Molecular Biology, Vol. 39, No. 5, 2005, pp. 727–740. Translated from Molekulyarnaya Biologiya, Vol. 39, No. 5, 2005, pp. 832–846.

GENOMICS. TRANSCRIPTOMICS. PROTEOMICS

UDC 577.1

Regulation of Nitrate and Nitrite Respiration in γ-Proteobacteria: A Comparative Genomics Study

D. A. Ravcheev[1,2], A. B. Rakhmaninova[1,2], A. A. Mironov[1,2,3], and M. S. Gelfand[1,2,3]

[1] *Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119992 Russia*
[2] *Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia*
e-mail: ravcheyev@iitp.ru
[3] *GosNIIGenetika State Research Center, Moscow, 117545 Russia*

Received March 9, 2005

Abstract—Nitrate and nitrite are the most preferable electron acceptors in the absence of molecular oxygen. In the γ-proteobacterium *Escherichia coli*, nitrate and nitrite respiration is regulated by two homologous transcription factors, NarL and NarP. Although this regulatory system was a subject of intensive research for more than 20 years, many key issues, including the structure of the NarL-binding site, are still unclear. Comparative genomics analysis showed that only NarP is responsible for regulation in most γ-proteobacteria. The NarP regulon was studied in ten genomes. Although its structure considerably differs among some genomes, the mechanism regulating the nitrate and nitrite reduction genes is highly conserved. A correlation was observed between the evolutionary changes in the nitrate and nitrite respiration system and the relevant regulatory system. Potential NarP-binding sites were found upstream of the gene for the global regulator FNR and the *syd*AB, *mdh*, and *suc*AB aerobic metabolism genes. It was assumed on the basis of this evidence that the role of NarP in regulating respiration changed during evolution. In total, 35 new operons were assigned to the generalized NarP regulon. Autoregulation of the *nar*QP operon was suggested for bacteria of the family Vibrionaceae.

Key words: comparative genomics, γ-proteobacteria, nitrate and nitrite respiration, NarP, NarL, NarX, NarQ

INTRODUCTION

Many prokaryotes, such as *Escherichia coli,* utilize various substrates for respiration, which ensures their survival under conditions changing within a broad range. Nitrate and nitrite respiration is a preferential source of energy in the absence of oxygen [1]. Respiration of this type is regulated in *E. coli* by a double two-component system, which includes the homologous sensory proteins NarQ and NarX and the homologous transcription factors NarL and NarP (Nar, nitrate reductase regulator). It is thought that such a regulatory system is necessary for modulating the respiratory system in the presence of two alternative electron acceptors, nitrate and nitrite [2]. Both NarL and NarP act to suppress or activate transcription, depending on the arrangement of their binding sites relative to the promoter of the target operon [1].

A partial division of functions between the homologous proteins is characteristic of this regulatory system (Fig. 1). For instance, NarX interacts exclusively with NarL and responds differently to nitrate and nitrite, while NarQ is capable of interacting with either regulator and similarly responds to both substrates [3]. On the other hand, NarP does not interact with the NarP-binding sites, while NarL interacts with both NarL-binding and its own sites [4] (Fig. 1). Thus, an integrated NarL–NarP regulon functions in *E. coli.* We divided its genes into the following groups.

Group 1 includes genes providing for reduction of nitrate and nitrite: *nar*GHJI, *nar*K [5], *nap*FDABCDEF–*ccm*ABCDEFGH [6], *nrf*ABCDEFG, and *nir*BDC–*cys*G [7]. Expression of these genes is regulated depending on the nitrate–nitrite ratio in the medium.

Group 2 includes genes coding for electron donor dehydrogenases: *fdn*GHI [8], *hya*ABCDEF, *hyb*OABCDEFG [9], and *nuo*ABCDEFGHIJKLMN [10]. Since the products of these genes are essential for the formation of complete respiratory chains, their expression is activated by NarL and NarP.

Group 3 includes genes coding for electron acceptor reductases: *dms*ABC [11], *tor*CAD [12], and *frd*ABCD [5]. Electron acceptor reductases encoded by these genes are utilized under anaerobic conditions in the absence of nitrate and nitrite. When they are present in the medium, nitrate and nitrite inhibit expression of these genes.

Group 4 includes the *dcu*B–*fum*B genes, which are involved in four-carbon dicarboxylic acid metabolism [13]. Their products are enzymes and transporters, which provide substrates for fumarate respiration; this

0026-8933/05/3905-0727

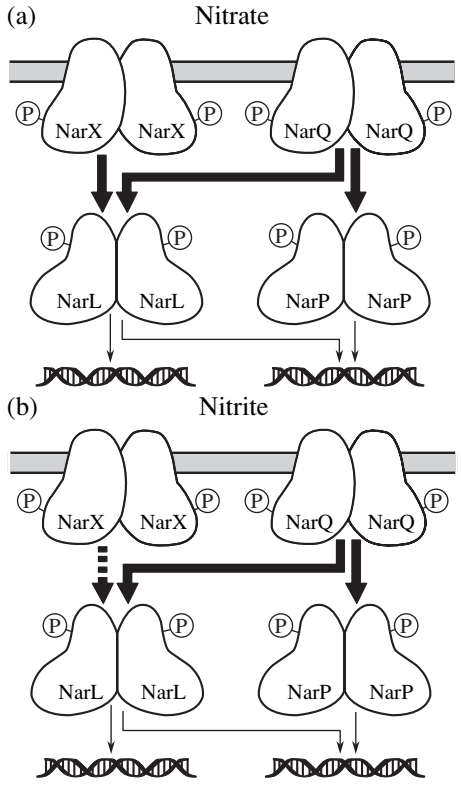

Fig. 1. System of differential response of *E. coli* cells to (a) nitrate and (b) nitrite. Thin arrows indicate binding sites for the regulatory proteins on DNA. Thick arrows show the interactions between the sensory and the regulatory proteins. Phosphorylation and dephosphorylation of the regulatory proteins are shown with solid and dashed arrows, respectively.

is far less efficient than nitrate and nitrite respiration. Hence, the genes are suppressed by NarL and NarP.

Group 5 includes fermentation enzyme genes, *foc*A–*pfl*B [14] and *adh*E [15]. Since fermentation is a less efficient mode of energy production compared to respiration, the genes are also suppressed in the presence of nitrate and nitrite.

Although regulation of nitrite and nitrate respiration has been a subject of intensive research over the past 20 years, many important issues are still poorly understood. The objective of this work was to study the regulation of nitrate and nitrite respiration in γ-proteobacteria. Focusing on organisms that possess only one regulator, NarP, we analyzed the structure of NarP regulons in the genomes of ten γ-proteobacteria and described a generalized NarP regulon for these organisms.

EXPERIMENTAL

Procedure of identifying genes involved in the generalized regulon. Potential NarP-binding sites (hereafter referred to as potential sites) were sought using a nucleotide position weight matrix [16], which was constructed as described below.

Genes were assigned to the regulon according to the results of correspondence testing. A gene was considered to belong to the regulon when a potential site was contained in its regulatory region and upstream of its orthologs from related genomes [16]. In this work, potential sites were sought in the region –400… +100 bp relative to start of translation. In the case of an unknown operon structure of a particular sequence, genes were assigned to one operon when they were transcribed in the same direction and were no more than 100 kb apart.

To identify the members of the generalized regulon, we used a new method based on pairwise comparisons of the genomes of organisms belonging to the same taxonomic group. Each genome was searched for potential sites. Putative members of the generalized regulon were selected according to the presence of potential sites upstream of orthologous genes. When a site was detected upstream of a particular gene in at least three genomes of organisms from one group, the gene was assigned to the generalized regulon. In addition, we checked whether such sites occur upstream of orthologs of the selected genes in organisms from other groups. This approach was used to examine the genomes of bacteria from the families Pasteurellaceae and Vibrionaceae.

