BIOLOGIJA. 2006. Nr. 4. P. 5–9
© Lietuvos mokslų akademija, 2006
© Lietuvos mokslų akademijos leidykla, 2006

Structural organization of the genomic region between genes *30* and *31* in T4-related bacteriophages LZ, LZ1 and LZ9

Laura Kalinienė*,

Lidija Truncaitė,

Aivaras Arlauskas,

Giedrė Stoškienė,

Aurelija Zajančkauskaitė,

Rimas Nivinskas

Institute of Biochemistry, Mokslininkų 12, LT-08662 Vilnius, Lithuania

We have determined the nucleotide sequence and organization of the genomic region between genes *30* and *31* in the T4-related bacteriophages LZ, LZ1 and LZ9. In case of T4, this region has been shown to contain eleven coding sequences and several regulatory elements, including two early, two middle and one late promoters. The sequence analysis of LZ, LZ1 and LZ9 revealed that the most diverged part of the genomic region *30–31* in these phages lies within the early gene cluster *30.9–30.3*. A new orthologue of T4 *30.8* was revealed in LZ1. Two individual open reading frames, *30.4A* and *30.4B*, replaced the counterpart of T4 gene *30.4* in LZ, while any counterparts of T4 genes *30.8* and *30.5* were absent at all. No analogues of T4 genes *30.7* and *30.8* were found in phage LZ9. One of two early promoters was absent in LZ and LZ9. Moreover, the intergenic region between genes *rIII* and *30.9* of LZ9 appears to contain the 153 nt insertion, forming new hypothetical ORF.

Key words: T4-related bacteriophages, phage LZ, phage LZ1, phage LZ9, genomic region *30–31*

INTRODUCTION

Bacteriophages are now recognized to be among the most numerous entities in the biosphere [1]. Despite the fact that T4-type phages appear to be an extremely widespread group of bacteriophages, with approximately 160 representatives reported in literature to date [2], complete genomes of only few T4-type bacteriophages are determined and available yet (http://phage.bioc.tulane.edu). The genomes of the T4-related phages share a common set of essential genes, but contain genomic loci with unique uncharacterized genes. Some of these phages have evolved different strategies for the control of gene expression [3, 4]. Although phylogenetically related, these phages can be subdivided into four groups: T-evens, pseudo T-evens, schizo T-evens and exo T-evens [1].

While best studied, bacteriophage T4 has still numerous genomic regions with the poorly characterized genes. Most of them are smaller-than-average genes that express in the early period of infection [5]. One of the genomic regions carrying uncharacterized genes is the region between two essential genes: DNA ligase gene *30* and capsid assembly gene *31*. This genomic region contains ten open reading frames (ORFs) as well as gene *rIII* [6–9]. The region is also known to harbor several regulatory elements: late promoter P_LrIII [10, 11] (located immediately downstream the termination codon of gene *31*), the rho-independent terminator (right after gene *30.9*) [6, 12], two strong early promoters (for transcription units starting from genes *30.8* or *30.7*) [6, 12–14], two middle promoters P_M30.2 and P_M30 (located upstream genes *30.2* and *30*, respectively) [15] and the putative stem-loop after gene *30.7*, which has been deduced from the DNA sequence. It is worth mentioning that all yet uncharacteriz-ed genes found in this genomic region were reported to be non-essential and to have no homologues in the NCBI protein database. Following the consideration that non-essential genes might be important for the adaptation of phages in their particular life-style [1], any orthologues of these genes might be a valuable tool for gene's function clarification, leading to a possible understanding of the whole process of biological adaptation of a phage itself. The genomic comparison of T4 and various T4-related phages is really helpful in this case.

The aim of our study was to compare the structural organization of the genomic region between genes *30* and *31* of phage T4 and phylogenetically relative bacteriophages LZ, LZ1 and LZ9. Furthermore, we wanted to check whether the regulatory elements found in T4 are also present in the phages tested.

MATERIALS AND METHODS

Phages and bacterial strains. T4-related phages LZ, LZ1 and LZ9 were kindly provided by Dr. K. Carlson.

