Molecular Biology, Vol. 37, No. 5, 2003, pp. 716–722. Translated from Molekulyarnaya Biologiya, Vol. 37, No. 5, 2003, pp. 843–849.

BIOINFORMATICS

UDC 577.1

Computer-Assisted Analysis of Regulation of the Glycerol-3-Phosphate Metabolism in Genomes of Proteobacteria

L. V. Danilova[1], M. S. Gelfand[2], V. A. Lyubetsky[1], and O. N. Laikova[2]

[1] *Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 101447 Russia; E-mail: dlv2k@mail.ru*

[2] *State Research Center GosNIIGenetika, Moscow, 113545 Russia*

Received April 04, 2003

Abstract—Comparative computer-assisted analysis was used to study putative GlpR regulons responsible for metabolism of glycerol and glycerol-3-phosphate in genomes of α-, β-, and γ-proteobacteria. New palindromic GlpR-binding signals were identified in γ-proteobacteria, consensus sequences being TGTTCGATAACGAACA for Enterobacteriaceae, wTTTTCGTATACGAAAAw for Pseudomonadaceae, and AATGCTCGATCGAGCATT for Vibrionaceae. The signals in α- and β-proteobacteria were also identified: they contained 3–4 direct TTTCGTT repeats separated by 3–4 nucleotide pairs.

Key words: GlpR, tandem repeats, computer-assisted analysis, operon structure, α-proteobacteria, β-proteobacteria, γ-proteobacteria

INTRODUCTION

The regulator GlpR, a member of the regulator family DeoR, controls the expression of the genes involved in metabolism of glycerol and glycerol-3-phosphate (G3P). The regulon GlpR is well studied in *Escherichia coli* [1–3], some data are available for *Pseudomonas aeruginosa* [4].

Glycerol is transferred to the cytoplasm from the outside by simplified diffusion provided by the *glpF* gene product, while G3P is actively transported by the *glpT* gene product. Intracellular glycerol is phosphorylated by glycerol kinase (*glpK*) into G3P, which can then be turned into dihydroxyacetone phosphate by one of the two G3P dehydrogenases available in *E. coli*: aerobic (*glpD*) or anaerobic (*glpA*). Besides these genes, the GlpR regulon in *E.coli* includes *glpQ* coding for periplasmic glycerophosphodiesterase, which hydrolyzes glycerol phosphodiesters and liberates G3P; *glpB* and *glpC* coding for the additional structural components of the anaerobic dehydrogenase; and *glpE*, *glpG*, and *glpX*, the function of which remains obscure. The above-mentioned genes are integrated into three loci on the chromosome of *E. coli*: *glpTQ*/*glpABC*, *glpEGR*/*glpD* and *glpFKX* (slash separates operons of opposite orientation). The gene GlpR has maximal affinity toward the regulatory region *glpD*. The G3P appears to be a true inducer of the regulon GlpR.

In this work we studied the structure of the GlpR regulons and analyzed the GlpR-binding signals in genomes of proteobacteria.