In the family Enterobacteriaceae, we examined the genomes of only two organisms, *Yersinia pestis* and *Y. enterocolitica*, and the above approach was thereby inapplicable. In this case, a gene was considered as a potential member of the generalized regulon when upstream potential sites were present in both genomes. Then, orthologous genes from other genomes were tested for such potential sites. When the sites were preserved, the gene was assigned to the generalized regulon.

Some enzymes of the respiratory system occur in several isoforms, which are products of paralogous genes. The regulation of such paralogs is considered separately (see Discussion).

Software. Potential sites were sought and orthologs analyzed using the GenomeExplorer program package [17]. Position weight matrices were constructed using the SignalX program [17]. A search for homologs in databases was performed using the BLAST program [18]. Multiple alignments of amino acid and nucleotide sequences were obtained using the ClustalX program [19]. Phylogenetic trees were constructed by the maximum likelihood method [20], using the program proml from the PHYLIP 3.63 package (http://evolution.genetics.washington.edu/). To construct Logo diagrams, which reflect the structure of a regulatory signal, we used the WebLogo program (http://weblogo.berkeley.edu/).

Subjects. We examined the genomes of 13 γ-proteobacteria. Complete genome sequences of *E. coli* K-12 (EC) [22], *Salmonella typhi* Ty2 (ST) [23],

Erwinia carotovora s. *atroseptica* (EO) [24], *Y. pestis* KIM (YP) [25], *Haemophilus ducreyi* 35000HP (HD), *Haemophilus influenzae* Rd (HI) [26], *Pasteurella multocida* Pm70 (PM) [27], *Vibrio cholerae* O1 (VC) [28], *Vibrio parahaemolyticus* RIMD 2210633 (VP) [29], and *Vibrio vulnificus* CMCP6 (VV) [30] were extracted from GenBank. The complete sequence of the *Y. enterocolitica* genome was obtained at http://www.sanger.ac.uk/. A draft sequence of the *Actinobacillus actinomycetemcomitans* HK1651 (AA) was obtained at http://www.genome.ou.edu/. A draft sequence of the *Vibrio fischeri* ES114 (VF) genome was extracted from the GOLD database [32].

RESULTS

Evolution of the Regulatory System

First, orthologs of the *E. coli* regulatory system genes were sought in the genomes of bacteria belonging to three families: Enterobacteriaceae (*S. typhi, E. carotovora, Y. pestis*, and *Y. enterocolitica*), Pasteurellaceae (*P. multocida, A. actinomycetemcomitans, H. ducreyi*, and *H. influenzae*), and Vibrionaceae (*V. cholerae, V. parahaemolyticus, V. vulnificus*, and *V. fischeri*). Genes of the dual regulatory system were found only in the genomes of *S. typhi* and *E. carotovora*, which are most closely related to *E. coli.* On the phylogenetic trees of NarL/NarP (Fig. 2a) and NarQ/NarX (Fig. 2b), proteins from different organisms formed two clusters separated by a long branch. It seems that duplication yielding a dual system took place before the divergence of evolutionary branches corresponding to the families Enterobacteriaceae, Pasteurellaceae, and Vibrionaceae. The genomes coding for a single system combine *nar*P with *nar*Q. The genomes of *Y. pestis* and *Y. enterocolitica* are exceptions, each coding for the unusual pair NarP–NarX (Fig. 2). In view of this, the NarX amino acid sequences of these organisms were considered in more detail.

The differential response to nitrate and nitrite is known to be due to the periplasmic domain of NarX [33] (Fig. 3a). However, no considerable difference in amino acid sequence has been observed between the periplasmic modules of NarX and NarQ [34].

On the other hand, structural features of the central module allow a distinction between NarX and NarQ. The amino acid sequence of the central module lacks homologs apart from NarX and NarQ. The function of the central module is still unclear. An important feature of the central module in *E. coli* NarX is a cluster of cysteines located in positions 308, 313, and 316. These cysteines are preserved in the NarX sequences encoded by other proteobacterial genomes; substitution of any cysteine of the cluster dramatically reduces the activity of the protein [34].

A multiple sequence alignment of NarX and NarQ of all organisms examined in this work is shown in Fig. 3b. The NarX proteins of *Y. pestis* and *Y. enterocolitica* lack the cysteine residues characteristic of NarX. Thus, although similar in sequence to NarX, the sensory proteins of bacteria of the genus *Yersinia* probably act as NarQ, suggesting a similarity of responses to nitrite and nitrate.

Construction of a Matrix for Detecting Potential NarP-Binding Sites

Since data on the structure of NarL-binding sites are discrepant [35, 36] and the regulation is difficult to analyze in the case of the dual system, we focused on the NarP-dependent regulation in the ten organisms whose genomes contain *nar*P but not *nar*L.

We observed that *nar*P is always present in genomes that contain genes for periplasmic nitrate and nitrite reductases (*nap* and *nrf*, respectively) and genes for heme transport into the periplasm (*ccm*) with various operon rearrangements (Table 1). The exceptions are the *Y. pestis, Y. enterocolitica*, and *V. cholerae* genomes, which lack the *nrf* genes. It has been shown with *E. coli* that expression of all these genes is regulated by NarP [6, 7]. The NarP-binding site is a palindrome with the consensus TACYYMTNNADRRGTA [4]. Sites of this type occur upstream of operons containing the *nap, nrf*, and *ccm* genes in various genomes (Table 2). We constructed a teaching sample, a matrix for detecting NarP-binding sites, and a Logo diagram (Fig. 4).

The genomes were searched for sites having a weight higher than the threshold (3.50). However, sites with a weight lower than the threshold were often found upstream of potential genes of the regulon in the *A. actinomycetemcomitans* genome. Hence the threshold was set at 3.25 for this genome. A search with these parameters revealed potential sites in upstream regions of approximately 400 genes in each genome. It is clear that the prediction was unreliable in some cases. Yet correspondence testing (see Experimental) leaves only a few false-positive predictions, as demonstrated more than once [37].

Structure of the Generalized NarP Regulon in the Genomes Examined

A search for regulon members with the use of our procedure (see Experimental) assigned 77 genes to the generalized regulon; these genes are organized in at least 29 operons. The generalized NarP regulon includes almost all genes belonging to the dual NarL–NarP regulon of the *E. coli* genome (Table 3). The only exceptions are the genes that lack orthologs in the genomes under study. These genes are from the *nar*GHJI, *nar*K, *hya, hyb, fum*B, and *nuo* operons.

