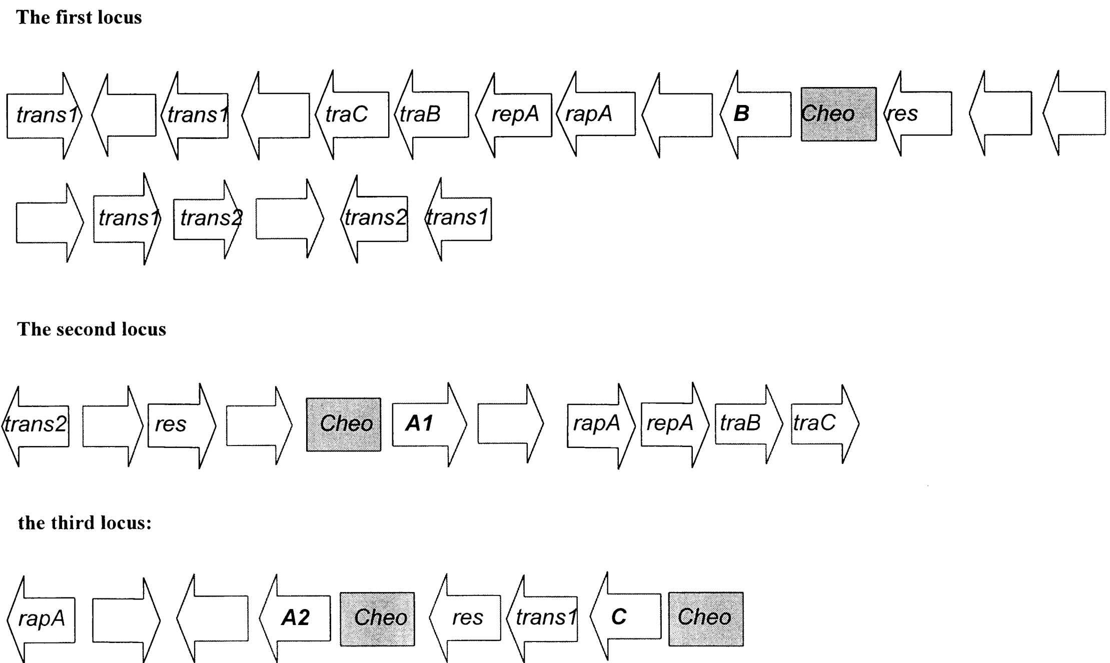


Fig. 6. *E. faecalis uvrX*-like genes are associated with transposon genes.

## Acknowledgements

This study was supported by grants from the Russian Fund of Basic Research (99-04-48247 and 00-15-99362), INTAS (99-1476) and the Howard Hughes Medical Institute (55000309). The unfinished bacterial genomes were obtained from the Web sites of The Institute of Genome Research (http://www.tigr.org) and the Sanger Center (http://www.sanger.ac.uk). We are grateful to Alexandra Rakhmaninova for discussions and critique.

## References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Blattner, F.R., Plunkett 3rd, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1474.

Cheo, D.L., Bayles, K.W., Yasbin, R.E., 1991. Cloning and characterization of DNA damage-inducible promoter regions from *Bacillus subtilis*. . J. Bacteriol. 173, 1696–1703.

Elledge, S.J., Walker, G.C., 1983a. The muc genes of pKM101 are induced by DNA damage. J. Bacteriol. 155, 1306–1315.

Elledge, S.J., Walker, G.C., 1983b. Proteins required for ultraviolet light and chemical mutagenesis: identification of the products of the umuC locus of *Escherichia coli*. J. Mol. Biol. 164, 175–192.

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

Felsenstein, J., 1989. PHYLIP – Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Gelfand, M.S., 1999. Recognition of regulatory sites by genomic comparison. Res. Microbiol. 150, 755–771.

Gelfand, M.S., Mironov, A.A., 1999. Computer analysis of regulatory signals in complete bacterial genomes. LexA and DinR binding sites. Mol. Biol. 33, 680–686.

Heidelberg, J.F., Eisen, J.A., Nelson, W.C., et al., 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406, 477–483.

de Henestrosa, A.R., Ogi, T., Aoyagi, S., Chafin, D., et al., 2000. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. Mol. Microbiol. 35, 1560–1572.

Kunst, F., Ogasawara, N., et al., 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390, 249–256.

Kim, S.-R., Maenhaut-Michel, G., Yamada, M., Yamamoto, Y., Matsui, K., Sofuni, T., Nohmi, T., Ohmori, H., 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.

Lewin, K.L., Harlow, G.R., Gregg-Jolly, L.A., Mount, D.W., 1994. Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in *Escherichia coli*. J. Mol. Biol. 241, 507–523.

Lodwick, D., Owen, D., Strike, P., 1990. DNA sequence analysis of the imp UV protection and mutation operon of the plasmid TP110: identification of a third gene. Nucleic Acids Res. 18, 5045–5050.

Mironov, A.A., Vinokurova, N.P., Gelfand, M.S., 2000. Software for analysis of bacterial genomes. Mol. Biol. 34, 222–231.

Ozawa, Y., Tanimoto, K., Fujimoto, S., Tomita, H., Ike, Y., 1997. Cloning and genetic analysis of the UV resistance determinant (*uvr*) encoded on the *Enterococcus faecalis* pheromone-responsive conjugative plasmid pAD1. J. Bacteriol. 179, 7468–7475.

Parkhill, J., et al., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413, 523–527.

Simpson, A.J., Reinach, F.C., Arruda, P., et al., 2000. The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis. Nature 406, 151–157.

Stover, C.K., Pham, X.Q., Erwin, A.L., et al., 2000. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. Nature 406, 959–964.

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

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Walker, C.G., 1996. The SOS response of *Escherichia coli*. *Escherichia coli* and *Salmonella*, Cellular and Molecular Biology, Vol. 1. ASM Press, Washington, DC, pp. 1400–1412 Chapter 89.

Yasbin, R.E., Cheo, D.L., Bayles, K.W., 1991. The SOB system of *Bacillus subtilis*: a global regulon involved in DNA repair and differentiation. Res. Microbiol. 142, 885–892.