MICROBIAL & COMPARATIVE GENOMICS
Volume 4, Number 1, 1999
Mary Ann Liebert, Inc.

Mixed-Function Supraoperons That Exhibit Overall Conservation, Albeit Shuffled Gene Organization, across Wide Intergenomic Distances within Eubacteria

GARY XIE,[1] THOMAS S. BRETTIN,[1] CAROL A. BONNER, and ROY A. JENSEN

ABSTRACT

Nearly identical mixed-function supraoperons (defined as nested transcriptional units encoding gene products that function in more than one biochemical pathway) have been found recently in *Pseudomonas stutzeri* and *Pseudomonas aeruginosa*. The *Pseudomonas serC(pdxF)-aroQ$_p$•pheA-hisH$_b$-tyrA$_c$-aroF-cmk-rpsA* supraoperon encodes 3-phosphoserine aminotransferase, a bidomain chorismate mutase/prephenate dehydratase, imidazole acetol-phosphate aminotransferase, cyclohexadienyl dehydrogenase, 5-enolpyruvylshikimate 3-phosphate synthase, cytidylate kinase, and 30S ribosomal protein S1. These enzymes participate in the biosynthesis of serine, pyridoxine, histidine, phenylalanine, tyrosine, tryptophan, and aromatic pathway vitamins and cytidylic acid, in addition to the general role of RpsA in the process of protein synthesis. Features that suggest supraoperon-wide translational coupling are the highly compressed intergenic spacing (including overlapping stop and start codons), as well as possible hairpin structures in mRNA, which could sequester many of the ribosome-binding sites. The *hisH-tyrA-aroF* segment corresponds to the distal genes of the classic *Bacillus subtilis* supraoperon. Extensive comparative analysis of the member genes of both the *Bacillus* and *Pseudomonas* supraoperons from organisms represented in the entire database revealed unmistakable organizational conservation of these genes across wide phylogenetic boundaries, although considerable gene shuffling was apparent. The persistence of *aroE-aroB, hisH$_b$-tyrA-aroF*, and *cmk-rpsA* throughout both the gram-negative and gram-positive assemblages of bacteria, but the absence in Archaea, suggests an ancestral gene organization that occurred in bacteria after the separation of the bacterial and archaeal domains. In gram-negative bacteria, the *hisH$_b$-tyrA$_c$-aroF* grouping may have been expanded (as with the *Pseudomonas* supraoperon) and then subsequently collapsed (as with the *Escherichia serC-aroF* supraoperon) via gene shuffling that is herein equated with gene fusion events.

Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida.
[1]*Present address*: Los Alamos National Laboratories, Los Alamos, New Mexico.

INTRODUCTION

Operon organization of genes

These modern times of whole-genome sequencing provide an unprecedented opportunity to evaluate gene organization. To what extent does a given cluster of genes remain linked together across what phylogenetic distance? What relationships exist between genes that might account for bringing them together and keeping them together?

Escherichia coli has provided the classic experimental system where multicistronic transcriptional units, such as the *lac* operon, the *trp* operon, and the *his* operon, were first elucidated. It initially was widely presumed that most genes specifying enzymes belonging to a particular biochemical pathway would be organized into operons dictating the formation of a single polycistronic mRNA. Such transcriptional units could be upregulated and downregulated via the interaction of *cis*-acting (e.g., operators) and *trans*-acting (e.g., activator or repressor proteins) regulatory elements, thus providing a mechanism to coordinate an appropriate rise and fall of all the enzymatic machinery cognate to a specific pathway.

It eventually became apparent that this attractive picture was an oversimplification. For example, in contrast to expectations raised by the elucidation of the *his* and *trp* operons in *E. coli*, the genes specifying the eight-step arginine biosynthesis pathway are scattered along the chromosome. Some of the classic operons, for example, the histidine operon, have subsequently proven to exhibit unanticipated features of complexity, such as internal promoters (Winkler, 1996), and in many cases, an unexpected admixture of genes of unknown function (Alifano et al., 1996). Furthermore, biochemical pathway gene organization in even fairly closely related organisms did not necessarily prove to be highly conserved. For example, *trp* pathway genes in *Pseudomonas aeruginosa* are scattered into three widely spaced groups rather than coexisting within one operon, as they are in *E. coli* (Crawford, 1989).