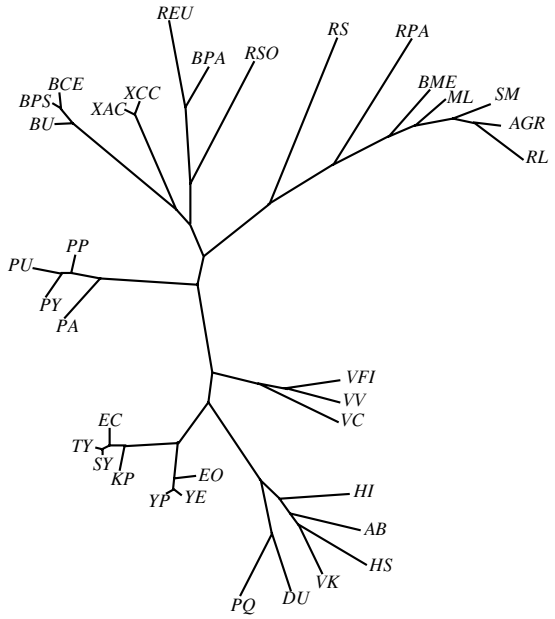

Fig. 1. Phylogenetic tree for homologs of the repressor GlpR in α-, β-, γ-proteobacteria. *EC, E. coli; TY, S. typhi; SY, S. typhimurium; KP, K. pneumoniae; YP, Y. pestis; YE, Y. enterocolitica; EO, E. carotovora; VC, V. cholerae; VV, V. vulnificus; VFI, V. fischeri; HI, H. influenzae; DU, H. ducrey; HS, Haemophilus somnus; VK, P. multocida; PQ, P. haemolytica; AB, A. actinomycetemconitans; PA, P. aeruginosa; PP, P. putida; PU, P. fluorescens; PY, P. syringae; BU, B. fungorum; BPS, B. pseudomallei; BCE, B. cepacia; XAC, Xanthomonas axonopodis; XCC, Xanhtomonas campestris; BPA, B. parapertussis; REU, R. eutropha; RSO, R. solanacearum; RL, R. leguminosarum; AGR, A. tumefaciens; SM, S. meliloti; ML, M. loti; BME, B. melitensis; RPA, R. palustris; RS, R. sphaeroides.*

0026-8933/03/3705-0716$25.00 © 2003 MAIK "Nauka/Interperiodica"

