J. Mol. Microbiol. Biotechnol. (2001) 3(4): 529-543.

JMMB Research Article

Regulation of Aromatic Amino Acid Biosynthesis in Gamma-Proteobacteria

E.M. Panina[1], A.G. Vitreschak[2], A.A. Mironov[3], and M.S. Gelfand[3]*

[1]Department of Mathematics, Moscow State University, Moscow, 119899, Russia
[2]Institute for Problems of Information Transmission, RAS, Moscow, 101447, Russia
[3]State Scientific Center GosNIIGenetika, Moscow, 113545, Russia

Abstract

Computational comparative techniques were applied to analysis of the aromatic amino acid regulons in gamma-proteobacteria. This resulted in characterization of the TrpR and TyrR regulons in the genomes of *Yersinia pestis*, *Haemophilus influenzae*, *Vibrio cholerae* and other bacteria and identification of new members of the PhhR regulon in the genome of *Pseudomonas aeruginosa*. Candidate attenuators were constructed for all studied genomes, including the *trpBA* operon of the very distantly related bacterium *Chlamidia trachomatis*. The *pheA* attenuator of *Y. pestis* is an integration site for the insertion element IS-200. It was shown that the triplication of the DAHP-synthase genes occurred prior to the divergence of families *Enterobacteriaceae, Vibrionaceae* and *Alteromonadaceae*. The candidate allosteric control site of the DAHP-syntheases was identified. This site is deteriorated in AroH of *Buchnera* sp. APS. The known DAHP-synthase of *Bordetella pertussis* is likely to be feedback-inhibited by phenylalanine, and the DAHP-synthase of *Corynebacterium glutamicum* could be inhibited by tyrosine. Overall, the most extensive regulation was observed in *Escherichia coli*, whereas the regulation in other genomes seems to be less developed. At the extreme, the tryptophan production in the aphid endosymbiont *Buchnera* sp. APS is free from transcriptional, attenuation, and allosteric control.

Introduction

Traditional analysis of complete bacterial genomes involves functional gene annotation and metabolic reconstruction. Recently, large-scale genomic analyses were performed in order to study the evolution of metabolic pathways (Galperin and Koonin, 1999; Forst and Schulten, 1999; Dandekar *et al.,* 1999). However, the evolution of regulation is still largely unexplored. One of the reasons for this is the absence of reliable algorithms for recognition of functional sites in DNA sequences (Fickett and Hatzigeorgiou, 1997; Frech *et al.,* 1997).

Availability of genomes of related bacteria allows one to apply the comparative approach to analysis of not only protein sequences, but regulatory patterns as well. At that, the search space is restricted to genes from a specific pathway, thus reducing the combinatorial space for false positives. Another important advantage of simultaneous analysis of several genomes is the possibility to distinguish true sites, occurring upstream of orthologous genes, from false positives scattered at random across the genome. The latter technique allowed us to analyze several regulons and obtain a number of promising predictions (Mironov *et al.,* 1999; Gelfand *et al.,* 1999; Gelfand *et al.,* 2000; Rodionov *et al.,* 2000), some of which were subsequently confirmed in experiments (Kreneva *et al.,* 2000). Similar, although not exactly identical, techniques were developed in (Stojanovich *et al.,* 1999; McGuire *et al.,* 2000; Ramirez-Santos, 2001). For a review of these and related studies see (Gelfand, 1999).

Here we study the evolution of regulation of the aromatic amino acid biosynthesis in the gamma subdivision of proteobacteria. This system is especially interesting, since it involves at least two transcriptional regulators, TrpR and TyrR, attenuation, and allosteric control by feedback inhibition. All these systems were well studied in *E. coli*, whereas the data for other bacteria are sporadic and incomplete. In this paper we consider the regulatory patterns in the bacteria whose genomes are complete or almost complete (*Salmonella typhimurium*, *Yersinia pestis*, *Haemophilus influenzae*, *Vibrio cholerae*, *Buchnera* sp. APS, and *Pseudomonas aeruginosa*). We also mention the regulatory signals in the available fragments of other genomes, if these signals provide additional insight into the evolution of regulatory interactions. We describe the possible site of allosteric control of DAHP-synthases and use it to predict modes of feedback inhibition of this enzyme in several bacterial species. Finally, we discuss the general trends in regulation of aromatic amino acid biosynthesis in this group of bacteria.

Biosynthesis of all three aromatic amino acids starts with the common pathway leading from phosphoenolpyruvate and erythrose 4-phosphate through 3-deoxy-D-arabino-heptulosonate-7-phosphate and shikimate to chorismate. Then the pathway divides into the terminal pathways, specific for each aromatic amino acid. Key enzymes of the latter pathways were extensively studied and all of them had been identified by early 1970s. There are three regulated transporter systems for aromatic amino acids, namely the general permease AroP, the tryptophan permease Mtr, and the tyrosine permease TyrP (Figure 1; for a review see Pittard, 1996).

The transcription of the genes for aromatic amino acid metabolism is regulated by two repressors, TrpR and TyrR

*For correspondence. Email misha@imb.imb.ac.ru; Fax. +7-(095)-315-05-01.

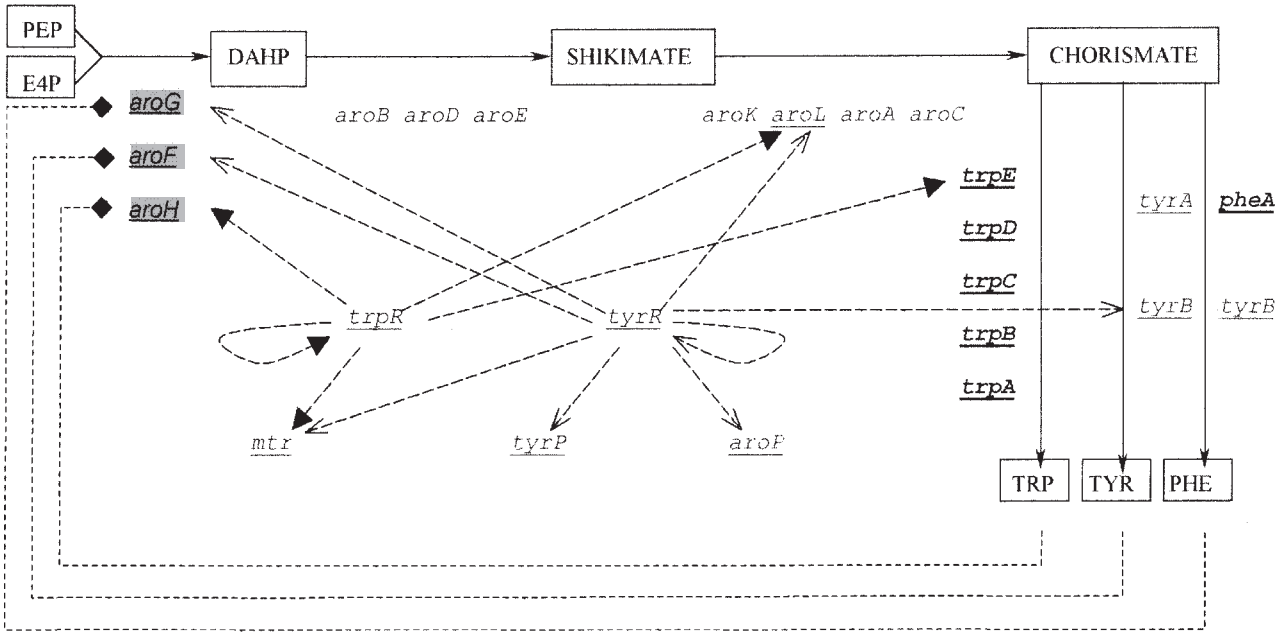


Figure 1. Genes encoding the enzymes of the aromatic amino acids biosynthesis pathway, their regulators and the transporters of tryptophan, tyrosin and phenylalanine. Regulation is shown by dotted lines. Underlined: genes regulated on the level of transcription (filled arrows: by TrpR; empty arrows: by TyrR); **boldface**: genes regulated by attenuation; shaded: enzymes inhibited by the end products (blunt arrows).

(Otwinowski *et al.,* 1988; Pittard and Davidson, 1991). In *E. coli* and *S. typhimurium* the operons *mtr* and *aroL* are regulated by both repressors (Heatwole and Somerville, 1992; Lawley and Pittard, 1994). Other regulatory mechanisms include transcriptional attenuation of the *trp* and *phe* operons (Jackson and Yanofsky, 1973; Yanofsky, 1981; Keller and Calvo, 1979), as well as two more aromatic amino acid-related operons, *pheST* encoding subunits of the phenylalanyl-tRNA synthetase, and *tnaAB* encoding enzymes of tryptophan catabolism (Grunberg-Manago, 1996). The immediate response to the changes in concentration of aromatic amino acids is provided by allosteric control of the first enzyme of the common pathway, DAHP-synthase, by the end products. *E. coli* has three DAHP-synthase isoenzymes, AroF, AroG and AroH, that are feedback inhibited by tyrosine, phenylalanin and tryptophan respectively (Camakaris and Pittard, 1974; Brown and Doy, 1966).