Mixed-function supraoperons

Even more surprising than the finding that genes directly related to a specific biochemical pathway were not always joined to a common promoter was the recognition that genes specifying apparently unrelated gene products sometimes coexisted within common transcriptional units. These have been variously termed mixed-function operons (Duncan and Coggins, 1986), complex operons (Tsui et al., 1994a), supraoperons (Tsui et al., 1994b), multifunctional operons (Man et al., 1997), or supraoperons (Henner and Yanofsky, 1993; Xie et al., 1999). The following nomenclature is suggested. Simple operons correspond to single transcriptional units encoding single pathway gene products. Mixed-function operons denote single transcriptional units encoding multiple pathway gene products. Supraoperons denote nested transcriptional units encoding single pathway gene products. Mixed-function supraoperons denote nested transcriptional units encoding multiple pathway gene products.

In one case, *Bacillus subtilis* has positioned a classic *trp* operon inside a larger transcriptional unit (Henner and Yanofsky, 1993). An upstream *aroGBH* operon overlaps the *trp* promoter, the *aroH* stop codon is within the *aroGBH* terminator, and the *aroGBH* terminator is synonymous with the *trp* attenuator. A third operon *($hisH_b$-$tyrA_p$-aroF)* downstream overlaps the *trp* operon, and readthrough *trp* transcripts can proceed through the *$hisH_b$-$tyrA_p$-aroF* operon.

Well-documented examples of mixed-function supraoperons in *E. coli* can be cited. (1) The *amiB-mutL-miaA-hfq-hflX-hflK-hflC* system governs functions that include cell wall hydrolysis *(amiB)*, DNA repair *(mutL)*, tRNA modification *(miaA)*, and proteolysis *(hlfX-hlfK-hflC)* (Tsui et al., 1994a,b). The supraoperon includes two additional genes of unknown function in the upstream region. Quite elaborate overall mechanisms of transcriptional control have been demonstrated. These include the use of multiple internal promoters, rho-dependent and rho-independent intraoperon attenuation, and RNA processing. (2) The *aroK-aroB-urf-dam-rpe-gph-trpS* system governs functions that include shikimate kinase *(aroK)*, dehydroquinate synthase *(aroB)*, DNA adenine methyltransferase *(dam),* D-ribulose-5-P epimerase *(rpe)*, 2-phosphoglycolate phosphatase *(gph)*, and tryptophanyl-tRNA synthetase *(trpS)* (Lyngstadass et al., 1995). (3) The *serC(pdxF)-aroF* system links the 3-phosphoserine aminotransferase step of serine biosynthesis and a transamination step of pyridoxine biosynthesis with the 5-enolpyruvylshikimate 3-phosphate (EPSP) syn-

thase step in the common early pathway of aromatic amino acid biosynthesis (Duncan and Coggins, 1986; Lam and Winkler, 1990).

The serC(pdxF)-aroF *mixed-function supraoperon*

Attempts to explain the *E. coli serC(pdxF)-aroF* gene organization in terms of functional interpathway relationships have been pursued quite extensively (Man et al., 1997). A rationale to explain why synthesis of SerC(PdxF) and AroF should be coordinated includes (1) their function in biosynthetic pathways (pyridoxine and aromatic amino acids), which, although formally separate, each draw on a common pool of erythrose-4-P, (2) their joint roles in the synthesis of an iron siderophore, enterochelin, which is derived from both serine and chorismate, and (3) their joint roles in the formation of L-tryptophan, as it is derived from both serine and chorismate. Thus, the *serC(pdxF)-aroF* operon may provide the means of coordinating the expression of these two genes so that enterochelin biosynthesis can proceed efficiently in response to iron starvation, as well as so that serine and chorismate availability can be tuned to the varying demands for tryptophan biosynthesis. Exactly how control of only a single gene from each of these complex pathways could coordinate flux to enterochelin and to L-tryptophan, however, is not at all clear.