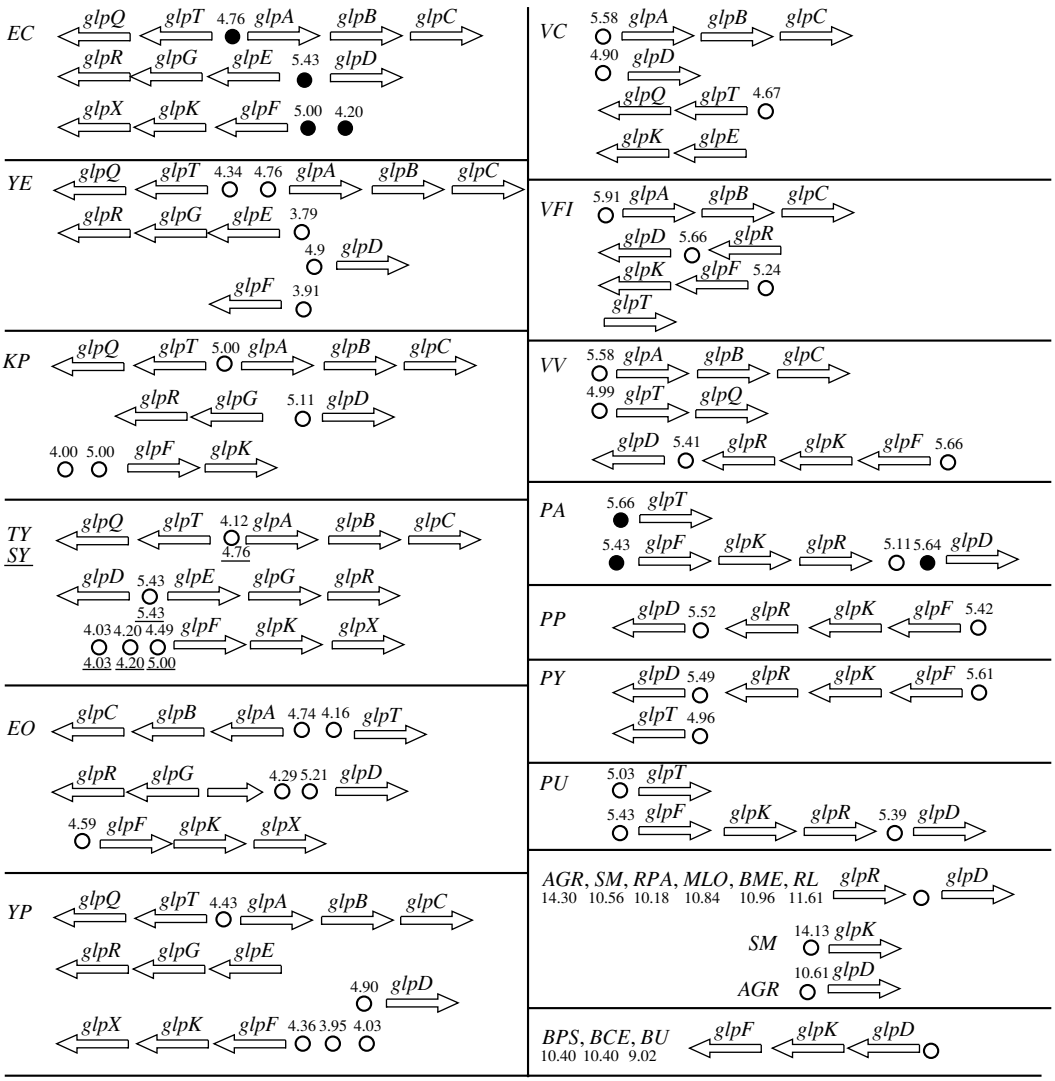

Fig. 2. Operon structure of the GlpR regulons in α-, β-, γ-proteobacteria. Shaded circles show known sites, end empty circles show predicted sites with their weights. Abbreviations, see Fig. 1.

EXPERIMENTAL

Genomes: *Escherichia coli* [5], *Salmonella typhi* [6], *S. typhimurium* [7], *Klebsiella pneumoniae* [8], *Erwinia carotovora*, *Yersinia pestis* [9], *Y. enterocolitica* [9], *Vibrio cholerae* [10], *V. vulnificus* [9] *V. fischeri*, *Pasteurella multocida* [11], *P. haemolytica*, *Haemophilus influenzae* [12], *H. ducrey*, *Actinobacillus actinomycetemcomitans* [16], *Pseudomonas aeruginosa* [13], *P. fluorescens* [14], *P. putida* [9], *P. syringae* [15], *Burkholderia fungorum*, *B. pseudomallei*, *B. cepacia*, *Bordetella parapertussis*, *Ralstonia eutropha*, *R. solanacearum* [9], *Mesorhizobium loti* [9], *Sinorhizobium meliloti* [9], *Rhizobium leguminosarum*, *Agrobacterium tumefaciens* [9], *Brucella melitensis* [9], *Rhodopseudomonas palustris* [9]. Close homologs of GlpR were found in many genomes, and a tree was generated (Fig. 1); all these genomes were studied. The ClustalW program was used to align the protein sequences [17], and the tree was generated with program PROML of the PHYLIP package [18].

Definitions of the genes in this work correspond to the names of their orthologs in *E. coli*.

Programs GenomeExplorer [19], SignalX [19], and IRSA [20] were used to search for the sites and to generate the learning sample and the positional weight matrix.

A matrix of positional weights was defined as:

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

where $N(b, k)$ is the number of occurrences of nucleotide b in position k. The weight of a putative signal is determined as the sum of positional weights of the