Results and Discussion

Operon structure

The operon *trpEDCBA* encodes all genes of the tryptophan biosynthesis terminal pathway. In *E. coli* it is subject to attenuation and repression by TrpR (Figure 2) (Jackson and Yanofsky, 1973; Bennett and Yanofsky, 1978). Its structure is well characterized (Yanofsky *et al.,* 1981) and tends to be conserved in the gamma-proteobacteria. Among the analyzed genomes, only *H. influenzae, P. aeruginosa* and *Buchnera* sp. have a different structure of the *trp* operon: in *H. influenzae* the operon is divided into two parts, *trpBA* and *trpEGDC*, with independent regulatory cassetes (see below), whereas in *P. aeruginosa* only the genes for the two tryptophan synthase subunits form an operon *trpBA*, and the remaining genes are scattered across the genome. *Buchnera* sp. has a chromosome-encoded operon *trpDCBA* and a plasmid-encoded operon *trpEG*.

A gene annotated as *trpH* in *S. typhimurium* (U92714) is transcribed divergently from the *trp* operon (Figure 2a). The function of this gene is not known. However, its position is conserved in *Enterobacteria* (*E. coli*, *S. typhimurium* and *Y. pestis*), *V. cholerae* and *S. putrefaciens*, which makes it likely that this gene is indeed involved in the tryptophan metabolism (Figure 2b; *cf.* Overbeek *et al.,* 1999).

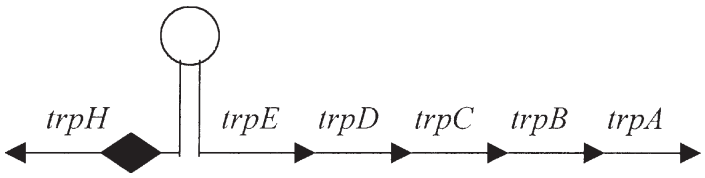
The two last multicistronic operons from the aromatic amino acid regulon in *E. coli* are *aroFtyrA* and *aroLM*. The second gene of the former, *aroM*, is an open reading frame with unknown function that does not have any orthologs in other studied genomes except *S. typhimurium*. The structure of *aroFtyrA* is conserved in *S. typhimurium*, *Y. pestis*, *V. cholerae*, and *S. putrefaciens.* There is no *tyrA* gene in the genome of *P. aeruginosa* and no *aroF* gene in the genome of *H. influenzae.* Both genes are absent in the sequenced part of the *A. actinomycetemcomitans* genome.

Control by Repression

The candidate TRP and TYR boxes upstream of aromatic amino acid synthesis genes from *Y. pestis*, *V. cholerae*, *H. influenzae*, *A. actinomycetemcomitans*, *S. putrefaciens*, *P. aeruginosa* are listed in Table 1. The *S. typhimurium* sites virtually coincide with those of *E. coli* and are not shown. The absence of some expected sites in the *A. actinomycetemcomitans* genome can be easily explained by the assumption that they are outside the sequenced fragments.

There is only a very weak candidate TRP box upstream of *aroH* in *Y. pestis* and *Erwinia herbicola* (U11066) (Figure

A



B

```
EC   -------MSDTNYAVIYDLHSHTTASDGCLTPEALVHRAVEMRVGTLAITDHDTTAAIAP
ST   ---MGAALSDTNYAVIYDLHSHTTASDGLLTPETLVHRAVEMRVGTLAITDHDTTAAIPA
YP   MKELTDATTLSTAFTLYDLHSHTTASDGTLTPSELVIRAAQMRVGVLAITDHDTTAGLAE
VC   -------------MRIDLHSHTTASDGRFTYQQLIDRAVSFEIDVLAITDHDTVAALAD
SP   ----------NTQSILADLHSHTTASDGHLTPSELVARALEKGVQLFAITDHDTVAGLSE
HI   ------------MTKKYDLHCHSTASDGVLSPTELVHRAYAQGVNVLALCDHDTIAGIDE
PA   --------------MRVDLHCHSTASDGVLAPAALVQRAHEKGVGLLALTDHDTLEGLPE
                     ***.*:***** ::   *: **    :  :*: ****  .:

EC   AREEISRSGLALNLIPGVEISTVWENHEIHIVGLNIDITHPLMCEFLAQQTERRNQRAQL
ST   AREEISRCGLALNLIPGVEISTVWENHEIHIVGLNIDIAHPAMRDFLAQQTHRRQARGRL
YP   AAAAIEQHQLALRLIPGVEISTLWENHEIHIVGLGIDVQHVSLRTLLEAQSEHRYQRAQE
VC   ARAYIAQQQYPLQLVNGIEISTVWQNKDIHIVGLNIDPNSEALGQLIARQQQRRVERAEL
SP   ARAFNQAQAEPLKLISGVEISTRWNSYDIHIVALNFDETHAELLRFLAHQRELRFQRAEE
HI   A--EIAAKEVGIELITGVEISTNWEGRGIHIVGLNFDKTHPKMTALLQSQKALREKRAVE
PA   A--HQAAARLGIELVNGVELSCTWGGATIHVLGYGFDLDAAPLRQATEALHRGRWARAEE
     *          :.*: *:*:*  * .  **::. .:*     :          *  *.

EC   IAERLEKAQIPGALEGAQRLAQGG-----AVTRGHFARFLVECGKASSMADVFKKYLARG
ST   IAERLEKAHIPGAWEGALRLANGG-----AVTRGHFARFLVECGKATTMAEVFKKYLARG
YP   ISTRLEKARIVGAWEGANRLANGG-----QVTRGHFARYLVELGLASNVGQVFKKYLAKG
VC   IAHRLQKATREGVLEEVQHIADGA-----PITRAHFAKWLVDNGYATNMQQVFKKYLTRD
SP   IGHRLAKAGIEGAYEGAKKLAAGA-----ALSRGHYARWLADNGHAVDMPSVFKRYLARG
HI   IGDKLEKAGIPNAYDGAKALADG------EVTRAHYARYLVQIGKVSNDGQAFKRYLGQG
PA   IDRRLAAKGMPGALEGARAIQQALGDSGNAPARPHFAEFLVQGGWVKDRADAFRKWLGAG
     *  :*      .. : .  :  .        :* *:*.:*.: * .    ..*:::*  .

EC   KTGYVPPQWCTIEQAIDVIHHSGGKAVLAHPGRYNLSAKWLKRLVAHFAEHHGDAMEVAQ
ST   KTGYVPPQWCTIEQAIDVIHHSGGXXVLAHPGRYDLSAKWLKRLVAHFADHHGDAMEVAQ
YP   KTGYVPAQWCTIEQAIDAIKQSGGQAVLAHPGRYDLTAKWLKRLLAHFAEHGGDAMEVAQ
VC   NPGYVPPNWCSMSEAVSAIHAAGGQAVLAHPGRYDLTAKWLKRLMTAFVEAGGDAMEVAQ
SP   KTGYVPNNWGDMASAIKVIHNAGGLAVLAHPSGYKLSAKWLKRLVREFKEAGGDAMEVVL
HI   KSAFVKAEWADIPTAIETIHAAGGIAIIAHPLRYNMTGKWVRKLIVDFKAWGGDGMEMAD
PA   KLGDVKQHWPTLEEVVGCLREARAWISLAHPYHYDFTRTKRRRLVADFLQAGGHAIEVVN
     : . *  .*  :  .:  :: : .   :***  *.:: .  ::*:  *    *..:*:.

EC   CQQSPNERTQLAALARQHHLWASQGSDFHQPCPWIELGRKLWLPAGVEGVWQLWE-QPQN
ST   CQQSPNERTQLATLARQHHLWASLGSDFHQPCPWIELGRKLWLPAGVEGVWQAWE-QPQI
YP   CQQAPHERALLGQYARDYHLLASQGSDFHQPCSWIELGRKLWLPAGVEPVWRDWPTEPAM
VC   PQQSPQEKRTLGDYAMEYQLLASQGSDFHYPSPWMELGRNLWLPAGVEPVWKDWGLSPSN
SP   GQQTVDDRANLVALSVQNQLLASIGSDFHFPSNWIELGKNLYQPQGIDWVWQSGSWVERP
HI   CGQTKDQRQMLARWAKEFDLQGSVGSDFHFPCGWIELGKNLDLVDSVIPVWEKF------
PA   GVQPAEQVGTLSILAREFGLMVTAGSDFHAPGDWSELGMYRAVPEDLPAMWPRFARHFRD
       *. .:   *   : :  *  : ***** *  * ***       .:  :*

EC   TTEREL----
ST   S---------
YP   PLV-------
VC   RVERLDEE--
SP   ----------
HI   ----------
PA   TERHSEAKGA
```

C

```
EC gggccttagtcgccga***ATGTACTAGAGAACTAGTGCAT***tagcttatttttttgttatcatg
ST ggggcttaaccgcagg**ATGTACTAGTAAACTAGTTTAA**tggattggttttt-gttatcatg
YP agggatgagtcgtcggATGTTTTAACTAAATATTTTCAtgagtgattatttgcggtttata
EH gggcctgactcgccgcATGTACTGATGGGTAACCGGCGctgaactgtaaa-gcgg-cgctg
BA -acttctgtcactgtatttttttagtatactattaaacttatcctccctaatattacaatt
```

Figure 2. A: Regulation of the operon *trpEDCBA* in *E. coli*. The attenuator structure and the TRP box (diamond) are shown. B: Alignment of orthologs of *trpH* from enteric bacteria. C: Tentative alignment of *aroH* upstream regions of enteric bacteria. EH: *Erwinia herbicola*. BA: *Buchnera aphidicola*; other genomes: see Data and Methods. Notation: ***BOLDFACE ITALIC***, TRP box (*E. coli*); **BOLDFACE**, predicted TRP box; CAPITALS, region homologous to the TRP box of *E. coli*.