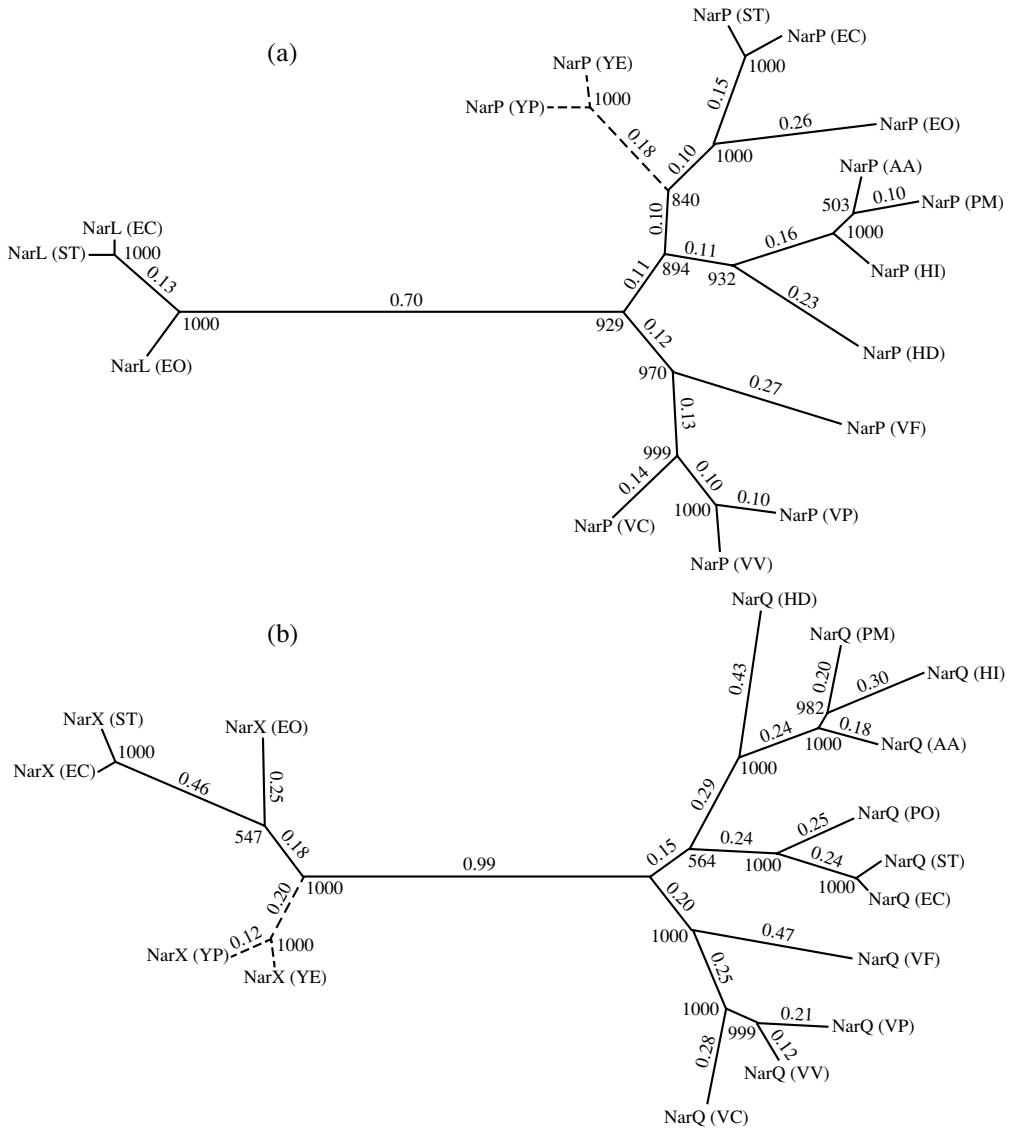

Fig. 2. Phylogenetic trees of (a) NarL/NarP and (b) NarX/NarQ. The trees were constructed by the maximum likelihood method. Here and in Fig. 5, species are designated as in the Experimental. The branches corresponding to *Y. pestis* and *Y. enterocolitica* proteins are shown with dashed lines. The expected proportion of amino acid substitutions is indicated for branches and the percentage of bootstrap replications (maximum, 1000), for knots.

Orthologs of the operon genes were found only in the *Y. pestis* and *Y. enterocolitica* genomes, but potential sites were not reliably detected in their upstream regions. Thus, the single two-component system regulates respiration in response to nitrate and nitrite in *Y. pestis* and *Y. enterocolitica.*

Several operons were included into the generalized NarP regulon on the basis of the formal criterion described in the Experimental. However, multiple sequence alignment of their regulatory regions showed that their potential sites are not conserved; i.e., the predictions are not sufficiently reliable. This was the case with the *ldh*A, *hem*R, *gcv*A, and *suc*AB operons. It is still advisable to experimentally verify the results of our computer analysis.

As new members of the generalized regulon were identified, two new functional groups were added to the above five groups (see Introduction). The seven groups are considered in detail below.

Regulatory protein genes. The most unexpected finding was the presence of potential sites upstream of the genes whose products are involved in regulating respiration. The *P. multocida, A. actinomycetemcomitans, H. influenzae, V. cholerae*, and *V. fischeri* genomes proved to contain potential sites upstream of *fnr*, which codes for a global respiration regulator

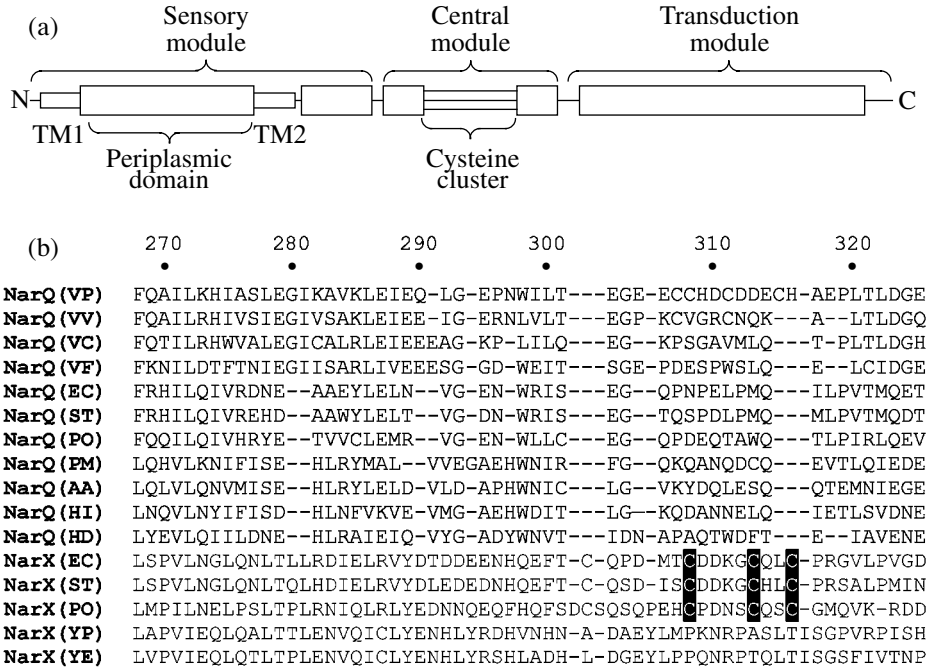

Fig. 3. (a) Structure of the NarX and NarQ sensory proteins [34]. TM, transmembrane segment. (b) Amino acid sequence alignment of a fragment of the central module of NarX and NarQ. Amino acid positions are shown at the top for *E. coli* NarX. The conserved cysteines are shaded.

responsible for switching between aerobic and anaerobic metabolism.

All genomes of Vibrionaceae species have potential sites upstream of the *nar*QP operon, in a highly conserved region downstream of the putative promoters. Another conserved region harbors potential binding sites for FNR (Fig. 5).

Nitrate and nitrite reduction genes. The genes directly involved in reduction of nitrate and nitrite form the core of the generalized NarP regulon. Some of these gene code for periplasmic nitrate reductase (*nap*) and for periplasmic (*nrf*) and cytoplasmic (*nir*) nitrite reductases; the others are responsible for exporting heme into the periplasm and the formation of the Nap complex (*ccm*). Potential sites upstream of the corresponding operons were detected in all cases.

Genes for electron donor dehydrogenases. It has been shown earlier that only the *fdn*GHI formate dehydrogenase operon is regulated by NarL and NarP [38]. This operon is homologous to the *fdo* operon and plays the same role. Potential sites upstream of the *fdo*

Table 1. Structures of the *nap*, *ccm*, and *nrf* operons

Genome	Operon		
	nap	*ccm*	*nrf*
Y. pestis, *Y. enterocolitica*	*napFDABC*	*ccmABCDEFGH*	Absent
P. multocida	*napFDAGHBC*	*ccmABCDEFGH*	*nrfABCDEXFGH*
A. actinomycetemcomitans	*napDAGHBC*	*ccmABCDEFGH*	*nrfABCDEXFGH*
H. influenzae	*napFDAGHBC*	*ccmABCDEFGH*	*nrfABCD nrfEXFGH*
H. ducreyi	*napFDAGHBC*	*ccmABCDEF ccmGH*	*nrfABCD nrfEXFGH*
V. vulnificus, *V. parahaemolyticus*	*napFDABC*	*ccmABCDEFGH*	*nrfA*↔*nrfBCDEXF nrfG*
V. cholerae	*napFDABC*	*ccmABCDEFGH*	Absent
V. fischeri	*napFDABC*	*ccmABCDEFGH*	*nrfA*↔*nrfBCDEXF nrfG*

Note: ↔, genes form a divergon.