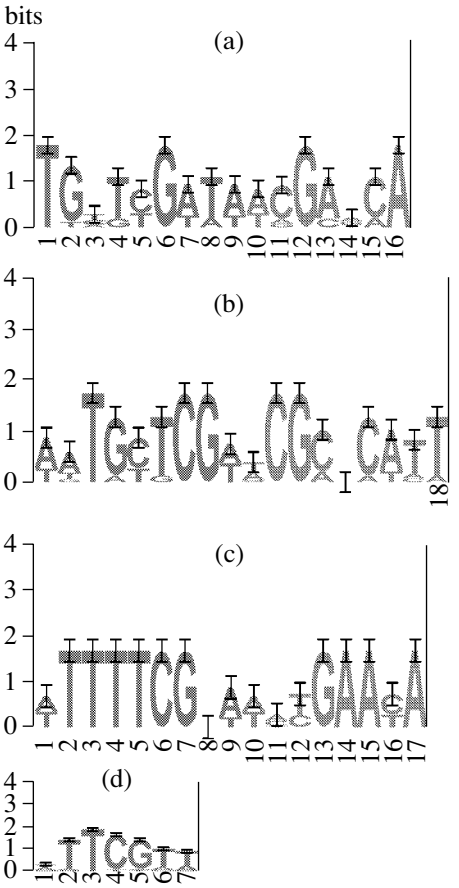

Fig. 3. Graph representation of the positional weight matrices obtained with the procedure described in [23] for different bacterial species. (a) *E. coli, E. carotovora, Y. enterocolitica, K. pneumoniae*; (b) *V. cholerae, V. vulnificus, V. fischeri*; (c) *P. aeruginosa, P. syringae, P. fluorescens, P. putida*; (d) *B. melitensis, M. loti, A. tumefaciens, S. meliloti, R. palustris*.

constituent nucleotides. The logarithm base is selected to provide normal distribution of the random nucleotide weights with zero average and variance equal 1.

All matrices of positional weights were applied to genomes aiming to search for new sites in the regions from –400 to +50 bp around the gene start point.

In cases when the operon structure of the given DNA fragment was unknown, the genes were assigned to one putative operon if they had the same reading direction and the distance between them was not larger than 100 bp. The start of this putative operon was defined as the gene with the upstream GlpR site (See Fig. 2).

RESULTS AND DISCUSSION

We found orthologs of the genes forming the GlpR regulon in all studied genomes. The regions upstream of these genes were analyzed as a learning sample using signal detection programs. The main taxonomic groups corresponding to the branches of the protein GlpR tree were analyzed separately.

γ-Proteobacteria, group Vibrionaceae. In this group of the three genomes: V. cholerae, V. vulnificus, and V. fischeri we succeeded to identify a 18-bp-long palindromic signal with consensus AATGCTCGATCGAGCATT (Fig. 3b). The identified sites and the matrix of positional weights are shown in Tables 1b and 2b, respectively. No putative sites were found by scanning of the genomes with this matrix.

γ-Proteobacteria, group Pseudomonadaceae. Then we studied four genomes of the group Pseudomonadaceae. A 18-bp palindromic signal wTTTTCGTATACGAAAAw was found in regulatory regions (Fig. 3c); this signal included the sites predicted earlier for *P. aeruginosa* [4] (Table 1c); The positional weight matrix was generated (Table 2c). This matrix with the limit of 4.9 detected new putative GlpR-binding sites shown in Table 1c.

α, β-Proteobacteria. In genomes of α-proteobacteria we found 3–4 tandem repeats of the motif TTTCGTT (Fig. 3d), following each other after 3–4 nucleotides (Table 1d); the matrix of positional weights was generated (Table 2d). Further analysis of *A. tumefaciens* found two orthologs of the gene *glpD*, that have tandem repeats in their regulatory regions. Similar repeats were found in bacteria of the genus *Burkholderia* from the group of β-proteobacteria (see Table 1d).

γ-Proteobacteria, group Enterobacteriaceae. First we analyzed regulatory regions of the four genomes: *E. coli*, *E. carotovora*, *Y. enterocolitica*, and *K. pneumoniae*. They were used to generate the learning sample which included already known sites from *E. coli* (Table 1a) with consensus TGTTCGATAACGAACA (Fig. 3a). This sample was used to generate the positional weight matrix (Table 2a) to search for palindromic signals longer than 16 bp. This matrix was used to scan the genomes of *Y. pestis*, *S. typhimurium*, and *S. typhi* with the limit of 4.1. The found sites are shown in Table 1a.