Transcription analysis suggests that the downstream *aroF* is controlled by transcription attenuation, a mechanism involving modulation of the efficiency of a transcription terminator (Man et al., 1997). There is a promoter in front of *serC(pdxF)*, an attenuator between *serC* and *aroF*, and a terminator following *aroF*. Transcriptional analysis revealed that two major transcripts were initiated from a promoter upstream from *serC(pdxF)*. About 88% of *serC* transcripts were present in single gene mRNA molecules that likely arose by rho-independent termination between *serC* and *aroF*. The remaining 12% of the transcripts continued through *aroF* and terminated at another rho-independent terminator near the end of *aroF*. Recently, it was reported that expression of the *serC(pdxF)-aroF* supraoperon is regulated over a 22-fold range by global regulatory mechanisms, with transcription being regulated positively by Lrp and negatively by CRP-cAMP (Man et al., 1997).

The characterization of mixed-function supraoperons in the foregoing examples reveals a general picture of a compact array of cistrons and a complexity of multiple, overlapping transcripts that arise as the result of internal promoters or internal attenuators or both. Thus, superimposed on the simplicity of coordinate gene expression from a whole-system readthrough transcript initiated at the far upstream promoter are possibilities for differential regulation of segmental gene combinations in response to appropriate control cues.

Supraoperon units and interlocking metabolic relationships

The preceding section indicates that known supraoperons contain genes whose relationships with one another are not always straightforward. However, metabolic pathways are a highly branched interwoven network, and unexpected coregulation of any given genes probably reflects little considered metabolic ties exerted at a more global level.

A putative mixed-function supraoperon recently found in *P. aeruginosa* and *Pseudomonas stutzeri* (Xie et al., 1999) contains genes of serine, pyridoxine 5′-phosphate, histidine, and aromatic amino acid biosynthesis. The latter pathways are treated in textbooks as separate pathways for practical reductionistic reasons. However, these pathways are, in fact, linked in many ways. Figure 1 is an attempt to illustrate some pathway relationships that are not ordinarily considered and that may prove relevant to the highly conserved linkages of certain interpathway genes described in this article. Figure 1 shows that erythrose-4-P is a common precursor of not only aromatic amino acids and aromatic vitamins but also of pyridoxal 5′-phosphate. L-Tryptophan and L-histidine both draw on a common pool of PRPP as an early substrate. Every molecule of biosynthetic L-tryptophan produced requires one molecule of L-serine input. As the release of glyceraldehyde-3-P in the tryptophan synthase reaction can be salvaged for recycling to serine biosynthesis, SerC(PdxF) can be viewed as the equivalent of a transamination step that is indirectly responsible for the amino group that is ultimately placed in the side-chain of tryptophan (Xie et al., 1999). Each turn of the cycle results in the net donation of an α-amino group to L-tryptophan that is derived from L-glutamate via the prior catalytic activity of SerC. Thus, the SerC(PdxF) aminotransferase participates not only in both L-serine and pyridoxal 5′-phosphate biosynthesis but also in tryptophan biosynthesis. The $HisH_b$ amino-