Table 2. Potential NarP-binding sites included in the teaching sample (for comments, see text; the position is indicated relative to the start of the open reading frame)

Genome	Operon	Site	Weight	Position
Y. pestis	*napFDABC*	T a ACTCT a a AGAGT a A	4.48	–184
	ccmABCDEFGH	TACC t c TAT A a g GGT A	4.67	–113
Y. enterocolitica	*napFDABC*	T a ACTCT a a AGAGT a A	4.48	–182
	ccmABCDEFGH	TACC t c TAT A a g GGT A	4.67	–113
		T A c T a C g T A a Gg A t T A	3.83	–79
P. multocida	*ccmABCDEFGH*	c ACCTCT a a AGAGGT a	4.80	–99
	nrfABCDEXFG	T A a CT c T a a A t AGg T A	4.78	–289
		TAC t T a T t t Ag A g GT A	4.71	–206
A. actinomycetemcomitans	*nrfABCDEXFG*	TAC t a AT a a AT a g GT A	4.32	–343
H. influenzae	*ccmABCDEFGH*	T A c t TCT a a AGA g t T A	4.79	–85
H. ducreyi	*napFDAGHBC*	T a c CT a a AT g g AG t g A	3.62	–81
	ccmABCDEFGH	TAC c TCT a a AGA a GT A	4.94	–86
		TAC t a T a a a a A a g GT A	3.82	–117
V. vulnificus	*napFDABC*	TACCCC c T A a GGGGT A	4.71	–126
	ccmABCDEFGH	T c a CTCT AT AGAGg t A	4.49	–147
	nrfBCDEXF	TACCCCT a a AGGGGT A	4.91	–304
	nrfA	TACCCCT t t AGGGGT A	4.91	–193
V. parahaemolyticus	*napFDABC*	TACCTC c T A a GAGGT A	4.97	–124
	napGH	TACCTC t T A t GAGGT A	4.88	–82
	nrfBCDEXF	TACC c CT a a AG t GGT A	4.88	–300
		T a c c TC t T Ag GA a t a A	4.21	–196
	nrfA	TACC a CT t t AGg GGT A	4.88	–190
		T t a t TC c T A a GA g g t A	4.21	–294
V. cholerae	*napF*	TACCTC c T A a GAGGT A	4.97	–79
		c A c C t CT t t AGg G c T a	4.16	–153
	ccmABCDEFGH	TAAC t a c a a a g g GT T A	4.23	–143
V. fischeri	*napF*	T A a CC a t T A t g GGg T A	4.23	–85
	ccmA	TACC t a c a a a g t GGT A	4.51	–190
	nrfA	T A a C a CT T AAG a Gg T A	4.93	–234
		TAC c t c T T AA t t a GT A	4.62	–144
		t ACC t c T t t A a t GGT g	4.33	–218
	nrfB	T A c C t CT T AAG t G t T A	4.93	–253
		TAC t a a T T AA g a g GT A	4.62	–343
		c ACC a t T a a Ag a GGT a	4.33	–269

operon were detected in the *Y. pestis* and *Y. enterocolitica* genomes. The *fdh*D gene with a potential site in the regulatory region was found in four genomes. The exact function of this gene is unknown; its product is presumably necessary for the formate dehydrogenase activity [39].

Another new member of the generalized regulon is the *nrq*ABCDEF operon, which codes for NADH dehydrogenase exporting sodium ions [40].

Genes involved in reduction of alternative electron acceptors. Potential sites were found upstream of the operons coding for dehydrogenases of alternative electron acceptors. This group includes the *tor*YZ and *cyd*AB operons, two new members of the regulon. The former codes for the second trimethylamine N-oxide reductase: *tor*Y and *tor*Z are paralogs of *tor*C and *tor*A, respectively [41]. The *cyd*AB operon codes for the cytochrome *d* oxidase complex. This enzyme catalyzes electron transfer from ubiquinol-8 to molec-

Fig. 4. Logo diagram for NarP-binding sites. Abscissa, nucleotide position. Ordinate, the information content (bits). The total height of the column shows the information content of the given position. The relative heights of the letters show the frequencies of particular nucleotides in the given position.

ular oxygen and thereby plays a role in aerobic respiration [42]. This operon is a member of the NarP regulon in many genomes (Table 3).

Four-carbon dicarboxylic acid metabolism genes. NarL-dependent regulation of the *dcu*B–*fum*B operon has been demonstrated experimentally in *E. coli* [13]. Some genomes have potential sites upstream of *dcu*B, while *fum*B orthologs were not found in the genomes examined. The generalized regulon includes *dcu*A, which is paralogous to *dcu*B and similarly codes for a dicarboxylic acid transporter [13]. In addition, potential sites were found upstream of *asp*A and two paralogs, *fum*B and *fum*C. The *fum*B and *fum*C genes each code for fumarase [43]. The *asp*A product is aspartate-ammonium lyase, another enzyme involved in metabolism of four-carbon dicarboxylic acids [44].

Fermentation enzyme genes. The *E. coli* NarL-NarP regulon includes *adh*E, coding for alcohol dehydrogenase [15], and *pfl*B, coding for pyruvate-formate lyase [14]. In *E. coli, pfl*B is transcribed together with the *foc*A formate transporter gene to yield the *foc*A–*pfl*B transcript or alone from its own promoter. The NarL-binding sites occur upstream of the *fuc*A–*pfl*B operon but not upstream of the internal promoter [14]. The *Y. pestis* and *Y. enterocolitica* genomes also contain a putative *foc*A–*pfl*B operon, suggesting a similar mode of transcription. This assumption is supported by the fact that potential NarP-binding sites were found upstream of *foc*A in these genomes (data not shown). Moreover, such sites were also found upstream of *pfl*B. The two genes are separated in the other genomes, and the upstream region of a single *foc*A lacks potential sites with considerable homology. Hence, it is possible to assume that the regulation of *pfl*B plays a key role. The generalized regulon probably includes *yfi*D, which is paralogous to *pfl*B and also codes for pyruvate-formate lyase [45].

Two other genes were assigned to the generalized NarP regulon on evidence of conservation of potential sites among different genomes: *ldh*A for lactate dehydrogenase [46] and *pgk* for phosphoglycerate kinase [47]. As follows from the results of multiple sequence alignment of the regulatory regions, potential sites found upstream of *pgk* are similar in position and overlap the putative promoters in the Pasteurellaceae genomes. Potential sites found upstream of *ldh*A vary in position in the Vibrionaceae genomes (data not shown), which makes the NarP-dependent regulation of this gene questionable.