Lowering the limit for genomes of the group Enterobacteriaceae results in considerable overprediction; however, it allows detection of rather weak sites upstream of the genes which form the G3P regulon (Table 1d). At the same time, even with a low limit putative sites can be detected not in all footprint regions shown in [1–3].

Since no data are available on the 3D structure of the regulators from the family DeoR, we have no reason to suggest that these regulators always form dimers which bind with the palindromic sites. Moreover, taking into account a tandem repeat signal identified by us for some other groups of bacteria, one may suggest that the site structure in Enterobacteriaceae is

Table 1. Sites upstream of the genes of the G3P regulon in the indicated bacterial species (the learning sample shown in bold)

Genome	Gene	Weight of site	Site
(a) Family Enterobacteriaceae			
E. coli	*glpD*	5.41	**T G T T C G A T A a C G A A C A**
″	*glpF*	4.99	**T G c T C G t T A a C G A t a A**
″	*glpT*	4.76	**T G T T t G A T t T C G c g C A**
E. carotovora	*glpD*	5.20	**T G c T C G A a A a C G A A C A**
″	*glpT*	4.72	**T G T T t G A T A a a G A g C A**
″	*glpF*	4.59	**T t c T C G t T t T C G c t C A**
K. pneumoniae	*glpD*	5.10	**T G a g C G A T A T C G A g C A**
″	*glpT*	5.00	**T G T T t G A T t T C G A g C A**
″	*glpF*	4.99	**T G c T C G t T A a C G A t a A**
Y. enterocolitica	*glpD*	4.89	**T G a g C G A a A a C G A A C A**
″	*glpT*	4.74	**c G c T C G t T A T g G A A C A**
E. coli	*glpF*	4.20	g G c g C G A T A a C G c t C A
E. carotovora	*glpD*	4.29	T G T T t G t T t T C G A t t A
″	*glpA*	4.16	T G T T C t A T t a C G A A C g
S. typhi	*glpD*	5.43	T G T T C G A T A a C G A A C A
″	*glpF*	4.49	T G c T C G t T A g C G A t a A
″	*glpF*	4.20	g G c g C G A T A a C G c t C A
″	*glpT*	4.12	T G T T t G A T t T C G c g C g
S. typhimurium	*glpD*	5.43	T G T T C G A T A a C G A A C A
″	*glpF*	5.00	T G c T C G t T A a C G A t a A
″	*glpT*	4.76	T G T T t G A T t T C G c g C A
″	*glpF*	4.20	g G c g C G A T A a C G c t C A
Y. enterocolitica	*glpA*	4.34	T G T T C c A T A a C G A g C g
Y. pestis	*glpD*	4.90	T G T T C G t T t T C G c t C A
″	*glpA*	4.43	T G T T t c t T A T C a A t C A
″	*glpF*	4.36	c G c T C G t T A a C G A t a A
(b) Family Vibrionaceae			
V. cholerae	*glpA*	5.57	**A A T G C T C G t T C G c G C t T T**
″	*glpD*	4.92	**A A T a t T C G A g C G c t C A T T**
″	*glpT*	4.56	**A t T G C T C G t T C G c c a t T T**
V. fischeri	*glpA*	5.91	**A A T G C g C G A a C G A G C A T T**
″	*glpD*	5.66	**A A T G t T C G t T C G c t C A T T**
″	*glpF*	5.24	**t g T G C T C G A a C G c t C A T T**
V. vulnificus	*glpF*	5.69	**t A T G C T C G A a C G c G C A T T**
″	*glpA*	5.66	**A A T G t T C G A a C G c t C A T T**
″	*glpD*	5.36	**A A T G C T C G t T C G A a C A a a**
″	*glpT*	5.02	**t t T G C T C G t T C G c a C A c T**
(c) Family Pseudomonadaceae			
P. aeruginosa	*glpD*	5.64	**A T T T T C G a A T t C G A A c A A**
″	*glpF*	5.43	**T T T T T C G a A a c t G A A c A A**
P. fluorescens	*glpF*	5.43	**T T T T T C G a A T c t G A A t A A**
″	*glpD*	5.39	**A T T T T C G c A a A t G A A c A T**
P. putida	*glpD*	5.52	**A T T T T C G c A a A C G A A c A T**

Table 1. (Contd.)