3A

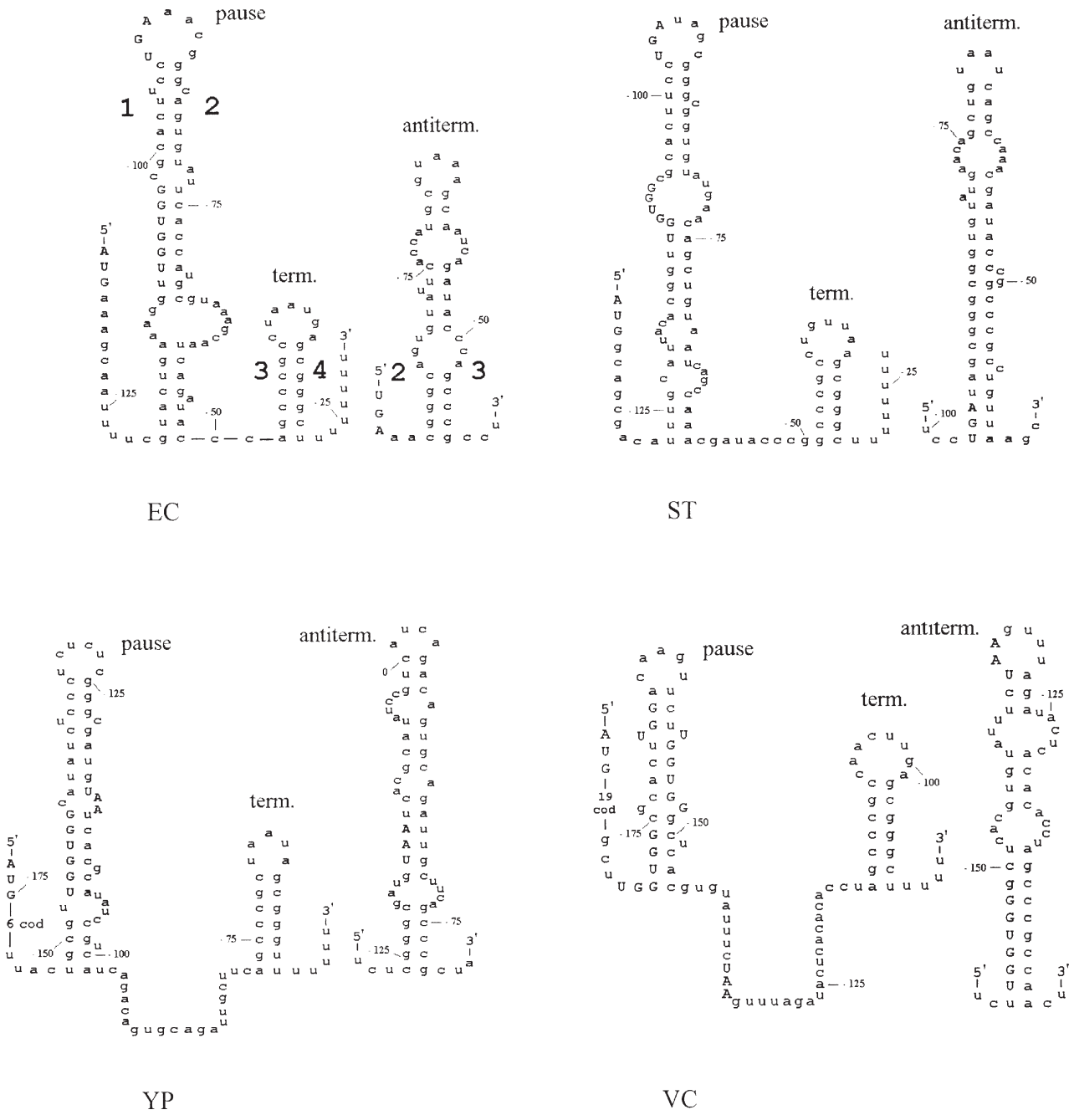
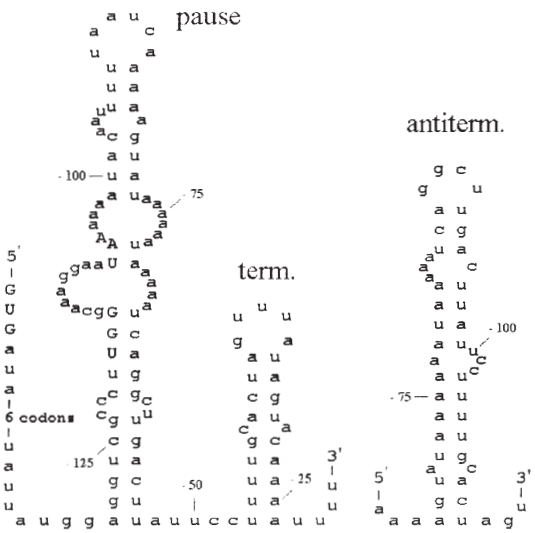
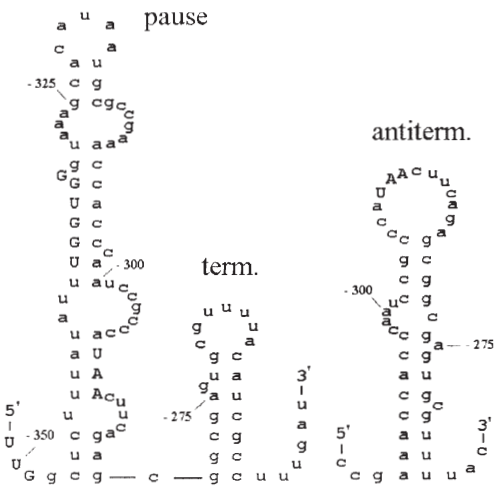


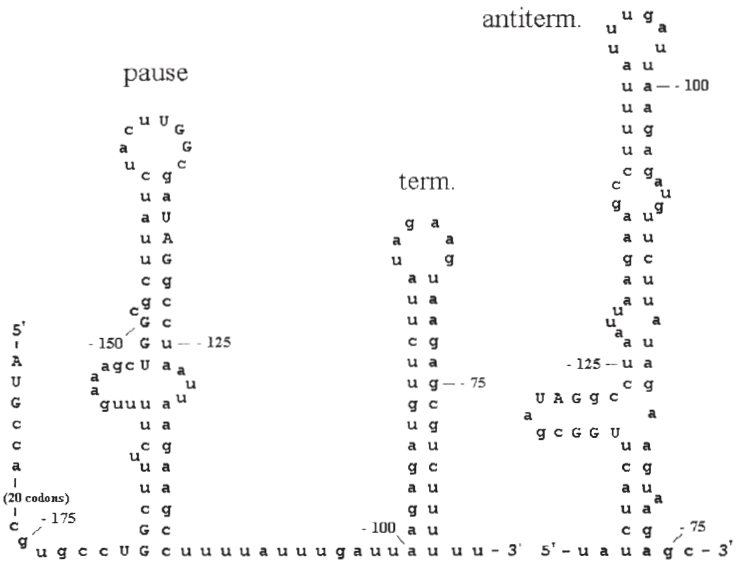
Figure 3. A: Attenuators of *trp* operons. Capitals: tryptophan, start and stop codons in the leader peptide. B. Attenuators of *pheA* operons. Capitals: phenylalanine, start and stop codons in the leader peptide. The arrow in the *E.coli* attenuator corresponds to the IS200 inserstion site of *Y. pestis*.



HI (*trpEGDC*)