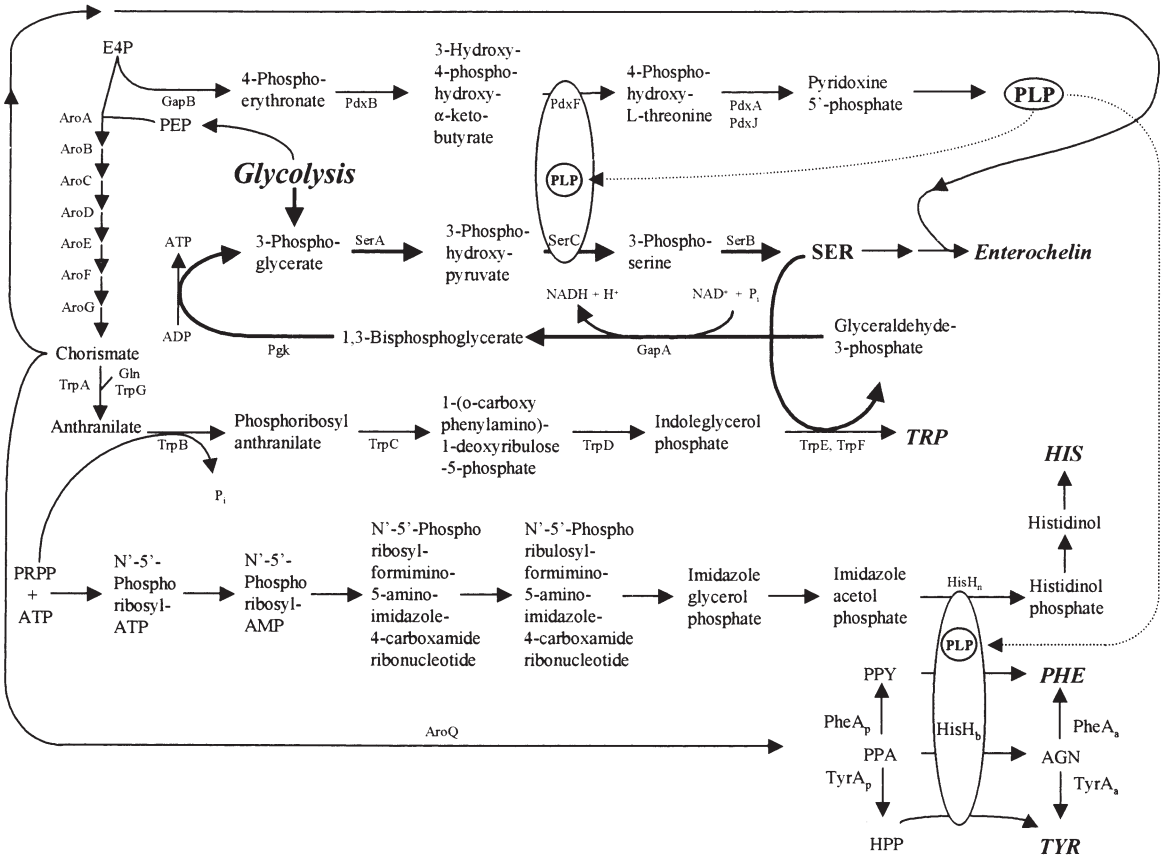


FIG. 1. Interlocking metabolic relationships of biosynthetic pathways leading to pyridoxal 5′-phosphate (PLP), serine (SER), histidine (HIS), enterochelin, and the aromatic amino acids. The serine salvage pathway recycles glyceraldehyde-3-phosphate to 3-phosphoglycerate via the two glycolytic enzymes, glyceraldehyde-3-P dehydrogenase (GapA) and phosphoglycerate mutase (Pgk). Ovals transect arrows depicting the several reactions catalyzed by 3-phosphoserine aminotransferase [SerC(PdxF)] and imidazole acetol phosphate aminotransferase ($HisH_b$). The $PheA_p$ and $PheA_a$ reactions can both be catalyzed by the broad-specificity PheC enzyme (Table 1). Likewise the $TyrA_p$ and $TyrA_a$ reactions can both be catalyzed by $TyrA_c$. E4P, erythrose-4-phosphate; PEP, phosphoenolpyruvate; PPA, prephenate; AGN, L-arogenate; PPY, phenylpyruvate; PHE, L-phenylalanine; HPP, 4-hydroxyphenylpyruvate; TYR, L-tyrosine; TRP, L-tryptophan; PRPP, phosphoribosylpyrophosphate; Gln, L-glutamine; Glu, glutamate; AroA through AroG (see Table 1); $HisH_n$, narrow-specificity imidazole acetolphosphate aminotransferase; AroQ, chorismate mutase (note the existence of two other homology classes denoted AroH and AroR (Gu et al., 1997); $PheA_p$, $TyrA_p$, $PheA_a$, and $TyrA_a$ (see Table 1); TrpA and TrpG, large and small subunits of anthranilate synthase; TrpB, anthranilate phosphoribosyl transferase; TrpC, phosphoribosyl anthranilate isomerase; TrpD, indoleglycerol phosphate synthase; TrpE and TrpF, β and α subunits of tryptophan synthetase; SerA, 3-phosphoglycerate dehydrogenase; SerB, 3-phosphoserine phosphatase; GapB, erythrose-4-P dehydrogenase; PdxB, 4-phosphoerythronate dehydrogenase; PdxA and PdxJ are thought to be involved in formation of the PNP pyridine ring from 1-deoxy-D-xylulose and 4-phosphohydroxy-L-threonine (Hill and Spenser, 1996).