Genes for tricarboxylic acid cycle enzymes. Two operons coding for tricarboxylic acid cycle enzymes are probably components of the generalized NarP regulon. The product of *mdh* is malate dehydrogenase [48], and the *suc*AB operon codes for two 2-oxoglutarate dehydrogenase subunits [49]. Potential sites

```
                                                          NarP
VP   AAATTGTAGTTTTTGACATATTTGTATAAATACCTCTTAGGAGGTAATTG-GGATGC
VV   AAATTGTAGTTTTTGACATATTTGTATAAATACCCCTTAGGGGGTAATTA-AGAGGC
VC   AAATTGTAGTTTTTGACATATTTGTATAAATACCTCTTAGGAGGTAATTG-GGGTGG
VF   GAATTTTAAATTTTGTCATATTTATTGATATACCCCATAATGGTTAAACATAGGTA-
       **** **  ***** ******* *  * ***** * **   * ***    *

                FNR
VP   AATATTGTTTTGGATCAATAAATTCCATCGCAGGTGATCACATATTGATAAAGTTGA
VV   AATGTTGTTTTAGATCAATAAATTCCGGCGAGAGTGATCACATATTGATATTGTTAA
VC   TGGATTGTTTTGGATCAATAAATTAACCTGCCCGTT-TCGGATAGTG-TGTAGCCCT
VF   AATTTTGTTTTAGATCAAT---TTTTAAAATAAATAA--AAACAACAAAATAGATCC
       ******* *******   **          *      * *        *

VP    AAGTCAATGGGTTCATTGCGTGTACAAGTTGGACAATTAATAGGCGAGGAACTC-T
VV    CAGAAAAACGTGAAACGGAGAGGCTTCGCTGC-CGTTTTGTTAGTGAAGAGGTTAT
VC    AA---AGAGGTG-------GAAGCTTGGTTA-----TCAATCAGAGA-GAGACTGC
VF    CATTTGCTTCTTT-----TAATCATAGGTCA--CAATTAACAAAC-ATGAAAC-AT
       *                                  *        * *
```

Fig. 5. Multiple sequence alignment of the regulatory region of the *nar*QP operon for bacteria of the family Vibrionaceae. The coding region is italicized, the start codon is in bold italics, potential NarP-binding sites are boldfaced, NFR-binding sites are framed, putative –35 and –10 promoter elements are underlined, and conserved positions are marked with asterisks.

Table 3. Potential sites upstream of the genes included in the generalized NarP regulon

Operon	Reference	Genome									
		Enterobacteriaceae		Pasteurellaceae				Vibrionaceae			
		YP	YE	PM	AA	HI	HD	VV	VP	VC	VF
Genes for regulatory proteins											
fnr	*	–	–	3.72 (–23)	3.41 (–30)	3.58 (–71)	–	–	–	4.84 (–60)	3.90 (–176)
narQP	*	–	–	–	–	–	–	4.73 (–188)	4.99 (–139)	4.99 (–94)	4.25 (–117)
Genes for nitrate and nitrite reduction											
napFDAGHBC	[6]	4.48 (–184)	4.48 (–182)	3.24 (–251)	3.47 (–135)	3.92 (–87)	3.62 (–81)	4.73 (–126)[1)]	4.99 (–124)[1)]	4.99 (–79)[1)]	4.25 (–85)[1)]
								3.73 (–101)[2)]	4.90 (–82)[2)]	4.17 (–153)[1)]	
ccmABCDEFGH	[6]	4.67 (–113)	4.67 (–113)	4.80 (–99)	3.57 (–99)	4.79 (–85)	4.95 (–86)	4.48 (–147)	3.53 (–145)	4.23 (–143)	4.52 (–190)
			3.83 (–179)								
nrfABCDEFG	[7]	0	0	4.97 (–289)	4.33 (–343)	3.52 (–71)	3.32 (–182)	4.92 (–304)[3)]	4.89 (–300)[3)]	0	4.94 (–253)[3)]
				4.72 (–206)	3.55 (–365)			3.87 (–202)[3)]	4.22 (–196)[3)]		4.64 (–343)[3)]
											4.34 (–269)[3)]
											3.64 (–165)[3)]
nirBDC-cysG	[7]	4.75 (–99)	4.75 (–101)	0	0	0	0	4.47 (–17)	4.47 (–17)	0	4.25 (–113)
								4.00 (–105)	4.00 (–107)		3.61 (–99)
Genes for reduction of alternative electron acceptors											
torCAD	[12]	0	0	4.14 (–268)	0	0	0	–	–	–	3.77 (–31)
torYZ	*	0	4.35 (–179)	0	3.64 (–63)	4.13 (–41)	3.57 (–201)	0	4.29 (–61)	4.56 (–45)	4.46 (–67)
dmsABC	[11]	3.90 (–178)	0	3.92 (–38)	3.68 (–44)	4.28 (–104)	0	0	–	0	4.42 (–87)
frdABCD	[5]	–	–	4.36 (–141)	4.20 (–270)	4.02 (–150)	–	3.54 (–85)	3.54 (–87)	4.36 (–134)	3.84 (–318)
						3.55 (–47)					
cydAB	*	4.02 (–372)	–	–	3.44 (–107)	3.60 (–133)	4.05 (–152)	4.59 (–389)	4.04 (–359)	3.52 (–101)	4.07 (–118)
							3.51 (–111)				3.84 (–78)
Genes for dehydrogenases of electron donors											
fdnGHI	[8]	0	0	3.68 (–234)	3.51 (–111)	–	0	0	0	0	0
fdoGHI	*	3.64 (–177)	3.64 (–167)	0	0	0	0	0	0	0	0
fdhD	*	3.64 (–10)	3.64 (–16)	3.68 (–24)	3.44 (–10)	–	–	0	–	0	0
nqrABCDEF	*	–	–	4.43 (–376)	4.36 (–351)	–	4.27 (–308)	–	–	–	3.69 (–251)
				4.13 (–351)							

Table 3. (Contd.)

Operon	Reference	Genome									
		Enterobacteriaceae		Pasteurellaceae				Vibrionaceae			
		YP	YE	PM	AA	HI	HD	VV	VP	VC	VF
Genes involved in four-carbon dicarboxylic acid metabolism											
dcuA	*	–	–	4.06 (–105)	3.45 (–19)	3.79 (–76)	–	–	3.65 (–22)	–	3.71 (–34)
											3.60 (–67)
dcuB	[13]	–	–	4.56 (–22)	4.64 (–56)	0	4.34 (–97)	–	–	–	3.55 (–196)
				3.65 (–105)	3.85 (–392)						
aspA	*	–	–	–	–	–	3.63 (–66)	4.89 (–224)	4.26(–226)	0	4.64 (–141)
fumC	*	–	–	3.65 (–64)	4.51 (–20)	3.62 (–62)	–	0	–	–	0
Genes for fermentation enzymes											
pflB	[14]	3.52 (–366)	3.70 (–22)	3.71 (–17)	3.71 (–17)	–	3.62 (–204)	4.11 (–23)	4.11 (–23)	–	4.22 (–270)
yfiD	*	–	0	–	0	–	3.98 (48)	–	–	3.65 (–220)	0
adhE	[15]	–	–	3.76 (–77)	4.13 (–50)	0	0	4.24 (–207)	3.98 (–191)	4.03 (–191)	3.56 (–234)
pgk	*	–	–	3.57 (–68)	–	3.50 (–71)	3.60 (–22)	–	–	–	–
ldhA	*	–	–	0	–	–	0	–	3.52 (–18)	3.89 (–153)	3.55 (–240)
Genes for tricarboxylic cycle enzymes											
mdh	*	–	–	4.25 (–21)	3.29 (–18)	4.33 (–20)	3.91 (–26)	–	–	–	–
sucAB	*	–	–	3.51 (–43)	–	3.94 (–21)	3.66 (–78)	–	–	–	–
Genes for molybdopterine cofactor synthesis											
moaABCDE	*	–	–	–	4.82 (–296)	3.59 (–323)	3.57 (–239)	3.89 (–255)	3.77 (–217)	4.19 (–233)	4.13(–257)
						3.59 (–305)					
gcvA	*	3.75 (–364)	–	3.57 (–141)	3.48 (–165)	3.57 (–130)	–	–	–	–	–
Genes for sulfur metabolism											
cysJIH	*	–	–	0	0	0	0	4.22 (–99)	4.22 (–27)	–	3.84 (–78)

Note: The genomes are designated as in the Experimental. The works that have assigned the genes to the NarL–NarP regulon in *E. coli* are referred to; * the assignment has not been demonstrated experimentally. The weight of a potential site and its position relative to the translation start (in parentheses) are indicated for each operon; (–), the site upstream of the operon is absent; 0, orthologs of the given genes were not found. Operon rearrangements (see Table 1) are indicated as follows: [1)] a site is upstream of the *napFDABC* operon, [2)] a site is upstream of the *napGH* operon, and [3)] a site is between the *nrfA* and *nrfBCDEXF* operons; the position is indicated relative to the start of *nrfB*.