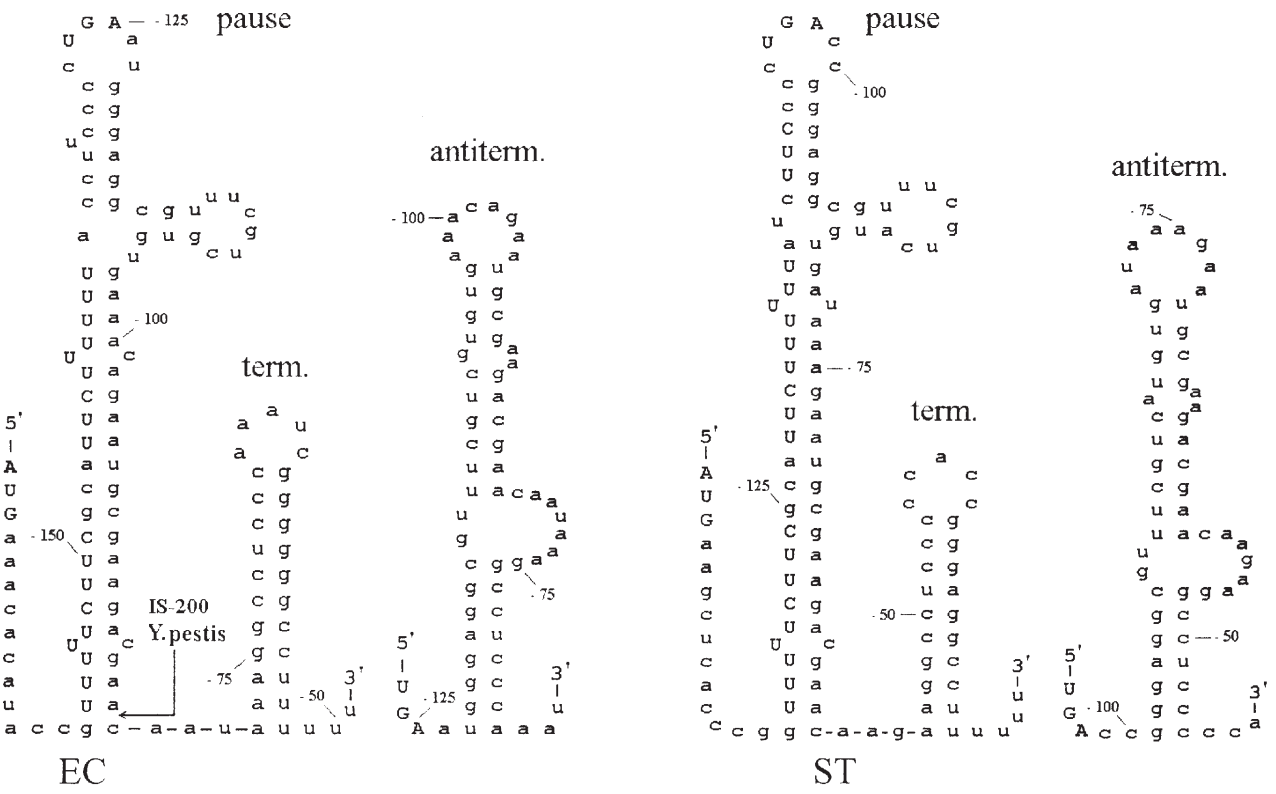


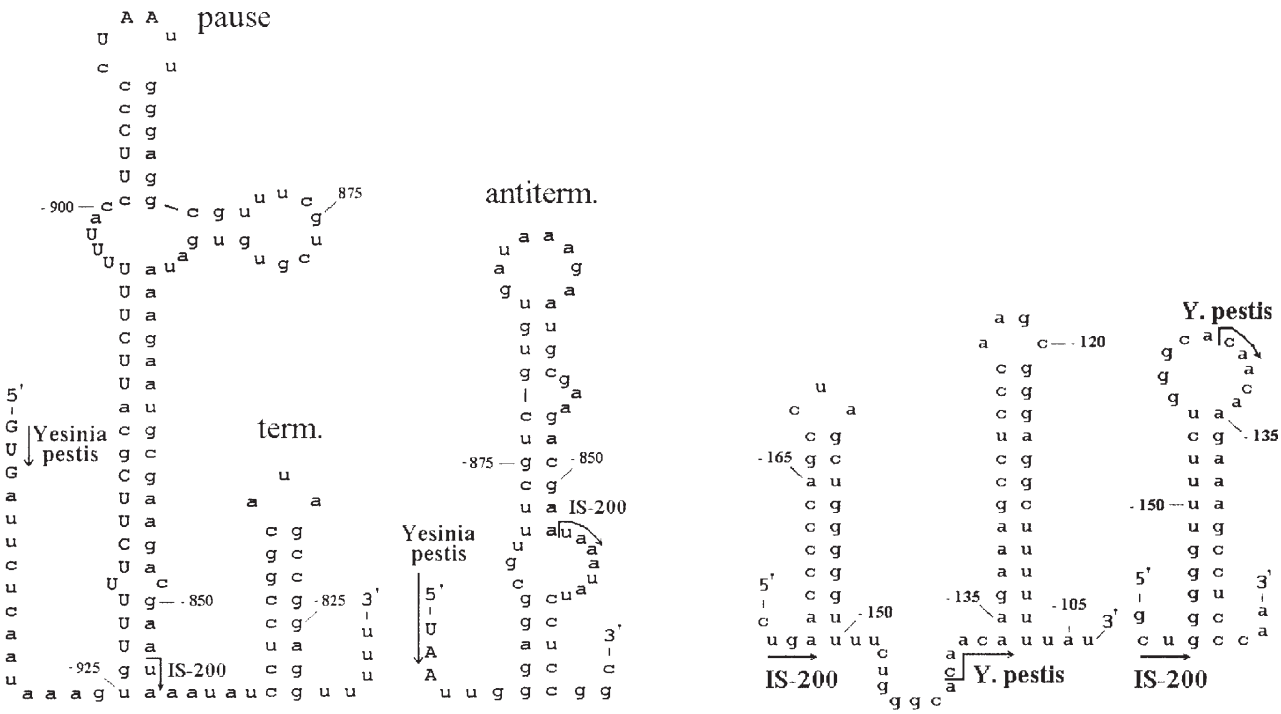
HI (*ydfGtrpBA*)



C. trachomatis (*trpBA*)

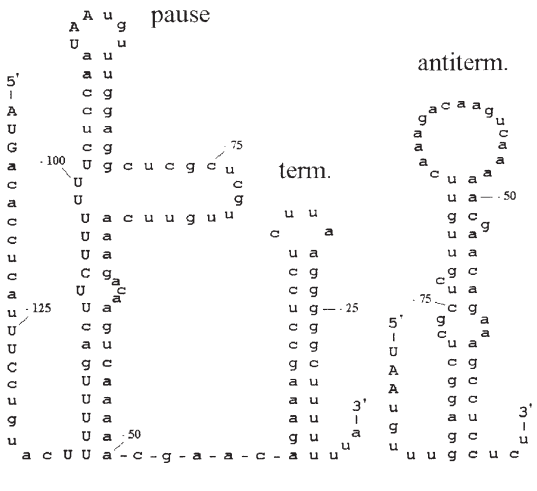
3B



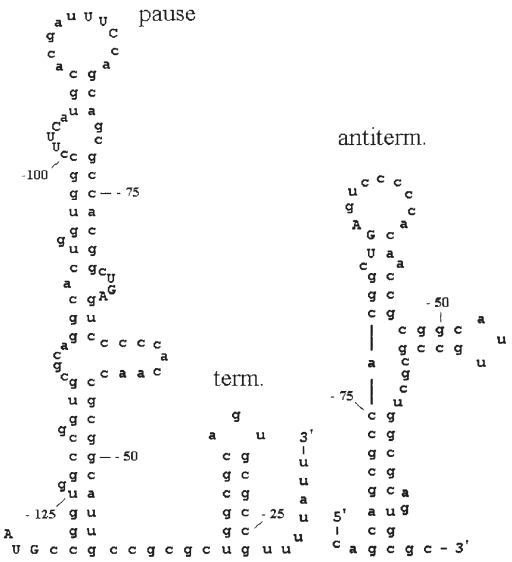


YP (distal structure)

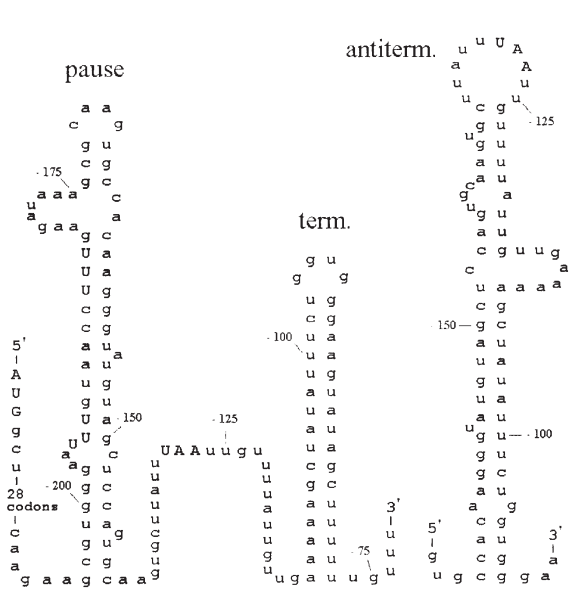
YP (proximal structure)



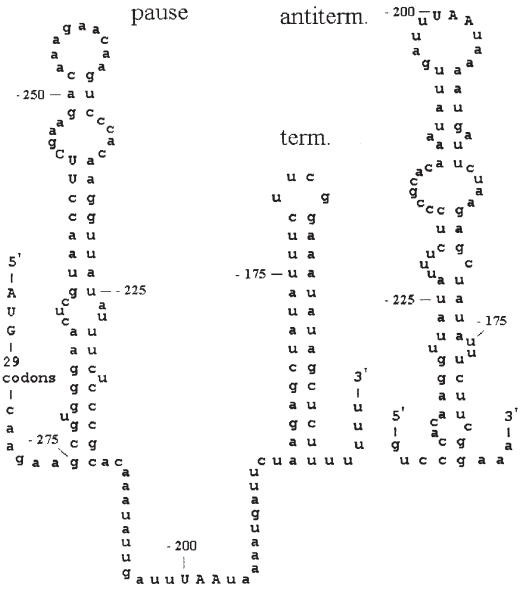
VC



Xanthomonas campestris



HI



AA

Table 1. TrpR and Tyr candidate binding sites and their scores. Boldface: TRP boxes. *Italic*: TYR boxes. Numbers: site scores.