* Corresponding author. E-mail: laura@bchi.lt

All phages were grown in *Escherichia coli* B^E (*sup⁰*) kindly supplied by Dr. L. W. Black. Phages were isolated in the Denver area from a sewage plant [16].

PCR and sequencing procedures. Initially, the DNA fragments of the T4-related bacteriophages were amplified by PCR using T4-specific primers. Later, primers based on the obtained LZ1 DNA sequences were synthesised and used to fill gaps of DNA sequences in the genomic region tested. The DNA templates for the PCR were denatured phage particles. The PCR was carried out as described [17]. The sequencing reactions were carried out using a CycleReader™ DNA sequencing kit (Fermentas AB). The oligonucleotide primers for the sequencing reactions were 5'-end labelled by T4 polynucleotide kinase (Fermentas AB) with [$g^{32}P$]ATP (Amersham Biosciences). Phage-specific primers for PCR and sequencing procedures were used as follows.

T4-specific primers: Pr.1, 5'–GCTCTGGGTCTTAA GCAGCC (252–272 nt of gene *31*); Pr.2, 5'-CAGAAG AACGTGACTACCGAG (100–121 nt of the non-coding region between genes *31* and *rIII*); Pr.3, 5'–CGTTG CAGTTCAAGAGCGTG (complementary to 18–38 nt of gene *rIII*); Pr.4, 5'-GCCTCTGGCTTCTGCTGCACG (191–212 nt of gene *rIII*); Pr.5, 5'-GCCAGCGCGTTTA GAATCACC (complementary to 45–66 nt of gene *30.9*); Pr.6, 5'-GAGCACGTGCGGTTCTTCGAG (121–142 nt of gene *30.9*); Pr.7, 5'-GTTCAATTTCGCATACTGATCC (complementary to 87–109 nt of gene *30.7*); Pr.8, 5'–CAGAGCTAAACCTGAGCATC (303–322 nt of gene *30.7*); Pr.9, 5'–GTGCAGGAAGGTCTTTAGGTGC (complementary to 61–82 nt of gene *30.6*); Pr.10, 5'–CGTACTGTAAATGGGCGTGATAC (144–167 nt of gene *30.6*); Pr.11, 5'–GAGGCTTATATTGTTTCCTGGG

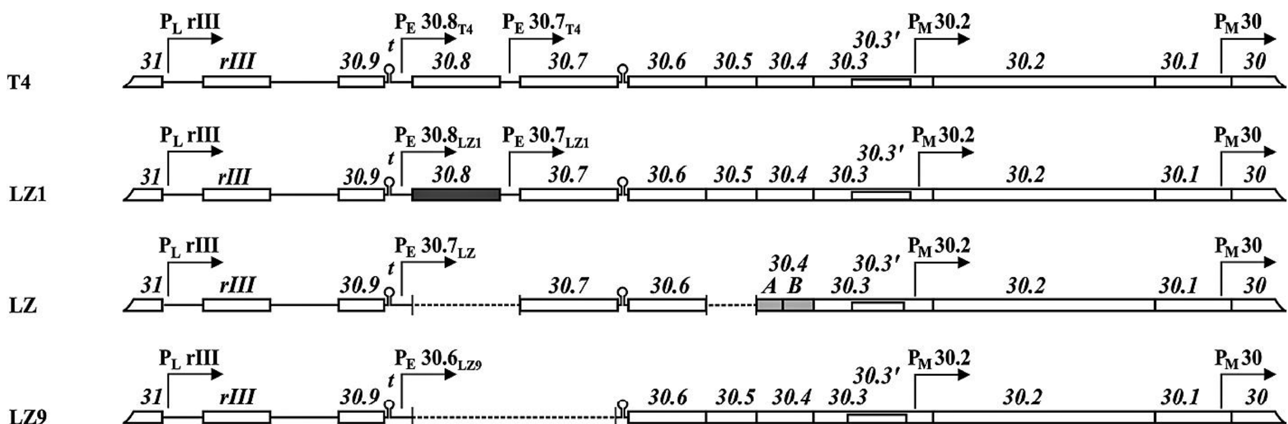

Fig. 1. Schematic outline of genomic region *30–31* of bacteriophages T4, LZ, LZ1 and LZ9. Shown are positions of genes, as well as the positions of late (P_L), early (P_E), middle (P_M) promoters, rho-independent terminators (*t*) and putative stem-loop structures (*s*). Grayed boxes represent genes with low homology to their T4 counterparts