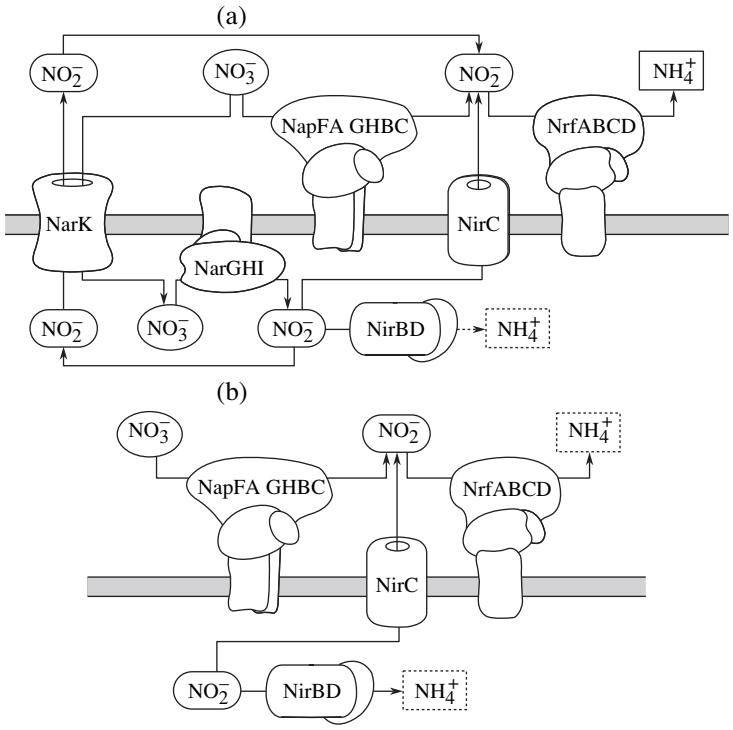

Fig. 6. Nitrate and nitrite reduction systems of (a) *E. coli* and (b) organisms under study.

upstream of these two operons were found only in the Pasteurellaceae genomes (Table 3). The sites upstream of *mdh* are similar in position, while the positions of the sites upstream of *suc*AB vary (data not shown).

Molybdopterine cofactor synthesis genes. Potential sites were found upstream of the *moa*ABDCE operon in seven genomes. The sites are similar in position in the four Vibrionaceae genomes and are scattered and less reliable in the three Pasteurellaceae genomes (data not shown). The products of the *moa*ABCDE operon are responsible for early steps of modybdopterine cofactor synthesis [50].

Some genomes have potential sites upstream of *gcv*A, but the site sequences are poorly conserved (data not shown). GcvA acts as a transcription factor to activate the *gcv*THP operon. The products of this operon are involved in hydrolyzing glycine, which is a source of folate-transferred monocarbon fragments [51]. Incorporation of such fragments transferred by tetrahydrofolates is an essential step in molybdopterine cofactor synthesis [50].

Sulfur metabolism genes. The NarP-dependent regulation revealed for the *cys*JIH operon is specific for species of the family Vibrionaceae. The protein encoded by the operon are involved in reducing 3'-phosphoadenyl sulfate to sulfide [52].

DISCUSSION

Structure of the Nitrate and Nitrite Respiration System in the Genomes Examined

An intricate system for reduction of nitrate and nitrite was formed in *E. coli* cells (Fig. 6a). Nitrate is reduced in the periplasm by the NapABCD complex and in the cytoplasm by the NarGHI complex. Cytoplasmic nitrate reduction is more energy-efficient owing to the specific organization of the respiratory chain, but the cytoplasm accumulates nitrite, which is toxic to the cell. Hence, the nitrite reductase complex NirBD is contained in the cytoplasm to play a protective role. In addition, nitrite is transported into the periplasm by NarK and NirC [1]. In the periplasm, nitrite is reduced by the membrane-associated Nrf complex.

The safer periplasmic pathway is utilized when the nitrate concentration is low. Higher concentrations suppress the periplasmic pathway and activate the cytoplasmic pathway [1]. Since such a system of nitrate and nitrite reduction needs fine adjustment, its regulatory system was duplicated during evolution.

The genomes under study contain only the genes for periplasmic nitrate reductase; i.e., only the safer pathway of nitrate reduction is realized in the respective organisms (Fig. 6b). A simpler, single regulatory system is sufficient in this case. NarQ, which similarly responds to nitrate and nitrite, is utilized as a sensor in this system. The genomes of bacteria of the genus

Yersinia possess *nar*X orthologs. However, certain changes in the encoded amino acid sequence, in particular, the lack of the conserved cysteines, suggest that the protein product acts as NarQ, without differentiation between nitrate and nitrite.

The *Y. pestis* and *Y. enterocolitica* genomes lack genes for periplasmic nitrite reductase, while the *V. cholerae* genome does not code for any nitrite reductase at all. Hence, the organization of the nitrite and nitrate reduction systems in the three organisms needs further investigation, which is beyond the scope of this work.

Autoregulation and Regulatory Cascades

As already mentioned, the Vibrionaceae genomes each contain potential sites upstream of the *nar*QP operon, suggesting its autoregulation. In the *E. coli* genome, the *nar*XL operon is autoregulated, while *nar*Q and *nar*P are not regulated by NarL or NarP [53]. The NarX–NarL system seems to play the central role in regulating nitrate and nitrite respiration in *E. coli.* In the other organisms under study, regulation is due to the single NarQ–NarP system, which takes on responsibility for the regulatory functions performed by the dual system in *E. coli.*

Potential sites were unexpectedly found upstream of *fnr* in five genomes (Table 3). The *fnr* product acts as a global respiration regulator and switches between aerobic and anaerobic metabolism [1].

It should be noted that expression of the *nar*XL operon is suppressed by FNR in *E. coli* [53]. Thus, the role of NarP in the organisms under study probably differs from its role in *E. coli.*

Aerobic Metabolism Genes Regulated by NarP

As mentioned above, the generalized NarP regulon includes several aerobic metabolism genes (*cyd*AB, *mdn*, and *suc*AB) in the organisms under study but not in *E. coli.* The *cyd*AB operon codes for aerobic reductase subunits and, in *E. coli*, is suppressed by FNR in response to a change to anaerobic conditions [42]. Still, distinct potential sites were detected upstream of the *cyd*AB operon in eight genomes.

In addition, two operons coding for tricarboxylic cycle enzymes were assigned to the generalized regulon in this work. Expression of the tricarboxylic cycle genes is regulated by the transcription factors ArcA and FNR in *E. coli* [54, 55]. However, potential sites were detected upstream of these genes in the genomes of bacteria of the family Pasteurellaceae.

The above findings, along with the presence of conserved NarP-binding sites upstream of *fnr*, suggest that the role of NarP in regulating respiration in the organisms under study is more significant than in *E. coli.*

Genes Involved in Molybdopterine Cofactor Synthesis, Heme Transport, and Sulfur Metabolism

It is known that the presence of the molybdopterine cofactor, heme, and iron–sulfur clusters is characteristic of the components of respiratory chains [1]. The molybdopterine cofactor is contained in some respiratory chain enzymes (reductases and dehydrogenases). For instance, the cofactor is essential for the function of the complexes encoded by the *nrf, nap, tor*CAD, *tor*YZ, *dms, fdn*, and *fdo* operons [1]. Since we predicted NarP-dependent regulation for all these operons, it is quite natural that the molybdopterine cofactor synthesis genes belong to the generalized regulon.