	aroG		*aroFtyrA*		*aroH*	
EC	*agtgtaaaaccccgtttacaca*	4.52	*agtgtaaatttatctatacaga*	5.28	**atgtactagagaactagtgcat**	**5.60**
			tgtgtaaataaaaatgtacgaa	4.83		
			tatggattgaaaactttacttt	4.03		
ST	*agtgtaaaagcccgtttacact*	4.37	*tgtgtaaagtttttgatacgaa*	4.46	**atgtactagtaaactagtttaa**	**5.74**
			tatgtaaagtttatttttacgta	4.98		
			tatggattgaaatctttacttt	4.18		
YP	*agcgataaagtagatttacagc*	3.74	*tgtgtaaattattctttacgat*	5.03	**atgttttaactaaatattttca**	**2.73**
			atagtaattatttatttacaaa	4.59		
			tatggattgaaaactttacatc	4.53		
VC	*ggtgtaatttattatttacatc*	5.16	*ggtgtaatttttaatttacatt*	5.13	no site	
			aatgttggtttaaatttacact	3.82		
			tgtgtaaataataaattacagg	4.92		
HI	**tcgaactagtttactagtacaa**	6.20	no gene		no gene	
AA	**ataaactcttttgctagtacaa**	5.15	no gene		no gene	
SP	no site		no site		no gene	

	aroL		*tyrB*		*trpEDCBA*	
EC	**ttgtactagtttgatggtatga**	5.49	*cgtgtttcaaaaagttgacgaa*	3.53	**tcgaactagttaactagtacgc**	6.04
	aatgtaatttattatttacact	5.16				
	agtggaatttttctttacaat	4.69				
ST	**ttgtactagttagatgatatga**	5.10	no gene		**ttgaaccagttaactagtacga**	6.14
	catgtaatgaaaaatttacagt	4.80				
	tgtgtaaattattatttacatt	5.43				
YP	no site		*gatggaaagttttaaagccgat*	3.49	**gtgaaccagttaactagtacac**	5.67
VC	no gene		no gene		**ccgcactagttaactagtacgc**	5.77
HI	no gene		no gene		**ttgcactagtttaatagtacaa** (*trpEDC*)	6.03
					ttgtactactttaatagtacaa (*trpBA*)	5.82
AA	no gene		no gene		no site	
SP	no gene		no gene		no site	

	aroP		*tyrP*		*mtr*	
EC	*gatgtaaacaaattaatacaac*	4.65	*attgtacatttatatttacacc*	5.01	**ttgtactcgtgtactggtacag**	5.77
	aacggaattgcaaacttacaca	3.83	*tatgtaacgtcggtttgacgaa*	3.76	*tctgtaaaatatatatacagc*	5.01
ST	*cttgtaaataaatcaatacaaa*	4.51	*actgtaaaatttccagtacacc*	4.37	**ttgtactcgtgtactggtacag**	5.77
	agcggaattgcaaacttacaca	3.92	*attgtaacgaccatttgacgaa*	3.69	*agcgtaaagtaaaatatacagc*	4.82
YP	*atcgtaaataaagcaatacaac*	4.28	*atcgtaaactataatttacact*	4.79	**gtgtaccattcagctagtacaa**	5.33
	agtgtaattaaatgattacatt	5.00	*atcgtaacgacgatttgacaca*	4.00		
VC	no gene		*agtgtaaattattaattacagt*	5.13	no site	
			ttcggatttaaatttttacaaa	4.16		
			aatgtataaaagagtttacaca	4.84		
HI	no gene		*actgtaaattatacaatacaat*	4.55	**gtgtactactataatagtgcaa**	5.04
			atcgtaaatttttttattacatc	4.66		
AA	no gene		*tatgttattaaaaaataaaacc*	3.46	no gene	
SP	no gene		no site		no site	

	tyrR		*trpR*	
EC	*tgtgtcaatgattgttgacaga*	4.56	**tctgactctttagcgagtacaa**	5.98
ST	*gctgtcaatatttgttgacaga*	4.75	no gene	
YP	*agtgtcacccaatcttgacggc*	3.97	**aagtactatttaactagtacaa**	5.87
VC	*actgtaccattttcgtgacact*	4.31	no gene	
HI	*tatgtaaaataatatttacact*	5.38	**atgcactagtttaatagtgtaa**	5.17
AA	*actgtaaataaaagttgccaaa*	4.47	no gene	
SP	*gctgtaaactaggcttgccact*	4.18	no gene	

3), and no box in *V. cholerae*. Both the phylogenetic analysis (Figure 4) and the analysis of the feedback inhibition site (below) show that these genes encode proteins regulated by tryptophan. Analysis of the DNA alignment (Figure 3) shows that the TrpR regulation probably was lost in *Y. pestis* and *E. herbicola* due to point mutations in the TRP box. No homologs of *trpR* were found in the genome of *Buchnera* sp. The region upstream of *aroH* in *Buchnera* sp. cannot be aligned with the *aroH* upstream regions of other enterobacteria.

No regulatory sites could be observed upstream of *mtr* and *aroH* in *V. cholerae*. The regulatory interactions in *P. aeruginosa* genes are absolutely different and are considered separately (below).

There exist strong candidate TRP boxes upstream of both *trpEGDC* and *trpBA* operons of *H. influenzae*, in the latter case the operon contains a recently inserted gene *ydfG* (Mironov *et al.,* 1999). The site upstream of *trpEGDC* could be homologous to the sites upstream of *trpEGDCBA*, whereas the site upstream of *ydfGtrpBA* should have arisen *de novo*. In both cases the candidate sites are downstream of the predicted leader peptides (see the next section) and

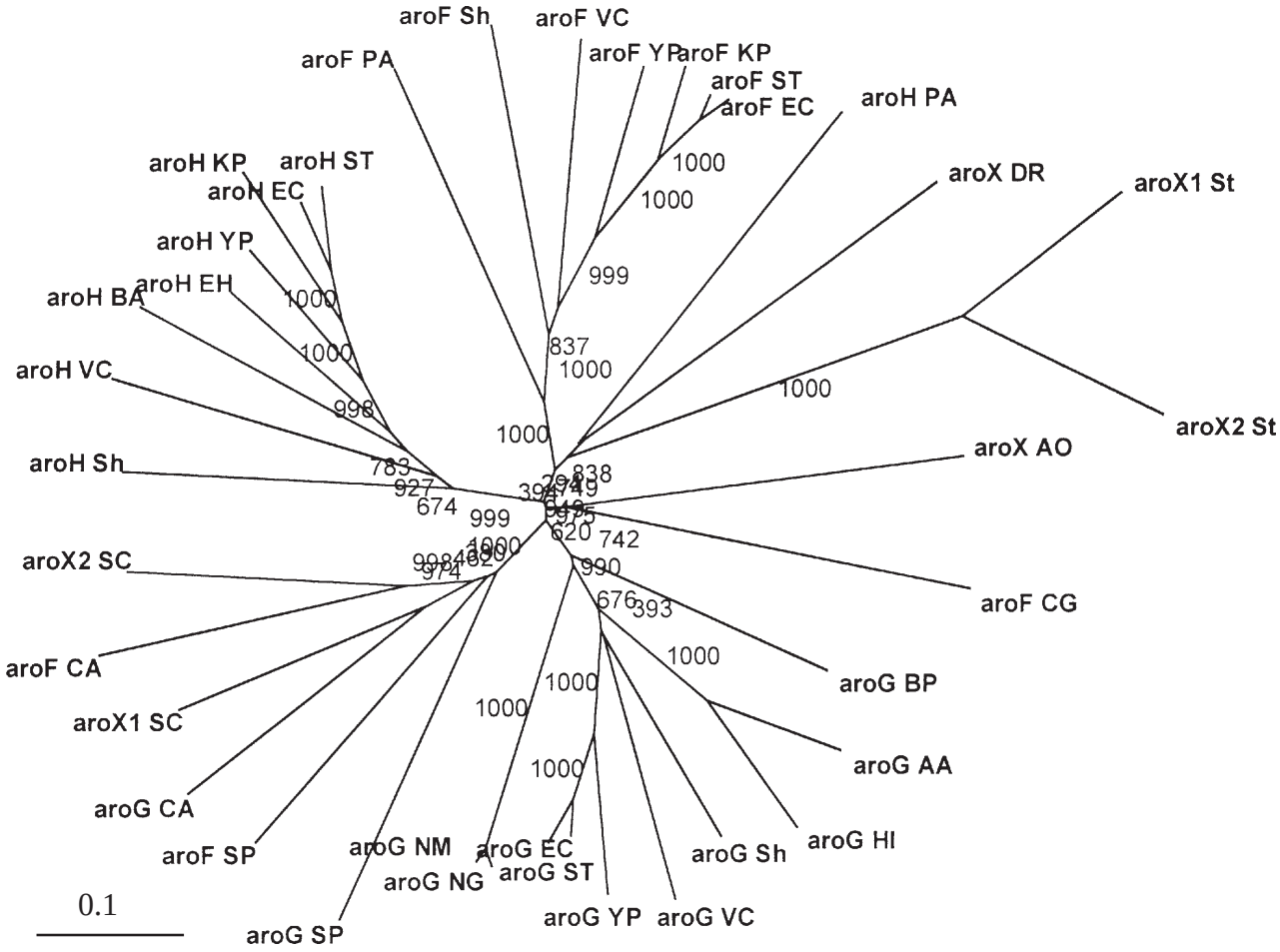


Figure 4. Evolutionary tree of DAHP-synthases. Numbers at internal nodes: bootstrap values. Additional notation: AM: *Amycolaptosis methanolica*, AO: *Amycolatopsis orientalis*, BA: *Buchnera aphidicola*, BP: *Bordetella pertussis*, CA: *Candida albicans*, CG: *Corynebacterium glutamicum*, DR: *Deinococcus radiodurans*, EH: *Erwinia herbicola*, NG: *Neisseria gonorrhoeae*, NM: *Neisseria meningitidis*, PA: *Pseudomonas aeruginosa*, SC: *Saccharomyces cerevisiae*, Sh: *Shewanella putrefaciens*, SP: *Schizosaccharomyces pombe*, St: *Streptococcus pneumoniae*. Numbers on internal nodes (bootstrap values fot 1000 replications).

thus within the transcribed region, which is unusual. There is a strong candidate TRP box upstream of the *trpBA* operon of *Pasteurella multodica*. Note that *P. multocida* is a close relative of *H. influenza* and thus the insertion of *ydfG* in the latter occured after divergence of these two genomes.