A

```
T4     MKTINLNAAVKTKCFNG----KYDETMWFLMAVEGDIIEVETTEGMGTDFTFTIQVHNFFTGW 59
LZ1    MK-INLNSYIKGKDHDGYKAIETKELQWHLYKEQFEFVDCMTPEGPSDDFSWKIVITNFFTGD 62
       ** ****: :* * .:*      : .*  *.*    : ::::  *.** . **::.* : *****

T4     IYELNTVIVGKI-----EQNELGEWHYVTARQR----AERLIEKMKKVGKLDMQHWKVVK 110
LZ1    TYELKTLILGKIRCETYEDKETGYSEDVTWYQNGRITADLLIEKMKAKGESNLDNWIKVA 122
        ***:*:*:***     *::* *  . **  *.     *: ******  *: ::::*  *
```

B

```
30.4A_LZ   MFKKLIQKLLGTEMVEVTYRVTDVSTLTEDHLEPYYDN--------------------------------- 38
              ::*:*::* : ::***:*** .. *::* **   .
30.4_T4    --MNIINKIFGIQYIKVTYKVTDKNPYTDEHEEPQVESIILER-DPNWPVEFRLPCYGHWADVEIISIENV 68
                                                   :* : : * . .:* *** **********.
30.4B_LZ   ------------------------------------MTIKMLKYDGGLSIEDRLPSYGHWADVEIIKY--- 32
```

Fig. 2. Comparision of the primary structure of T4 gp30.8 with LZ1 gp30.8 (**A**) and T4 gp30.4 with gp30.4A and gp30.4B of phage LZ (**B**). Sequences were aligned using the ClustalW program. A dash indicates a space, which was inserted in the sequence to preserve the alignment. An asterisk (*) means that the residues in the column are identical; a double dot (:) indicates the conserved substitutions; a dot (.) indicates the semi-conserved substitutions

(115–137 nt of gene *30.5*); Pr.12, 5'–CGGATAACCC AAATATACGACG (complementary to 151–172 nt of gene *30.3*); Pr.13, 5'-CAAGATACTCCCCGGCGTGG (254–273 nt of gene *30.3*); Pr.14, 5'-GAAAGGAAAA GTTTTGAAGC (347–367 nt of gene *30.3*); Pr.15, 5'-AGGAAGGCCTGATTGCCAGC (complementary to 44–63 nt of gene *30.2*); Pr.16, 5'-ATGTCTGAA ACGGATGCGTC (complementary to 754–774 nt of gene *30.2*); Pr.17, 5'-GCTCAGCGAGTTTATACATGG (543–564 nt of gene *30.2*); Pr.18, 5'-GCTCGGAATT CTGGATTGCATC (complementary to 76–98 nt of gene *30.1*); Pr.19, 5'-CCCACGAGAATAAGTCAGACG (complementary to 59–80 nt of gene *30*).

LZ1-specific primers: Pr.1, 5'-GACGGCTATAAAGC TATCGAAAC (42–65 nt of gene *30.8*); Pr.2, 5'-GCAAGGATGAGTCCCTAAATCG (complementary to 221–243 nt of gene *30.7*).

RESULTS AND DISCUSSION

PCR analysis using the oligonucleotide primers based on phage T4 genome sequences allowed us to suppose that most of the genes located between DNA ligase gene *30* and a capsid assembly gene *31* were present in the genomes of phages tested. To compare the genomic region *30–31* of the bacteriophages tested with that of T4, we sequenced the PCR-generated LZ, LZ1 and LZ9 DNA fragments. Figure 1 shows the schematic outline based on the sequence data obtained. The outline shows that the most diverged part of the sequenced genomic region lies within the early gene cluster *30.9–30.3* in case of all three phages examined.