Genome	Gene	Weight of site	Site
″	*glpF*	5.42	**TTTTTCGTt Tc t GAAt AA**
P. syringae	*glpF*	5.61	**TTTTTCGTt TACGAAt AT**
″	*glpD*	5.49	**ATTTTCGg Aa A t GAAc AT**
P. aeruginosa	*glpT*	5.66	T T T T T C a T t TACGAAAAA
″	*glpD*	5.11	A T g T T C G T t T c a GAAAAA
P. fluorescens	*glpT*	5.03	A T T T T C G g t a ACGAAAc T
P. syringae	*glpT*	4.96	T T T T T C t g t a A t GAAAAT
			(d) α-, β-Proteobacteria
A. tumefaciens	*glpD*	14.30	**g TTCGTTt a t TTTC t TT t gac a TTCGTTt tgt TTTCGc T**
″	*glpD*	10.61	**TTTCGTTt g a c a TTCGTTt t g t CTTCGAA**
B. melitensis	*glpD*	10.96	**TTTCGTTt g a t TTTCa TTt gc TTTCGTa**
M. loti	*glpD*	10.84	**TTTCGTTt g a c a TTCGTTat g a g TTCGa a**
R. leguminosarum	*glpD*	11.61	**a TTCGTTt g a c a TTCGTa t t c c TTTCGTT**
R. palustris	*glpD*	10.18	**TTTCGTTt t g g TT t GTg c t t t a TTCGTT**
S. meliloti	*glpK*	14.13	**TTTCGTTt g a c a TTCGTTt t t c Ta TCTa t tgaa g TCGTT**
″	*glpD*	10.56	**a TTCGTTt g a c a TTCGa a at a t TTTCGc T**
B. pseudomallei	*glpD*	10.40	T T T C G a T t a t g T T C G T T a aa T T T C G a a
B. cepacia	*glpD*	10.40	T T T C G a T t c c g T T C G T T a aa T T T C G a a
B. fungorum	*glpD*	9.02	T T T C G a a t a t g T T C a T T a aag T T C G a a
			(e) Putative sites in Enterobacteriaceae (strong overprediction)
K. pneumoniae	*glpF*	4.00	G G c g C G A a A a C G c t C A
S. typhi	*glpF*	4.03	T t c a C G t a A a C G c g C A
S. typhimurium	*glpF*	4.03	T t c a C G t a A a C G c g C A
Y. enterocolitica	*glpF*	3.91	A G c T t G A T A a C a A t a A
″	*glpE*	3.79	T t a g C a A T A T g G A A C A
Y. pestis	*glpF*	4.03	T a c g C G A a A a C G c t C A
″	*glpF*	3.95	T t c T C G t T t T C G c t C g

also a tandem repeat. Alignment of the regions upstream of *glpD* in bacteria of this group: *E. coli*, *S. typhimurium*, *K. pneumoniae*, *Y. enterocolitica* shows both palindromic symmetry and repeats (Table 3). However, we have not succeeded in generating the matrix of positional weights for these regulatory regions, assuming either palindromic symmetry or tandem repeat symmetry.

Figure 2 shows examples of operons from some of the studied genomes regulated by repressor GlpR and having a respective site. Shadowed circles are known sites, and empty circles are predicted sites with their weights.

In conclusion, we have identified putative binding signals for GlpR in Vibrionaceae and Pseudomonadaceae (palindrome), and also in α-, β-proteobacteria (tandem repeat). The situation of Enterobacteriaceae remains unclear. Alignment of the regulatory regions and their analysis with the signal-detecting programs allow one to suggest the existence of either a palindromic signal or a signal in the form of a phased direct repeat. We cannot exclude a possibility that the GlpR monomers in these bacteria can bind with regulatory sites in various orientations, forming cooperative complexes (as, e.g., regulators NarL [21] and FUR [22]). This situation may be clarified by the new experimental data, sequencing of the new genomes from this group, or by analysis of other regulons of the family DeoR.