transferase is competent to participate in the biosynthesis of L-phenylalanine and L-tyrosine, in addition to L-histidine (Gu et al., 1995; Jensen and Gu, 1996). The pathway end product, pyridoxal 5′-phosphate, which depends on SerC(PdxF) for its biosynthesis, is itself an essential cofactor for SerC(PdxF) (and for $HisH_b$) function. L-Serine and chorismate of the aromatic amino acid pathway are both precursors for synthesis of the iron siderophore, enterochelin.

Some of the foregoing interlocking relationships apply to a broader distribution of organisms than others. Thus, at one extreme, the input of L-serine into L-tryptophan biosynthesis appears to be universal. On

the other hand, some relationships are much more narrowly distributed; for example, enterochelin is made by enteric bacteria but not by *Pseudomonas* species. (However, another siderophore compound made by *P. aeruginosa* called pyoverdine [Merriman et al., 1995; Wendenbaum et al., 1983] does resemble enterochelin in that both chorismate and serine provide precursor input.)

GENE AND GENE PRODUCT ACRONYMS

In this report, a comparative analysis of gene organization is pursued extensively. The contemporary erratic naming of genes in different organisms is an increasingly awkward problem when such comparisons are attempted, and a universal naming system is inevitable. Therefore, uniform acronyms are proposed as set forth in Table 1. Many of these acronyms have already been established by Gu et al. (1997) and Subramaniac et al. (1998). Thus, genes encoding enzymes in the common pathway portion of aromatic biosynthesis *(aroA→aroG)* or of tryptophan biosynthesis *(trpA→trpF)* are named in order of reaction sequence. The nonhomologue classes of *aroC* are referred to as $aroC_{I}$ and $aroC_{II}$. Genes encoding isoenzyme paralogs of shikimate kinase in *E. coli* are referred to by use of subscripts ($aroE_{K}$ and $aroE_{L}$). In other organisms it is generally not clear whether *aroE* corresponds to *E. coli* $aroE_{K}$ or $aroE_{L}$, although with the many cases of linkage with *aroB*, an identity with $aroE_{K}$ is suggested. Genes encoding homologue enzymes with different substrate specificities are distinguished with lowercase subscript denotations. Genes that specify fusion proteins having multiple catalytic domains corresponding to single gene counterparts elsewhere are named to identify each domain (separated by bullets as suggested by Crawford, 1989. Thus, $AroQ_{p}$• and $AroQ_{t}$• specify isoenzyme domains of chorismate mutase located on the bidomain $AroQ_{p}$•PheA and $AroQ_{t}$•$TyrA_{c}$ proteins of enteric bacteria, respectively. Regulatory paralogs of AroA are distinguished with capitalized subscript denotations that indicate allosteric specificity (Subramaniac et al., 1998); for example, $aroA_{Y}$ encodes an *E. coli* isoenzyme of DAHP synthase that is sensitive to feedback inhibition by tyrosine.

In this article the gene families represented by the individual gene members of the *Pseudomonas* supraoperon are analyzed comprehensively. This is followed by an overview of the dynamics of gene organization that can be inferred from information in the database about the genes that reside in the *Pseudomonas* and *Bacillus* supraoperons.