The *H. ducreyi, H. influenzae,* and *H. multocida* genomes contain potential sites upstream of *hem*R, which presumably codes for a hemine transporter [26]. However, the sites upstream of this operon are extremely divergent (data not shown), and the assignment of *hem*R to the generalized regulon is questionable.

It is unclear what is the role of potential sites found upstream of the *cys*JHI operon, which is responsible for reduction of sulfur compounds. As already mentioned, the products of these genes are involved in reducing 3'-phosphoadenyl sulfate to sulfide [52]. Reduced sulfur compounds may be necessary for recharging the iron–sulfur clusters in components of respiratory chains.

Regulation of Duplicated Genes

Many genes involved in respiration are duplicated in *E. coli* [1]. In this work, we considered four such duplications: *fdn*GHI/*fdo*GHI, *tor*CAD/*tor*YZ, *pfl*B/*yfi*D, and *dcu*A/*dcu*B. All of these pairs have one feature in common: the two operons code for functionally similar products and substantially differ in expression regulation [1, 8, 13, 45].

Yet the regulatory mechanisms might greatly change during evolution. In view of this, we considered general regulation of the two operons: when a site was conserved upstream of one of the two operons, they were both assigned to the generalized regulon. For instance, potential sites were found upstream of the *fdo*GHI operon in the *Y. pestis* and *Y. enterocolitica* genomes and upstream of the *fdn*GHI operon in the *P. multocida* and *A. actinomycetemcomitans* genomes. The other genomes lack either potential sites upstream of these operons or the operons themselves. In total, potential sites were detected in the four genomes, and, consequently, the two operons were assigned to the NarP regulon.

The *tor*CAD and *yfi*D operons were assigned to the regulon on similar grounds (Table 3).

Regulation of the NADH Dehydrogenase Genes

The genomes of organisms of the family Pasteurellaceae contain potential sites upstream of the *nqr*ABCDEF operon, which codes for NADH dehydrogenase exporting sodium ions. This enzyme was first revealed in the marine bacterium *Vibrio alginolyticus* [40]. In this bacterium, Na^+ is exported against a concentration gradient on account of NADH oxidation, while electrons are transferred to ubiquinones; i.e., the Nqr complex acts as a component of the respiratory chain. Orthologs of the *nqr* genes were found in all genomes under study, but NarP-dependent regulation of the operon is specific for the family Pasteurelaceae.

In *E. coli*, proton-exporting NADH dehydrogenase is encoded by the *nuo* operon, whose transcription is activated by NarL and FNR [10]. The *nuo* operon is present in the *Y. pestis* and *Y. enterocolitica* genomes, but its upstream region lacks potential sites. Since the *nuo* and *nqr* operon genes are nonhomologous, this case provides an example of a nonhomologous substitution with a partial change of the function.

Taxon-Specific Regulation

The ten γ-proteobacterial genomes examined in this work were not studied in terms of NarP-dependent regulation earlier. Moreover, these genomes are far more poorly studied in general, compared to the *E. coli* genome. In view of this, we used a new method to identify members of the generalized regulon.

Our study of the generalized regulon initially started with a search for potential sites in a well-known genome. When sites were conserved among closely related organisms, the corresponding gene was assigned to the regulon [56]. This approach was inapplicable in this work, because two transcription factors, NarL and NarP, simultaneously regulate nitrate and nitrite respiration in *E. coli* and there are almost no reliable data on the structure of NarL-binding sites. Hence, we used a method based on pairwise comparisons of genomes of organisms belonging to particular taxonomic groups (see Experimental).

The new method allowed us to reveal several new members of the generalized regulon. The NarP-dependent regulation of some operons proved to be family-specific. For instance, the involvement of NarP in regulating the *fum*C, *pgk, mdh*, and *suc*AB operons is specific for the family Pasteurellaceae. The regulation of the *ldh*A and *cys*JHI operons and autoregulation of the *nar*QP operons is specific for the family Vibrionaceae. The regulation of some other genes is not strongly specific for a particular family. Yet these genes were assigned to the generalized regulon, because they were identified as regulon members in one taxonomic group and found to have upstream potential sites in the genomes of organisms from another group. These genes include *fnr, asp*A, and *gcv*A.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute (grant no. 55000309), the Russian Science Support Foundation, and the programs Molecular and Cell Biology and The Origin and Evolution of the Biosphere of the Russian Academy of Sciences.

REFERENCES

1. Gennis R.B., Stewart V. 1996 Respiration. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology.* Ed. Neidhart F.C. Washington: ASM Press, pp. 217–286.
2. Stewart V., Rabin R.S. 1995 Dual sensors and dual response regulators interact to control nitrate- and nitrite-responsive gene expression in *Escherichia coli*. In: *Two-Component Signal Transduction*. Eds. Hoch J.A., Silhavy T.J. Washington: ASM Press, pp. 233–252.
3. Rabin R.S., Stewart V. 1993. Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in *Escherichia coli* K-12. *J. Bacteriol*. **175**, 3259–3268.
4. Darwin A.J., Tyson K.L., Busby S.J., Stewart V. 1997. Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement. *Mol. Microbiol*. **25**, 583–595.
5. Li J., Kustu S., Stewart V. 1994. *In vitro* interaction of nitrate-responsive regulatory protein NarL with DNA target sequences in the *fdnG, narG, narK* and *frdA* operon control regions of *Escherichia coli* K-12. *J. Mol. Biol*. **241**, 150–165.
6. Darwin A.J., Ziegelhoffer E.C., Kiley P.J., Stewart V. 1998. Fnr, NarP, and NarL regulation of *Escherichia coli* K-12 *napF* (periplasmic nitrate reductase) operon transcription *in vitro. J. Bacteriol*. **180**, 4208.
7. Wang H., Gunsalus R.P. 2000. The *nrfA* and *nirB* nitrite reductase operons in *Escherichia coli* are expressed differently in response to nitrate than to nitrite. *J. Bacteriol*. **182**, 5813–5822.
8. Wang H., Gunsalus R.P. 2003. Coordinate regulation of the *Escherichia coli* formate dehydrogenase *fdnGHI* and *fdhF* genes in response to nitrate, nitrite, and formate: Roles for NarL and NarP. *J. Bacteriol*. **185**, 5076–5085.
9. Richard D.J., Sawers G., Sargent F., McWalter L., Boxer D.H. 1999. Transcriptional regulation in response to oxygen and nitrate of the operons encoding the [NiFe] hydrogenases 1 and 2 of *Escherichia coli. Microbiology.* **145**, 2903–2912.
10. Bongaerts J., Zoske S., Weidner U., Unden G. 1995. Transcriptional regulation of the proton translocating NADH dehydrogenase genes (*nuoA*-N) of *Escherichia coli* by electron acceptors, electron donors and gene regulator. *Mol. Microbiol*. **16**, 521–534.
11. Bearson S.M., Albrecht J.A., Gunsalus R.P. 2002. Oxygen and nitrate-dependent regulation of *dmsABC* operon expression in *Escherichia coli*: Sites for Fnr and NarL protein interactions. *BMC Microbiol*. **2**, 13.
12. Iuchi S., Lin E.C. 1987. The *nar*L gene product activates the nitrate reductase operon and represses the fumarate

reductase and trimethylamine N-oxide reductase operons in *Escherichia coli. Proc. Natl. Acad. Sci. USA*. **84**, 3901–3905.