The high level of sequence conservation in this region was observed in bacteriophage LZ1. The deduced primary structures of gprIII, gp30.9 and gp30.5 of LZ1 shared a 100% sequence identity with their respective T4 homologues. Moreover, LZ1 gp30.7, gp30.6, gp30.5, gp30.4, gp30.3, gp30.3' [18], gp30.2 and gp30.1 seem to show a ~95% homology with their corresponding T4 counterparts. Therefore, indeed surprising was the fact that gp30.8 of LZ1 appeared to show only a 38% amino acid sequence identity with T4 gp30.8 (Fig. 2A). The deduced primary structure of LZ1 gp30.8 revealed that this protein is by 12 aa longer than that of T4. It is worth mentioning that the most distant orthologue of T4 gp30.8 was found in RB69, showing a 40% identity with its T4 counterpart [14]. The sequence alignment of LZ1 gp30.8 and RB69 gp30.8 revealed a ~55% amino acid identity (data not shown).

Notable sequence and organization differences in the region were observed in case of phage LZ (Fig. 1). This phage appeared not to have any analogues for T4 genes *30.8* and *30.5*. Meanwhile the counterpart of T4 gene *30.4* appears to be replaced by two individual partially overlapping open reading frames 30.4A and 30.4B encoding hypothetic proteins of 38 aa and 32 aa, respectively. The deduced primary structure of LZ gp30.4A shows an approximate 38% amino acid identity with first 36 N-terminal amino acids of T4 gp30.4, whereas gp30.4B of LZ, aligned with the C-terminal adequate part of T4 gp30.4, shows ~61% amino acid identity (Fig. 2B). The predicted primary structure of both gprIII and gp30.9 of LZ were identical with their respective T4 homologues, while the deduced amino acid sequences for gp30.7, gp30.6, gp30.3, gp30.3', gp30.2 and gp30.1 of this phage retained a >92% sequence identity compared with these of T4.

In case of bacteriophage LZ9, we did not detect any counterparts of T4 genes *30.8* and *30.7*. Meanwhile, the genes that were present in the genomic region *30–31* of LZ9 showed a high percentage of similarity with their respective T4 homologues. When translated, LZ9 gprIII and gp30.9 were 100% identical to their T4 counterparts. The predicted amino acid sequences of gp30.6, gp30.5, gp30.4, gp30.3, gp30.3', gp30.2 and gp30.1 showed a >91% amino acid identity with those of T4.

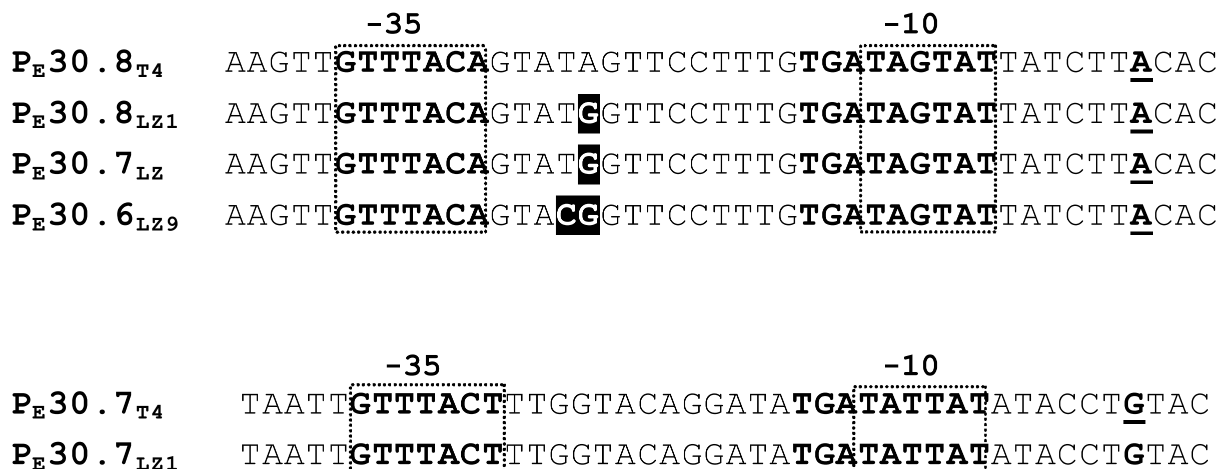

Fig. 3. Comparison of the nucleotide sequences of T4 early promoters P_E30.8 and P_E30.7 with these found in phages LZ, LZ1 and LZ9. Different nucleotides are given in black background. The –35 and –10 consensus elements of early promoters are boxed. The sequences of –35 and the extended –10 elements are shown in bold. The initiating nucleotide is shown in bold and underlined