THE *serC* GENE FAMILY

SerC members comprise one protein family within a huge aminotransferase superfamily assemblage (Mehta et al., 1993). A multiple alignment of 17 deduced amino acid sequences is presented in Figure 2, which includes SerC representatives from eukarya (animals, plants, and microorganisms), bacteria, and archaea. A total of 21 residues are invariant throughout the phosphoserine aminotransferase family, including the 4 that are invariant throughout the larger aminotransferase superfamily (Mehta et al., 1993). The latter four residues, which include the active site lysine, are marked with asterisks in Figure 2. An additional 9 residues are invariant except for SerC from *Methanosarcina barkeri*, the single archaeon representative available. It is curious that current genome annotations of three archaeon organisms indicate the presence of *serA* and *serB* orthologues but absence of a *serC* orthologue. As *nifS* is an established, albeit remote, *serC* homologue (Mehta and Christen, 1993) and as these archaeon organisms process *nifS* genes, the latter might be functional equivalents of *serC*. Because the three sequences at the bottom of Figure 2 exhibit only 20% or less identity with biochemically established phosphoserine aminotransferases, it must be conceded that these might in fact have some other substrate specificity. The latter qualification aside, the consensus motif identified as a signature for phosphoserine aminotransferase (Van der Zel et al., 1989) and marked in Figure 2 is no longer absolute with the inclusion of the two most outlying groups shown at the bottom of the Figure 2 dendrogram. Various short motifs are strongly conserved (e.g., FxxGP near the N-terminus), however.

It would be interesting to know the extent to which the various *serC* homologues other than *E. coli serC* are competent for PdxF function. Our functional complementation results (Xie et al., 1999) with an *E. coli serC(pdxF)* mutant have indicated that the *P. stutzeri* SerC enzyme must have a broad specificity that includes 3-hydroxy-

Table 1. Guide to Gene Designations

Gene designations			
Homology group	*Gene members*	*Corresponding gene products (abbreviations used)*	*Gene designations in common use*
aroA	$aroA_F$	PHE-sensitive DAHP synthase	*aroG (Escherichia)*
	$aroA_W$	TRP-sensitive DAHP synthase	*aroH (Escherichia)*
	$aroA_Y$	TYR-sensitive DAHP synthase	*aroF (Escherichia)*
aroB	*aroB*	3-Dehydroquinate synthase	*aroB (Escherichia)*
aroC	$aroC_I$	3-Dehydroquinate dehydratase (biosynthetic)	*aroD (Escherichia)*
	$aroC_{II}$	3-Dehydroquinate dehydratase (catabolic)	
aroD	*aroD*	Shikimate dehydrogenase	*aroE (Escherichia)*
aroE	$aroE_K$	Shikimate kinase I	*aroK (Escherichia)*
	$aroE_L$	Shikimate kinase II	*aroL (Escherichia)*
aroF	*aroF*	5-Enolpyruvylshikimate-3-phosphate (EPSP) synthase	*aroA (Escherichia)*
aroG	*aroG*	Chorismate synthase	$aroC_{II}$ *(Escherichia)*
hisH	$hisH_b$	Wide-specificity imidazole acetolphosphate aminotransferase	*hisH (Zymomonas)*
	$hisH_n$	Specific imidazole acetolphosphate aminotransferase	*hisC (Escherichia)*
			his8 (Saccharomyces)
aroQ	$aroQ_t$•	Chorismate mutase domain of T-protein (CM-T)	*tyrA (Escherichia)*
	$aroQ_p$•	Chorismate mutase domain of P-protein (CM-P)	*pheA (Escherichia)*
	$aroQ_f$	Monofunctional chorismate mutase (CM-F)	
	•$aroQ_d$	Chorismate mutase domain of aroA•$aroQ_d$	*aroG (Bacillus)*
aroH	*aroH*	Monofunctional chorismate mutase	*aroH (Bacillus)*
tyrA	$tyrA_c$	Cyclohexadienyl dehydrogenase (CDH)	*tyrC (Zymomonas)*
	$tyrA_p$	Prephenate dehydrogenase (PDH)	*tyrA (Escherichia)*
	$tyrA_a$	Arogenate dehydrogenase (ADH)	
	$tyrA_x$	Substrate specificity not established	
pheA	$pheA_p$	Prephenate dehydratase (PDT)	*pheA (Escherichia)*
	$pheA_a$	Arogenate dehydratase (ADT)	
pheC	*pheC*	Cyclohexadienyl dehydratase (CDT)	*pheC (Pseudomonas)*