13. Golby P., Kelly D.J., Guest J.R., Andrews S.C. 1998. Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli. J. Bacteriol*. **180**, 6586–6596.
14. Kaiser M., Sawers G. 1995. Nitrate repression of the *Escherichia coli pfl* operon is mediated by the dual sensors NarQ and NarX and the dual regulators NarL and NarP. *J. Bacteriol*. **177**, 3647–3655.
15. Chen Y.M., Lin E.C. 1991. Regulation of the *adhE* gene, which encodes ethanol dehydrogenase in *Escherichia coli. J. Bacteriol*. **173**, 8009–8013.
16. Mironov A.A., Koonin E.V., Roytberg M.A., Gelfand M.S. 1999. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res*. **27**, 2981–2989.
17. Mironov A.A., Vinokurova N.P., Gelfand M.S. 2000. Software for analyzing bacterial genomes. *Mol. Biol*. **34**, 253–262.
18. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z M.W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389–3402.
19. 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.
20. Felsenstein J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol*. **266**, 418–427.
21. Schneider T.D., Stephens R.M. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res*. **18**, 6097–6100.
22. Blattner F.R., Plunkett 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.
23. Deng W., Liou S.R., Plunkett G., Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R. 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J. Bacteriol*. **185**, 2330–2337.
24. Bell K.S., Sebaihia M., Pritchard L., Holden M.T., Hyman L.J., Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J., Mungall K., Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J., Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H., Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S., Salmond G.P., Birch P.R., Parkhill J., Toth I.K. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. USA*. **101**, 11105–11110.
25. Deng W., Burland V., Plunkett G., Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R., Perry R.D. 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol*. **184**, 4601–4611.
26. Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. **269**, 496–512.
27. 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.
28. Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae. Nature*. **406**, 477–483.
29. Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S., Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T. 2003. Genome sequence of *Vibrio parahaemolyticus*: A pathogenic mechanism distinct from that of *V. cholerae. Lancet*. **361**, 743–749.
30. Kim Y.R., Lee S.E., Kim C.M., Kim S.Y., Shin E.K., Shin D.H., Chung S.S., Choy H.E., Progulske-Fox A., Hillman J.D., Handfield M., Rhee J.H. 2003. Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect. Immunol*. **71**, 5461–5471.
31. Benson D.A., Boguski M.S., Lipman D.J., Ostell J., Ouellette B.F., Rapp B.A., Wheeler D.L. 1999. GenBank. *Nucleic Acids Res*. **27**, 12–17.
32. Bernal A., Ear U., Kyrpides N. 2001. Genomes OnLine Database (GOLD): A monitor of genome projects worldwide. *Nucleic Acids Res*. **29**, 126–127.
33. Stewart V., Chen L.L., Wu H.C. 2003. Response to culture aeration mediated by the nitrate and nitrite sensor NarQ of *Escherichia coli* K-12. *Mol. Microbiol*. **50**, 1391–1399.
34. Stewart V. 2003. Biochemical Society Special Lecture: Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem. Soc. Trans*. **31**, 1–10.
35. Baikalov I., Schroder I., Kaczor-Grzeskowiak M., Grzeskowiak K., Gunsalus R.P., Dickerson R.E. 1996. Structure of the *Escherichia coli* response regulator NarL. *Biochemistry*. **35**, 11053–11061.
36. Maris A.E., Sawaya M.R., Kaczor-Grzeskowiak M., Jarvis M.R., Bearson S.M., Kopka M.L., Schroder I., Gunsalus R.P., Dickerson R.E. 2002. Dimerization

allows DNA target site recognition by the NarL response regulator. *Nature Struct. Biol*. **9**, 771–778.

37. Gelfand M.S., Laikova O.N. 2003. Prolegomena to the evolution of transcriptional regulation in bacterial genomes. In: *Frontiers in Computitional Genomics*. Eds. Galperin M.Y., Koonin E.V. Wymondham, U.K.: Caiser Acad. Press, pp. 195–216.
38. Darwin A.J., Li J., Stewart V. 1996. Analysis of nitrate regulatory protein NarL-binding sites in the *fdnG* and *narG* operon control regions of *Escherichia coli* K-12. *Mol. Microbiol*. **20**, 621–632.
39. Plunkett G., Burland. V., Daniels D.L., Blattner F.R. 1993. Analysis of the *Escherichia coli* genome: 3. DNA sequence of the region from 87.2 to 89.2 minutes. *Nucleic Acids Res*. **21**, 3391–3398.
40. Tokuda H., Nakamura T., Unemoto T. 1981. Potassium ion is required for the generation of pH-dependent membrane potential and delta pH by the marine bacterium *Vibrio alginolyticus. Biochemistry.* **20**, 4198–4203.
41. Gon S., Patte J.C., Mejean V., Iobbi-Nivol C. 2000. The *torYZ (yecK-bisZ)* operon encodes a third respiratory trimethylamine N-oxide reductase in *Escherichia coli. J. Bacteriol*. **182**, 5779–5786.
42. Cotter P.A., Chepuri V., Gennis R.B., Gunsalus R.P. 1990. Cytochrome *o* (*cyoABCDE*) and *d* (*cydAB*) oxidase gene expression in *Escherichia coli* is regulated by oxygen, pH, and the *fnr* gene product. *J. Bacteriol*. **172**, 6333–6338.
43. Tseng C.P., Yu C.C., Lin H.H., Chang C.Y., Kuo J.T. 2001. Oxygen- and growth rate-dependent regulation of *Escherichia coli* fumarase (FumA, FumB, and FumC) activity. *J. Bacteriol*. **183**, 461–467.
44. Spiro S., Guest J.R. 1991. Adaptive responses to oxygen limitation in *Escherichia coli. Trends Biochem. Sci*. **16**, 310–314.
45. Wyborn N.R., Messenger S.L., Henderson R.A., Sawers G., Roberts R.E., Attwood M.M., Green J. 2002. Expression of the *Escherichia coli yfiD* gene responds to intracellular pH and reduces the accumulation of acidic metabolic end products. *Microbiology.* **148**, 1015–1026.
46. Jiang G.R., Nikolova S., Clark D.P. 2001. Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli. Microbiology.* **147**, 2437–2446.
47. Alefounder P.R., Perham R.N. 1989. Identification, molecular cloning and sequence analysis of a gene cluster encoding the class II fructose 1.6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde 3-phosphate dehydrogenase of *Escherichia coli. Mol. Microbiol*. **3**, 723–732.
48. Park S.J., Cotter P.A., Gunsalus R.P. 1995. Regulation of malate dehydrogenase (*mdh*) gene expression in *Escherichia coli* in response to oxygen, carbon, and heme availability. *J. Bacteriol*. **177**, 6652–6656.
49. Park S.J., Chao G., Gunsalus R. 1997. Aerobic regulation of the *sucABCD* genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: Roles of ArcA, Fnr, and the upstream *sdhCDAB* promoter. *J. Bacteriol*. **178**, 4138–4142.
50. Rajagopalan K.V. 1996 Biosynthesis of the molybdenum cofactor. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology.* Ed. Neidhart F.C. Washington: ASM Press, pp. 674–679.
51. Green J.M., Nichols B.P., Matthews R.G. 1996. Folate biosynthesis, reduction, and polyglutamylation. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology.* Ed. Neidhart F.C. Washington: ASM Press, pp. 665–673.
52. Kredich N.M. 1996. Biosynthesis of cysteine. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology.* Ed. Neidhart F.C. Washington: ASM Press, pp. 514–527.
53. Darwin A.J., Stewart V. 1995. Expression of the *narX, narL, narP*, and *narQ* genes of *Escherichia coli* K-12: Regulation of the regulators. *J. Bacteriol*. **177**, 3865–3869.
54. Lynch A.S., Lin C.C. 1996. Responses to molecular oxygen. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology.* Ed. Neidhart F.C. Washington: ASM Press, pp. 1526–1537.
55. Cronan J.E., Laporte D. 1996. Tricarboxylic acid cycle and glyoxylate bypass. In: *Escherichia coli and Salmonella. Cellular and Molecular Biology.* Ed. Neidhart F.C. Washington: ASM Press, pp. 206–216.
56. Ravcheev D.A., Gelfand M.S., Mironov A.A., Rakhmaninova A.B. 2002. The purine regulon of gamma-proteobacteria: A detailed description. *Genetika*. **38**, 1203–1214.