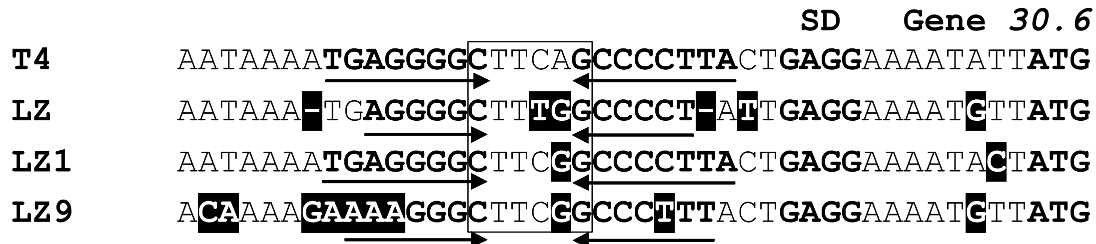

Fig. 4. Comparison of non-coding region upstream gene *30.6* of phages T4, LZ, LZ1 and LZ9. Different nucleotides are given in black background. A dash indicates a space, which was inserted to preserve the alignment. The initiation codon of gene *30.6*, the GAGG motifs of the SD sequences, and the inverted repeats of palindromic sequences are given in bold. Convergent arrows underline the inverted repeats of palindromic sequences; the loop motifs are boxed

The alignment of nucleotide sequences of the intergenic region *rIII-30.9* of bacteriophages LZ, LZ9 and T4 revealed the presence of insertion in both phages tested. The intergenic region between genes rIII and *30.9* of bacteriophage LZ appeared to contain the insertion of 44 nt, whereas phage LZ9 contained a 153 nt insertion forming a new 141 nt hypothetical ORF.

As already mentioned, bacteriophage T4 employs seven regulatory elements in the genomic region between genes *30* and *31*. Comparative sequence alignment revealed some changes in the regulatory sequences of all three phages analyzed. Phages LZ and LZ9 seem to harbor only one of the two early promoters found in the region *30–31* of T4. In both cases, the retained P_E was very similar to $P_E30.8$ of T4 (Fig. 3), showing a difference of only one ($P_E30.7_{LZ}$) or two ($P_E30.6_{LZ9}$) bases in the spacing region between –10 and –35 elements. In case of bacteriophage LZ9, $P_E30.6_{LZ9}$ appears to be situated between the sequences encoding two stable RNA hairpins. It must be noted that LZ9 shares the same structural features of the intergenic region between genes *30.9* and *30.6* found in bacteriophage LZ10 [14]. Since the $P_E30.6_{LZ10}$ showed no functional disturbances [14], we may suggest that $P_E30.6_{LZ9}$ is active as well. The –10 and –35 elements of both $P_E30.7_{LZ1}$ and $P_E30.8_{LZ1}$ appeared to be identical with their respective T4 analogues, with a single base substitution found in the spacing region of $P_E30.8_{LZ1}$ (Fig. 3). Furthermore, despite several nucleotide differences and deletions observed in the intergenic region between genes *30.7* and *30.6*, palindromic sequences giving rise to a stable RNA stem-loop structure are present in all phages analyzed (Fig. 4). Moreover, in case of phages LZ1 and LZ9 the secondary RNA structure has the predominant tetraloop sequence UUCG which confers an exceptional stability to the RNA hairpins [19]. Meanwhile the sequences for the rho-independent terminator, both middle promoters and late promoter, were found to be 100% identical to these present in the genomic region *30–31* of T4.

From these considerations we may conclude that the genomic region between genes *30* and *31* of all three bacteriophages tested, despite various divergences both in size and nucleotide sequence, employ almost the same regulatory pattern of gene regulation as does bacteriophage T4. Furthermore, the presence of a ~100% identity of genes *rIII* and *30.9* among phages LZ, LZ1, LZ9 and T4 implies the importance of the corresponding gene proteins under certain growth conditions.