Fused genes	*Corresponding gene products*	*Current gene designations*
$aroQ_p$•*pheA*	Chorismate mutase•prephenate dehydratase (P-protein)	*pheA (Escherichia)*
$aroQ_t$•$tyrA_c$	Chorismate mutase•cyclohexadienyl dehydrogenase (T-protein)	*tyrA (Escherichia)*
aroA•$aroQ_d$	DAHP synthase•chorismate mutase	*aroG (Bacillus)*
$tyrA_x$•$aroQ_p$•*pheA*	Prephenate dehydrogenase•chorismate mutase•prephenate dehydratase	*pheA (Archeaoglobus)*

4-phospho-α-ketobutyrate (the substrate used by PdxF). We searched the *P. aeruginosa* genomic database for a *serC* paralog that might, in contrast to the results with *P. stutzeri*, encode a narrow specificity PdxF. A search in all six frames did not reveal any paralog candidate. This, plus the presence of a *pdxB* gene in the *P. aeruginosa* genome, implies that both *P. stutzeri* and *P. aeruginosa* possess *E. coli*-like *serC(pdxF)* genes.

The dendrogram of Figure 2 does not always show SerC relationships that mirror the phylogeny of the home organisms. For example, the *Mycobacterium* SerC representatives cluster with the archaeon homologue rather than with other bacterial homologues. (Alternatively, this might be explained by a different substrate specificity, as discussed previously.) *Bacillus* SerC representatives cluster with higher plant SerC homologues rather than with other eubacteria, which could imply an endosymbiotic relationship.

THE *aroQ* GENE FAMILY

Chorismate mutases in nature fall into three distinct homology families (Gu et al., 1997, and refs. therein). Members of the AroQ protein family exhibit low pairwise identities, but conserved catalytic residues es-

FIG. 2. Multiple alignment of deduced amino acid sequences of SerC. The PILEUP software program of GCG was used to obtain the multiple alignment shown. The dendrogram generated is shown at the upper left. Immediately to the right of the dendrogram are individual designations and accession or contig numbers. Ending residue numbers are shown at the right of each horizontal data block. Invariant residues within all six clusters compared are boxed, and highly conserved residues are printed in boldface type. The four invariant residues in the group of vitamin B_6-dependent aminotransferases, including the active site lysine, are marked with an asterisk. Center dots indicate gaps introduced to optimize the alignment. In many aminotransferases, the pyridoxal-5′-phosphate-binding Lys has been identified by chemical analysis. The overlining indicates the phosphoserine aminotransferase protein signature LKGHR $(X)_2$GG(X)R (Van der Zel et al., 1989). Ocu, *Oryctolagus cuniculus*; Cel, *Caenorhabditis elegans*; Sol, *Spinacia oleracea*; Ath, *Arabidopsis thaliana*; Bci, *Bacillus circulans*; Bsu, *Bacillus subtilis*; Eco, *Escherichia coli*; Sga, *Salmonella gallinarum*; Yen, *Yersinia enterocolitica*; Pst, *Pseudomonas stutzeri*; Hin, *Haemophilus influenzae*; Nme, *Neisseria meningitidis*; Spo, *Schizosaccharomyces pombe*; Sce, *Saccharomyces cerevisiae*; Mtu, *Mycobacterium tuberculosis*; Mle, *Mycobacterium leprae*; Mba, *Methanosarcina barkeri*.