The unique DAHP-synthases from *H. influenzae* and *A. actinomycetemcomitans* belong to the AroG lineage (Figure 4) and have typical phenylalanine inhibition sites (below). However, in both genomes the *aroG* genes are preceded by strong TRP boxes, but have no candidate TYR boxes (Mironov *et al.,* 1999).

Control by Attenuation

Attenuators regulate two aromatic amino acid biosynthetic operons, *trp* and *phe*, the operon *pheST* encoding subunits of the phenylalanyl-tRNA synthetase, and the catabolic operon *tnaAB* (Landick and Yanofsky, 1996; Keller and Calvo, 1979; Grunberg-Manago, 1996). In the first three cases attenuation involves formation of alternative RNA hairpins dependent on the relative rates of transcription and translation. The rate of translation of the leader peptide, in turn, depends on the concentration of aminoacylated tRNA and thus on the availability of the amino acid. The latter operon, *tnaAB*, is regulated via rho-dependent attenuation.

We have applied the comparative methods to predict attenuators of the *trp, phe, pheST*, *tnaAB* operons in *S. typhi, Y. pestis, V. cholerae, H. influenzae, A. actinomycetemcomitans, X. campestris*, as well as bacterium from a different taxonomic group, *Chlamydia trachomatis*. There are no candidate attenuators in *Buchnera* (Shigenomu *et al.,* 2000).

Attenuators of the *trp* and *phe* operons are described in this section, whereas the data for *pheST* and *tnaAB* operons are not shown, since no novel observations could be made. The general properties of the candidate attenuators are summarized in Table 2.

Operon *trp*

Attenuators of this operon in *S. typhi, Y. pestis, V. cholerae* are similar to the *E. coli* attenuator despite divergence of

Table 2. General properties of attenuators of aromatic amino acid operons. (a) Lengths of the structure elements. Cells: Length of the leader peptide in codons / Number of regulatory codons / Distance between the stop of the leader peptide and the antiterminator helix / Distance between the last regulatory codon and the antiterminator helix. * - the last regulatory codons are excessive and the distance is given for the codon shown in parentheses. (b) Free energy of the attenuator hairpins. Hairpins: ps – pause, term – terminator, ant – antiterminator. *phe* operon of *Y. pestis*: first line – distal attenuator, second – proximal attenuator. *H. influenzae*: first line – attenuator of the operon *trpEGDC*, second – attenuator of the operon *ydfGtrpBA* operon.

a:

	trpEDCBA	*pheA*	*pheST*
E. coli	45/2/2/14	48/6/1/13	45/5/5/14
S. typhi	45/2/2/14	48/7/3/14*(6)	45/5/5/14
Y. pestis	63/2/overlap/15	51/7/3/15*(6)	45/5/5/14
V. cholerae	111/5/overlap/11*(3)	48/6/5/14	
H. influenzae	*trpEGDC*: 57/1/22(?)/32 *trpBA*: 63/2/overlap/17	174/2/overlap/23	78/3/overlap/13*(2)
A. actinomycetemcomitans	not sequenced?	174/1/overlap/23	not sequenced?
X. campestris	not sequenced?	66/2/overlap/20*(1)	not sequenced?
C. trachomatis	*trpBA*: 114/2/7/13	none	none

b:

operons	*trp* operon			*phe* operon			*pheST* operon		
hairpins	ps	term	ant	ps	term	ant	ps	term	ant
E.coli	-18.6	-11.4	-13.5	-25.5	-16.1	-21.2	-32.5	-10.0	-16.4
S. typhi	-10.0	-9.1	-15.9	-25.5	-16.4	-19.6	-23.8	-12.7	-19.6
Y. pestis	-13.2	-8.8	-13.9	-25.7	-10.7	-11.7	-26.8	-10.7	-12.5
				-13.3	-14.8	-12.8			
V. cholerae	-10.7	-11.4	-19.4	-14.8	-12.1	-11.7	-	-	-
H. influenzae	-4.0	-7.3	-4.8	-10.1	-12.1	-14.4	-22.2	-5.2	-6.0
	-5.8	-5.8	-7.5						
A. actinomycetemcomitans	-	-	-	-17.8	-13.1	-5.0	-	-	-
X. campestris	-	-	-	-12.6	-5.0	-16.0	-	-	-
C. trachomatis	-10.7	-12.0	-11.6	-	-	-	-	-	-

the sequence (Figure 2a). The *trp* operon of *H. influenzae* is split into two parts (see above). Both of them have potential attenuators in the upstream region (Figure 2a). However, the candidate TRP box occurs after the attenuator (Table 3). This is somewhat unusual, although not exceptional: other examples of transcription respressor sites within transcripts are known, in particular, PUR box within *purB* gene of *E. coli* (He and Zalkin, 1992) or CRE box within *gntR* gene of *B. subtilis* (Miwa and Fujita, 1990). The high scores of the candidate TRP boxes and the existence of alternative RNA structures in both operons make it unlikely that these predictions are false positives. This raises the question of the mechanism of regulation. Neither a leader peptide, not candidate attenuator were found upstream of the *trpBA* operon of *Pasteurella multodica*.

We have also predicted the attenuator of the *trpBA* operon of *C. trachomatis* (Figure 2a). Its structure is similar to that of enteric bacteria, suggesting the same mode of regulation.

Table 3. Attenuator and TRP box positions upstream of the tryptophan operons. Begining of the attenuator is defined as the start codon; end of the attenuator, as the position of the polyT tract.

	Attenuator position	TRP box position
EC	-135…-25	-183
ST	-139…-25	-176
YP	-177…-55	-285
VC	-244…-90	-296
HI (*trpEGDC*)	-161…-25	-47
HI (*ydfGtrpBA*)	-351…-275	-48

Operon *pheA*

In *Enterobacteria*, *V. cholerae* and *Xanthomonas campestris*, the distance between the pause stem-loop and the terminator is several nucleotides (Figure 2b), whereas in *Pastereullaceae* (*H. influenzae* and *A. actinomycetemcomitans*) it exceeds 25 nucleotides (Figure 2b). However, conservation of the structure between the latter genomes makes us confident in the prediction.

In *Y. pestis*, the *phe* attenuator is disrupted by insertion of IS200 element so that the transposase gene of IS200 is on the complementary strand. The terminal stem-loops of IS200 (Mahillon and Chandler, 1998) substitute for the disrupted elements of the attenuator retaining the alternative structure at both parts (Figure 2b). The distal candidate attenuator contains all necessary structural elements, but it is very distant from the *pheA* gene, so that transcription from the original promoter regulated by the distal attenuator generates an untranslated leader of more than 800 nucleotides. The proximal structure contains all RNA elements, but no candidate leader peptide, as it completely lacks open reading frames. Prior to the IS200 insertion, the *Y. pestis* attenuator has been similar to the *E. coli* one with several nucleotide mismatches.

Allosteric Control

Escherichia coli and related bacteria have three DAHP-

```
                  ****              **********
AroH_EC      LDMVTGQFIAD       HMFLSPDKDGQMTIYQT
AroH_YP      LNMVTGQYIAD       HMFLSPDKTGQMTIYQT
AroH_EH      LDMVIGQFIAD       HMFLSPDKLGQMTIYQT
AroH_BA      LDMVIGQFIAD       HLFFAPNKDGQMTINHT
AroH_VC      LDMITGQYIAD       HYFYSPDKNGRMTVYRT
AroH_SP      LDMVNGQYIAD       HIFYSPDKDGAMSVYRT

AroG_EC      LDMITPQYLAD       HCFLSVTKWGHSAIVNT
AroG_ST      LDMITPQYLAD       HCFLSVTKWGHSAIVNT
AroG_YP      LDMITPQYLAD       HCFLSVTKWGHSAIVNT
AroG_VC      LDMITPQYVAD       HHFLSVTKFGHSAIVET
AroG_SP      LDMITPQYVAD       HHFLSVTKFGHSAIVST
AroG_AA      LDMITPQYLAD       HHFLSVTKFGHSAIVST
AroG_HI      LDMITPQYLAD       HYFLSVTKFGHSAIVST
AroG_NG      LDMITPQYYAD       HHFLSVTKAGHSAIVHT
AroG_NM      LDMITPQYYAD       HHFLSVTKAGHSAIVHT
AroG_BP      LDMITPQYIAD       HHFLSVTKGGHSAIVST

AroF_EC      LDPNSPQYLGD       HRFVGINQAGQVALLQT
AroF_ST      LDPNSPQYLGD       HRFVGINQAGQVALLQT
AroF_YP      LDPNSPQYLGD       HRFMGINQSGQVCLLQT
AroF_VC      LDPISPQYLAD       HRFMGINREGQVALLTT
AroF_SP      LDPISPQYISE       HRFMGINQQGQVALLQT
AroF_PA      LDPISPQYLQD       HRFLGINQQGGVSIVTT
AroG_CG      LEPNSPQYYAD       HFFFGTSDDGALSVVET

AroG_PA      LQPLAAGYFDD       HRHFGLDPHGHPALIET
AroF_DR      LDPFAPQYLFD       HAFFTIDEDGRAAIVHT
```

Figure 5. Alignment of the feedback inhibition site in bacterial DAHP-synthases. Notation: see the legend to Figure 4.

synthase isoenzymes AroF, AroG and AroH feedback inhibited by tyrosine, phenylalanin and tryptophan respectively (Camakaris and Pittard, 1974; Doy, 1967) It has been suggested that appearance of the most recently diverged DAHP-synthase-PHE (aroG) coincided with the enteric lineage (Ahmed *et al.,* 1988). However, there are genes encoding all three izoenzymes in the genomes of *V. cholerae* and *S. putrefaciens*, and thus the duplication leading to AroG occured before divergence of *Enterobacteriaceae, Vibrionaceae* and *Alteromonadaceae*. On the other hand, *Buchnera* sp. has only one DAHP-synthase (Kolibachuk *et al.,* 1995; Shigenobu *et al.,* 2000) belonging to the enteric AroH subfamily (Figure 1).