Table 2. Matrix of positional weights for the GlpR signal in different bacteria

(a) *E. coli*, *E. carotovora*, *Y. enterocolitica*, *K. pneumoniae*				(b) *V. cholerae*, *V. vulnificus*, *V. fischeri*				(c) *P. aeruginosa*, *P. syringae*, *P. fluorescens*, *P. putida*			
a	c	g	t	a	c	g	t	a	c	g	t
−0.16	−0.16	−0.16	0.48	0.35	−0.24	−0.24	0.13	0.25	−0.25	−0.25	0.25
−0.23	−0.23	0.38	0.09	0.29	−0.30	−0.02	0.03	−0.15	−0.15	−0.15	0.46
0.06	0.13	−0.34	0.15	0.05	−0.22	−0.22	0.39	0.03	−0.31	0.11	0.18
−0.24	−0.24	0.12	0.36	−0.06	−0.25	0.37	−0.06	−0.15	−0.15	−0.15	0.46
−0.24	0.36	−0.24	0.12	0.00	0.15	−0.18	0.04	−0.15	−0.15	−0.15	0.46
−0.25	−0.06	0.37	−0.06	−0.26	−0.26	0.24	0.27	−0.15	0.46	−0.15	−0.15
0.28	−0.26	−0.26	0.25	−0.16	0.47	−0.16	−0.16	0.18	−0.24	0.31	−0.24
0.15	−0.25	−0.25	0.35	−0.16	−0.16	0.47	−0.16	0.01	−0.09	0.01	0.07
0.35	−0.25	−0.25	0.15	0.22	−0.11	−0.30	0.20	0.29	−0.24	−0.24	0.20
0.25	−0.26	−0.26	0.28	0.20	−0.30	−0.11	0.22	0.20	−0.24	−0.24	0.29
−0.06	0.37	−0.06	−0.25	−0.16	0.47	−0.16	−0.16	0.07	0.01	−0.09	0.01
0.12	−0.24	0.36	−0.24	−0.16	−0.16	0.47	−0.16	−0.24	0.31	−0.24	0.18
0.36	0.12	−0.24	−0.24	0.27	0.24	−0.26	−0.26	−0.15	−0.15	0.46	−0.15
0.15	−0.34	0.13	0.06	0.04	−0.18	0.15	0.00	0.46	−0.15	−0.15	−0.15
0.09	0.38	−0.23	−0.23	−0.06	0.37	−0.25	−0.06	0.46	−0.15	−0.15	−0.15
0.48	−0.16	−0.16	−0.16	0.39	−0.22	−0.22	0.05	0.18	0.11	−0.31	0.03
				0.03	−0.02	−0.30	0.29	0.46	−0.15	−0.15	−0.15
				0.13	−0.24	−0.24	0.35	0.25	−0.25	−0.25	0.25
(d) *B. melitensis*, *M. loti*, *A. tumefaciens*, *S. meliloti*, *R. palustris*											
0.19	−0.33	−0.10	0.24								
−0.11	−0.40	−0.11	0.63								
−0.26	−0.26	−0.26	0.79								
−0.33	0.71	−0.33	−0.04								
−0.11	−0.40	0.63	−0.11								
0.03	−0.06	−0.49	0.51								
0.19	−0.49	−0.19	0.49								

Table 3. Alignment of the regions upstream of the *glpD* gene in *E. coli*, *S. typhimurium*, *K. pneumoniae*, *Y. enterocolitica* (palindromic regions in bold, repeats underlined; weight calculated for palindromic site)