ACKNOWLEDGEMENTS

We thank Dr. Lindsay W. Black for the *E. coli* strain and Dr. Karin Carlson for the phages. This work was supported in part by the Lithuanian State Science and Studies Foundation.

Received 17 February 2006
Accepted 11 August 2006

References

1. Desplats C, Krisch HM. Res Microbiol 2003; 154: 259–67.
2. Ackermann AW, Krisch HM. Arch Virol 1997; 142: 2329–45.
3. Desplats C, Dez C, Tétart F, Eleaume H, Krisch HM. J Bacteriol. 2002; 184: 2789–804.
4. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczypinski B, Fraser CM. J Bacteriol 2003; 185: 5220-33.
5. Kawabata T, Arisaka F, Nishikawa K. Gene 2000; 259: 223–33.
6. Nivinskas R, Zajančkauskaitė A, Raudonikienė A, Vitėnienė I. DNA Sequence. J. DNA Sequencing Mapp 1992; 2: 405–9.
7. Raudonikenė A, Nivinskas R. Gene 1992; 114: 85–90.
8. Zajančkauskaitė A, Raudonikienė A, Nivinskas R. Gene 1994; 147: 71–76.
9. Zajančkauskaitė A, Malys N, Nivinskas R. Gene 1997; 194: 157–62.
10. Nivinskas R, Black LW. Mol Biol 1988; 22: 1507–16.
11. Driukas A, Nivinskas R. Biologija 1998; 1: 34–36.
12. Gram H, Liebig HD, Hack A, Niggemann E, Ruger W. Mol Gen Genet 1984; 194: 232–40.
13. Liebig HD, Rüger W. Bacteriophage T4 early promoter regions. Consensus sequences of promoters and ribosome-binding sites. J Mol Biol 1989; 208**:** 517–36.
14. Truncaitė L, Zajančkauskaitė A, Arlauskas A, Nivinskas R. Virology 2006; 344: 378–90.

15. Truncaitė L, Zajančkauskaitė A, Nivinskas R. J Mol Biol 2002; 317: 179–90.
16. Kutter E, Gachechiladze K, Poglazov A, Marusich E, Shneider M, Aronsson P, Napuli A, Porter D, Mesyanzhinov V. Virus Genes 1996; 11(2/3): 285–97.
17. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Science 1988; 239: 487–91.
18. Kolesinskiene G, Zajančkauskaite A, Arlauskas A, Klausa V, Nivinskas R. Biologija 2003; 2: 15–18.
19. Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, Stormo G, d'Aubenton-Carafa Y, Uhlenbeck OC, Tinoco Jr I, Brody EN, Gold L. Proc Natl Acad Sci USA 1988; 85: 1364–68.

Laura Kalinienė, Lidija Truncaitė, Aivaras Arlauskas, Giedrė Stoškienė, Aurelija Zajančkauskaitė, Rimas Nivinskas

LZ, LZ1 IR LZ9 BAKTERIOFAGŲ GENOMO SRITIES TARP *30* IR *31* GENŲ STRUKTŪRINĖS ORGANIZACIJOS TYRIMAS

S a n t r a u k a

Šiame darbe buvo nustatyta LZ, LZ1 ir LZ9 bakteriofagų genomo srities tarp *30* ir *31* genų nukleotidų seka ir išaiškinta šios srities struktūrinė organizacija. Didžiausi struktūriniai pokyčiai visų tirtų fagų minėtame regione buvo aptikti tarp *30.9* ir *30.3* genų. LZ fage neaptikta T4 atitikmenų *30.8* ir *30.5* genams, o LZ9 fage – *30.8* bei *30.7* genams. Nustatyta, kad LZ1 fago *30.8* genas, lyginant su T4 fago *30.8* genu, yra itin pakitęs. LZ fago tirtoje genomo srityje vietoj vieno *30.4* geno, aptinkami du – *30.4A* bei *30.4B* genai. Vieno iš dviejų ankstyvųjų promotorių, esančių T4 genomo srityje *30–31*, neaptikta LZ ir LZ9 faguose. LZ9 bakteriofago genomo srityje tarp *rIII* ir *30.9* genų aptikta 153 bazių porų insercija, kurioje rastas naujas hipotetinis ASR.