Experiments have shown that mutations of Val and Gly at positions 147, 149 in AroH and Pro at position 148 in AroF are sufficient to lift feedback inhibition by tryptophan and tyrosine respectively (Ray *et al.,* 1988; Weaver and Herrmann, 1990). Figure 5 shows alignment of this site together with a second, spatially close region. Together these two regions form the amino acid binding pocket.

The analysis of patterns in both feedback inhibition sites of the unique DAHP-synthase in *H. influenzae* and *A. actinomycetemcomitans* confirms that these enzymes belong to Phe-dependent family (Figure 5). Thus the unexpected TrpR binding sites in their upstream regions probably indicate the change of regulatory system.

The second site of AroH in *Buchnera* lost some residues absolutely conserved in other tryptophan-inhibited DAPH-synthetases. This can be explained by selection for tryptophan production in these bacteria, causing possible loss of feedback inhibition (cf. above).

The experimental evidence about inhibition of DAHP-synthases in *Pseudomonas aeruginosa* is contradictory. One of them (AroF) is probably inhibited by tyrosine, whereas the other, annotated as AroG, has been claimed to be tryptophan-dependent (Whitaker *et al.,* 1982) or phenyalanine-dependent (Maksimova *et al*, 1991). Analysis of the feedback inhibition site or the phylogenetic tree cannot resolve this problem.

Finally, we have considered DAHP-synthases from

Table 4 . PhhR candidate binding sites in *P. aeruginosa*. Positions are given relative to the start of translation.

operon	sites	score	position
phhABC	*tccgtcaagaatatgtgacagt*	4.24	-81
	ttcgtaaggaaaactttacgaa	3.95	-8
ppdAaroP	*actgtaaagataactttacgaa*	5.01	-214

other bacteria. Analysis of the specific patterns allows us to predict phenylalanine inhibition for the DAHP-synthase from *Bordetella pertussis*. The first site in the DAHP-synthase from *Corynebacterium glutamicum* coincides with the pattern in tyrosine-inhibited isoenzymes from gamma-proteobacteria, although the second site is less conserved. However, this fact and position of this protein in the phylogenetic tree makes it likely that this is AroF rather than AroG, as annotated in GenBank.

Transcriptional Regulation in *Pseudomonas aeruginosa*

P. aeruginosa has no orthologue of TrpR. Instead, the *trpBA* operon is regulated by TrpI, a LysR-type activator whose gene is transcribed divergently (Auerbach *et al.,* 1993). The binding sites of TrpI in the *trpI-trpBA* intergenic region have been completely described in (Olekhnovich and Gussin, 1998). No additional sites could be found upstream of other tryptohan biosynthesis genes.

The orthologue of TyrR in the genome of *P. aeruginosa* is known as PhhR. It was shown to activate transcription of the adjacent divergently transcribed operon *phhABC* that encodes enzymes of the L-phenilalanine catabolism (Song and Jensen, 1996). PhhR binds two sites upstream of *phhA* that are similar to the TYR boxes of *E. coli*. Moreover, if transferred to *E. coli*, PhhR can substitute TyrR as a repressor of the *aroFtyrA* operon, although it appears to be uncapable of replacing TyrR as an activator of *mtr* (Song and Jensen, 1996).

Using the TYR box profile, we have identified another likely member of the PhhR regulon. This operon consists of two genes. The upstream gene encodes a homologue of the p-hydroxyphenylpyruvate dioxygenase (*ppdA*) that transforms L-tyrosine into p-hydroxyphenylpyruvate at the first step of the L-tyrosine catabolism pathway. The second gene is an orthologue of *aroP* and thus is likely to encode a general aromatic amino acids transporter (Table 4).

Conclusion

This study demonstrates that evolution of interactions in the aromatic amino acids regulon is somewhat more complicated that it was thought before. Indeed, identification of all three DAPH-synthases in *V. cholerae* and *S. putrefaciens* demonstrates that these isoenzymes diverged prior to filiation of *Enterobacteriaceae*. On the other hand, *Buchnera* sp., an early branching member of *Enterobacteriaceae*, has lost AroF and AroG.

Another example of regulatory evolution is the change of regulation of the *aroG* gene in *Pasteurellaceae*. Less striking, but important from the biological point of view, is the apparent loss of many regulatory sites, both in DNA and in proteins. Several transcription factor binding sites in the *Y. pestis* genome are deteriorated and probably not functional; transcriptional regulation in *V. cholerae* seems to be much less extensive than in *E. coli* and other *Enterobacteriaceae*.

It has been noted that the genes encoding the aromatic amino acid metabolism enzymes in *Buchnera* sp. probably are not subject to control on the level of transcription, both by repression and attenuation (Shigenobu *et al.,* 2000). We report deterioration of the allosteric control site in the DAPH-synthase. Thus the aromatic amino acid biosynthesis in these aphid endosymbionts seems to be free from any known form of control, probably reflecting selection towards maximum efficiency of the essential amino acid production.

Transcriptional regulation of the aromatic amino acid metabolism in *Pseudomonas aeruginosa* is different from that in other gamma purple bacteria. Instead of tryptophan repressor TrpR, it has activator TrpI, whereas PhhR, the orthologue of TyrR, regulates catabolic, rather than anabolic genes. We have identified an additional candidate PhhR binding site upstream of an operon encoding a homologue of p-hydroxyhenylpyruvate dioxygenase (*ppdA*), the first enzyme of the L-tyrosine catabolism pathway, and a candidate aromatic amino acid transporter.

Many results reported here are tentative and require experimental verification. In particular, the importance of the second part of the candidate allosteric control site should be demonstrated by direct analysis, e.g. site-specific mutation studies. The consequences of IS200 insertion in the *pheA* attenuator in *Y. pestis* cannot be predicted by purely computational approaches. The predicted transcription factor binding sites look relevant based on comparative and functional reasoning, but still have to be checked in an experiment.

Of particular interest are the interactions between different mechanisms of regulation, in particular, transcriptional regulation and attenuation. In *H. influenzae*, the candidate TRP box is located downstream of the predicted attenuator in both *ydgFtrpBA* and *trpEGDC* operons. No attenuator could be predicted upstream of the *trpBA* operon in *Pasteurella multocida*, although there is a strong candidate TRP box. Conversely, a perfect attenuctor can be formed upstream of the *trpBA* operon of *Chlamidia trachomatis*. It is likely that this operon has been horizontally transferred from some gamma-proteobacterium (Stephens, R.S. *et al.,* 1998). Indeed, it has no homologue in the genome of a closely related *Chlamidia pneumoniae* (Kalman S. *et al.,* 1999), whereas in *C. trachomatis* it is located immediately downstream of the repressor gene *trpR*. Gamma-proteobacteria is the only taxonomic group known to contain the *trpR* gene. On the other hand, in this case we could not identify a TrpR binding site.

In this study, as well as in analysis of other regulatory systems of gamma-proteobacteria, in particular, purine, arginine, heat shock, damage repair, and sugar metabolism regulons (Mironov *et al.,* 1999; Gelfand, 1999, and unpublished observations) we have noted that the systems of *E. coli* are the most complicated. Many genes transcriptionally regulated in *E. coli* have no candidate binding sites for respective factors in other genomes. This

could reflect the variability of the environments inhabited by this bacterium and flexibility of its responses to the changing conditions. However, this also could be an artifact of the approach. Indeed, the comparative technique allows one to find only the conserved cores of regulons, whereas the genome-specific sites cannot be identified. The *E. coli* genome is studied in detail, and thus many genome-specific sites have been found experimentally, whereas specific sites in other genomes have not been discovered, not can they be identified by the comparative computational analysis. As more genomes are sequenced, the probability of finding conserved sites in a subset of newly sequenced genomes increases.

Overall, although this study seems to raise more questions than answers, we feel that most predictions made here are true. The remaining uncertainties in evolution of regulatory interactions in this part of methabolism will be resolved as more data becomes available. On the other hand, the obtained predictions should aid in functional studies, allowing one to set up specific experiments directed towards resolution of the remaining uncertainties and ambiguities.

Data and Methods

The complete genome sequences of *Escherichia coli* (EC), *Haemophilus influnzae* (HI), *Buchnera* sp. APS (BA) and *Pseudomonas aeruginosa* (PA) were extracted from GenBank (Benson *et al.,* 1999). The partially sequenced genomes of *Salmonella typhimurium* (ST), *Yersinia pestis* (YP), *Vibrio cholerae* (VC), *Actinobacillus actinomycetemcomitans* (AA), *Shewanella putrefaciens* (SP) were obtained from the TIGR WWW site (http://www.tigr.org). In several cases additional analyses were applied to sequence fragments extracted from GenBank (Benson *et al.,* 1999).

Profiles for recognition of TrpR and TyrR binding sites (resp., TRP and TYR boxes) were taken from (Mironov *et al.,* 1999). Positional nucleotide weights in these profiles are defined as:

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

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

$$Z(b_1 \ldots b_L) = \Sigma_{k=1\ldots L}\, W(b_k,k),$$

where k is the length of the site.