Genome	Gene	Weight of site	Site
E. coli	*glpD*	10.0	aat a**tgttcgat**<u>**aacgaaca**</u>ttt a<u>**tgagctt**</u>**t**<u>**aacgaaag**</u>tgaat
S. typhimurium	*glpD*	10.0	att a**tgttcgat**<u>**aacgaaca**</u>ttttt<u>**tgaactt**</u>**t**<u>**aacgaaag**</u>tgcaa
K. pneumoniae	*glpD*	9.8	ata g**tgagcgat**<u>**atcgagca**</u>ttt a<u>**tgagctt**</u>**a**<u>**aacgaaag**</u>tgtga
Y. enterocolitica	*glpD*	8.9	atcg**tgagcgaa**<u>**aacgaaca**</u>tta a<u>**agagctgtt**</u>**t**<u>**cgaaca**</u>ttt gg

ACKNOWLEDGMENTS

The authors are grateful to V.Yu. Makeev, A.A. Mironov, and D.A. Rodionov for helpful discussion of the results, and to K.Yu. Gorbunov for his assistance in developing the algorithm.

This work was supported by the Howard Hughes Medical Institute (grant 55000309).

REFERENCES

1. Weissenborn D.L., Wittekindt N., Larson T.J. 1992. Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J. Biol. Chem.* **267**, 6122–6131.
2. Larson T.J., Cantwell J.S., van Loo-Bhattacharya A.T. 1992. Interaction at a distance between multiple operators controls the adjacent, divergently transcribed

glpTQ-glpABC operons of *Escherichia coli* K-12. *J. Biol. Chem*. **267**, 6114–6121.

3. Yang B., Larson T.J. 1996. Action at a distance for negative control of transcription of the *glpD* gene encoding sn-glycerol 3-Phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol*. **178**, 7090–7098.
4. Schweizer H.P., Po C. 1996. Regulation of glycerol metabolism in *Pseudomonas aeruginosa*: characterization of glpR repressor gene. *J. Bacteriol*. **178**, 5215–5221.
5. Blattner F.R., Plunkett G. 3rd, Bloch C.A. *et al.* 1997. The complete genome sequence of *Escherichia coli* K-12. *Science*. **277**, 1453–1474.
6. Parkhill J., Dougan G., James K.D. *et al*. 2000. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar *typhi* CT18. *Nature*. **413**, 848–852.
7. McClelland M., Sanderson K.E., Spieth J. *et al.* 2001. Complete genome sequence of *Salmonella enterica* serovar *typhimurium* LT2. *Nature*. **413**, 852–856.
8. http://genome.wustl.edu
9. http://www.ncbi.nlm.nih.gov/Genbank/index.html
10. Parkhill J., Wren B.W., Thomson N.R. *et al.* 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*. **413**, 523–527.
11. 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.
12. May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida* Pm70. *Proc. Natl. Acad. Sci. USA*. **98**, 3460–3465.
13. 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.
14. http://spider.jgi-psf.org/JGI_microbial/html/
15. http://www.tigr.org
16. http://www.genome.ou.edu/act.html
17. Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**, 4673–4680.
18. Lim A., Zhang L. 1999. WebPHYLIP: a web interface to PHYLIP. *Bioinformatics*. **15**, 1068–1069.
19. Mironov A.A., Vinokurova N.P., Gelfand M.S. 2000. Software for analysis of bacterial genomes. *Mol. Biol.* **34**, 253–262.
20. Danilova L.V., Gorbunov K.Yu., Gelfand M.S., Lyubetsky V.A. 2001. An algorithm to detect regulatory signals in DNA sequences. *Mol Biol*. **35**, 987–995.
21. Gennis R.B., Stewart V. 1996. *Escherichia coli* and *Salmonella*. Cellular and Molecular biology. Washington DC: ASM Press. 2822 p.
22. Panina E.M., Mironov A.A., Gelfand M.S. 2001. Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res*. **29**, 5195–5206.
23. http://www.bio.cam.ac.uk/seqlogo/logo.cgi