Multiple protein alignments and phylogenetic trees were constructed using CLUSTAL (Thompson *et al.,* 1997) and plotted with GeneTree (Page, 1998). Analysis of the protein 3D structure was done using RASMOL (http://www.umass.edu/microbio/rasmol/). Genomic analyses (protein similarity search, analysis of orthology, DNA profile search) were done using GenomeExplorer (Mironov *et al.,* 2000).

Candidate attenuators were predicted by manual folding of gene upstream regions with the *E. coli* attenuators as templates. The free energy of the secondary structures was estimated using the rules from (Freier *et al.,* 1986).

Acknowledgements

We are grateful to Eugene Koonin, Yury Kozlov, and Alexandra Rakhmaninova for useful discussions. This study was partially supported by grants from the Merck Genome Research Institute (244), the Russian Fund of Basic Research (99-04-48247 and 00-15-99362), the Russian State Scientific Program “Human Genome”, INTAS (99-1476), and the Howard Hughes Medical Institute (55000309).

References

Ahmad, S., Rightmire, B., Jensen, R.A. 1986. Evolution of the regulatory isozymes of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase present in the *Escherichia coli* genealogy. J. Bacteriol. 165: 146-154.

Auerbach, S., Gao, J., Gussin, G.N. 1993. Nucleotide sequences of the *trpI*, *trpB*, and *trpA* genes of *Pseudomonas syringae*: positive control unique to fluorescent pseudomonads. Gene 123: 25-32.

Bennett, S.N. and Yanofsky, C., 1978. Sequence analysis of operator constitutive mutants of the tryptophan operon of *Escherichia coli*. *J. Mol. Biol.* 121: 179-192.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouelette B.F., Rapp, B.A. and Wheeler, D.L. 1999. GenBank. Nucleic Acids Res. 27: 12-17.

Brown, K.D. and Doy, C.H. 1966. Control of three isoenzymic 7-phospho-2-oxo-3-deoxy-Darabino-heptonate-D-erythrose-4-phosphate lyases of *Escherichia coli* W and derived mutants by repressive and “inductive” effects of the aromatic amino acids. Biochim. Biophys. Acta 118: 157-172.

Camakaris, J. and Pittard, J. 1974. Purification and properties of 3-deoxy-D-arabinoheptulosonic acid-7-phosphate synthetase (trp) from *Escherichia coli.* J. Bacteriol. 120: 590-597.

Dandekar, T., Schuster, S., Snel, B., Huynen, M., Bork, P. 1999. Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem J. 343:115-24.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., Turner, D.H. 1986. Improved free-energy parameters for prediction of RNA duplex stability. Proc Natl Acad Sci USA 83: 9373-9377.

Fickett, J.W. and Hatzigeorgiou, A.G. 1997. Eukaryotic promoter recognition. *Genome Res.* 7: 861-878.

Forst, C.V. and Schulten K. 1999. Evolution of metabolisms: a new method for the comparison of metabolic pathways using genomics information. J. Comput. Biol. 6: 343-360.

Frech, K.,Danescu-Mayer, J., Werner, T. 1997. Software for the analysis of DNA sequence elements of transcription. Comput. Appl. Biosci. 13: 89-97.

Galperin, M.Y. and Koonin, E.V. 1999 Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. Genetica 106:159-70

Gelfand, M.S., Koonin, E.V., Mironov, A.A. 2000 Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. Nucleic Acids Res. 28: 695-705.

Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I., Perumov, D.A. 1999. A conserved RNA structure element

involved in the regulation of bacterial riboflavin synthesis genes. Trends Genet. 15: 439-442.

Grunberg-Manago, M. 1996. Regulation of the expression of aminoacyl-tRNA synthetases and translation factors. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology. Neidhardt F.C. (ed.), ASM Press, Washington DC, pp. 1432-1457.

He B, Zalkin H. 1992. Repression of *Escherichia coli purB* is by a transcriptional roadblock mechanism. J. Bacteriol. 174: 7121-7127.

Heatwole, V.M., and Somerville, R.L. 1992. Synergism between the Trp repressor and Tyr repressor in repression of the *aroL* promoter of *Escherichia coli* K-12. J. Bacteriol. 174: 331-335.

Jackson, E.N. and Yanofsky C. 1973. The region between the operator and first structural gene of the tryptophan operon of *Escherichia coli* may have a regulatory function. J. Mol. Biol. 76:89-101

Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger L., Grimwood, J., Davis, R.W., Stephens, R.S. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nature Genetics 21: 385-389.

Keller, E.B. and Calvo, J. 1979 Alternative secondary structures of leader RNAs and the regulation of the *trp*, *phe*, *his*, *thr*, and *leu* operons. Proc. Natl. Acad. Sci. USA 76: 6186-6190.

Kolibachuk, D., Rouhbakhsh, D., Baumann, P. 1995. Aromatic amino acid biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids): cloning and sequencing of a DNA fragment containing *aroH-thrS-infC-rpmI-rplT*. Curr Microbiol. 30: 313-316.

Kreneva, R.A., Gelfand, M.S., Mironov, A.A., Jomantas, J.A., Kozlov, Y.I., Mironov, A.S., Perumov, D.A. 2000. Phenotype of *Bacillus subtilis* with inactivated *ypaA*. Genetika 36: 1166-1168.

Landick, R., Charles, L., Turnbough, J. R., Yanofsky, C. 1996. Transcription attenuation. In: *Escherichia coli* and *Salmonella.* Cellular and Molecular Biology. Neidhardt FC (eds), ASM Press, Washington DC, pp.1263-1286.

Lawley, B. and Pittard, A.J. 1994. Regulation of *aroL* expression by TyrR protein and Trp repressor in Escherichia coli K-12. J. Bacteriol. 176: 6921-6930.

Mahillon, J., Chandler, M. 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. 62: 725-774.

Maksimova, N.P., Olekhnovich, I.N. and Fomichev, Iu.K. 1991 Regulation of the synthesis of 3-deoxy-D-arabinoheptulosonate-7-phosphate-synthase in Pseudomonas bacteria. Genetika 27: 217-221.

McGuire, A.M., Hughes, J.D. and Church, G.M. 2000. Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. Genome Res. 10: 744-757.

Mironov, A.A., Koonin, E.V., Roytberg, M.A. and Gelfand, M.S. 1999. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. Nucleic Acids Res. 27: 2981-2989.

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

Miwa, Y., Fujita, Y. 1990. Determination of the cis sequence involved in catabolite repression of the *Bacillus subtilis gnt* operon; implication of a consensus sequence in catabolite repression in the genus *Bacillus*. Nucleic Acids Res. 18: 7049-7053.

Olekhnovich, I. and Gussin, G.N. 1998. Recognition of binding sites I and II by the TrpI activator protein of *Pseudomonas aeruginosa*: efficient binding to both sites requires InGP even when site II is replaced by site I. Gene 223: 247-255.

Otwinowski, Z. *et al*. 1988. Crystal structure of the *trp* repressor/operator complex at atomic resolution. Nature 335: 321-329.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., Maltsev, N. 1999. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* 96: 2896-2901.

Page, R.D. 1998. GeneTree: comparing gene and species phylogenies using reconciled trees. Bioinformatics 14: 819-820.

Pittard, A.J. 1996. Biosynthesis of aromatic amino acids. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology. Neidhardt F.C. (ed.), ASM Press, Washington DC, pp. 458-484.

Pittard, A.J. and Davidson, B.E. 1991. TyrR protein of Escherichia coli and its role as repressor and activator. Mol. Microbiol. 5: 1585-1592.

Ramirez-Santos, J., Collado-Vides, J., Garcia-Varela, M. and Gomez-Eichelmann, M.C. 2001. Conserved regulatory elements of the promoter sequence of the gene rpoH of enteric bacteria. Nucleic Acids Res. 29: 380-386.

Ray, J.M., Yanofsky, C. and Bauerle, R. 1988. Mutational analysis of the catalytic and feedback sites of the tryptophan sensitive 3-deoxy-D-arabinoheptulosonate-7-phosphate of *Escherichia coli.* J. Bacteriol. 170: 5500-5506.

Rodionov, D.A., Mironov, A.A., Rakhmaninova, A.B. and Gelfand M.S. 2000. Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. Mol. Microbiol. 38: 673-683.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H. 2000. Genome sequencing of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407: 81-86.

Song, L., Jensen, R.A. 1996. PhhR, a divegently transcribed activator of the phenylalanine hydroxylase gene cluster of *Pseudomonas aeruginosa*. Mol. Microbiol. 22: 497-507.

Stephens, R.S., Kalman, S., Lammel, C., Fan, J. Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V., Davis, R.W. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282: 754-759.

Stojanovic, N., Florea, L., Riemer, C., Gumuchio, D., Slightom, J., Goodman, M., Miller, W. and Hardison, R. 1999. Comparison of five methods for finding conserved sequences in multiple alignments of gene regulatory regions. Nucleic Acids Res. 27: 3899-3910.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F and Higgons., 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.
Weaver, L. M. and Herrmann, K.M. 1990. Cloning of an *aroF* allele encoding a tyrosine-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. J. Bacteriol. 172: 6581-6584.
Whitaker, R.J., Fiske, M.J. and Jensen, R.A. 1982. *Pseudomonas aeruginosa* possesses two novel regulatory isozymes of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. J. Biol. Chem. 257: 12789-12794.
Yanofsky C. 1981. Attenuation in the control of expression of bacterial operons. Nature 289: 751-758.
Yanofsky, C., Platt, T., Crawford, I.P., Nichols, B.P., Christie, G.E., Horowitz, H., VanCleemput, M., Wu, A.M. 1981. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli.* Nucleic Acids Res. 9: 6647-6668.