tablished by x-ray crystallography of Eco $AroQ_p$• have facilitated recognition of homology. Gu et al. (1997) presented an analysis of 14 proteins belonging to the AroQ protein family. These included the $AroQ_p$• domains of six P-proteins, the $AroQ_t$• domains of four T-proteins, the *B. subtilis* $AroQ_d$• domain fused to a catalytic domain for 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase, and three monofunctional $AroQ_f$ species (two having cleavable signal peptides). Figure 3 displays a multiple alignment updating an expanded AroQ protein family that contains 22 members.

In one case (Archaeoglobus fulgidus), •AroQ• is the central catalytic domain of an apparently trifunctional protein that possesses an N-terminal TyrA domain and a C-terminal PheA domain.

Interestingly, two new substrate specificities are now represented. *Streptomyces pristinaespiralis* uses a mutase reaction encoded by *papB* in a pathway generating 4-dimethylamino phenylalanine, a precursor of the antibiotic pristinamycin (Blanc et al., 1997). PapB converts 4-amino-4-deoxy chorismate to 4-amino-4-deoxy prephenate. Thus, PapB recognizes a substrate with a 4-amino substituent, whereas AroQ recognizes a 4-hydroxy substituent in an otherwise identical molecule. As Mtu $AroQ_f$ and Ehe $AroQ_f$ cluster with Spr PapB and as both *Mycobacterium tuberculosis* and *Erwinia herbicola* possess other AroQ paralogs, Mtu $AroQ_f$, and Ehe $AroQ_f$ could quite possibly have PapB substrate specificity. However, Ehe $AroQ_f$ has been shown to catalyze the chorismate mutase reaction in vitro (Xia et al., 1993a). PapB differs from its two closest homologues in its lack of a cleavable signal peptide.

The second substrate specificity is represented by PchB from *P. aeruginosa*. The two proteins most similar to PchB are from *Vibrio vulnificus* and *Pseudomonas fluorescens* and are probably functionally similar orthologues. PchB catalyzes a step in salicylate biosynthesis in *P. aeruginosa* (Serino et al., 1995). Serino et al. (1995) conclude that PchA converts chorismate to isochorismate and PchB converts isochorismate to salicylate (and pyruvate). The latter reaction is analogous to that of chorismate lyase, which converts chorismate to 4-hydroxybenzoate (and pyruvate). We suggest an alternative. PchB catalyzes a mutase reaction using isochorismate as substrate and producing isoprephenate. This reaction has been demonstrated in higher plants by Zamir et al. (1993). For the following step of salicylate production, PchA has an N-terminal extension that is absent in a number of other isochorismate synthases, and we suggest that this may comprise a catalytic domain for what might be termed isoprephenate lyase. Although the N-terminal region of PchA shows no obvious homology to chorismate lyase, as one might expect, such facile reactions exhibit great divergence (or probability for independent origin), as already illustrated by chorismate mutase orthologues and analogues.

THE *pheA* GENE FAMILY

Modular organization

Figure 4 shows the modular organization shared by all known PheA proteins. An N-terminal catalytic domain (C-domain) is joined to a carboxy-terminal allosteric domain (R-domain). Xia et al. (1992) showed that when 260 bp was excised from the *E. herbicola* •PheA domain at the 3′-end, catalytic competence was retained, but allosteric effects were lost. A discrete allosteric domain is consistent with the mutation analysis results of Nelms et al. (1992). The lengths of the catalytic domains are quite uniform (except for the N-terminal extension found for the *Xanthomonas* enzyme) (Gu et al., 1997). Nine C-domain residues are invariant, and many other residues are highly conserved. All of the PheA proteins that have been studied at the enzyme level are specific for prephenate. However, the higher plant *(Arabidopsis)* PheA protein may prove to be an arogenate-specific protein, as arogenate dehydratase, but not prephenate dehydratase, is readily found in higher plants (Jung et al., 1986). The single broad-specificity cyclohexadienyl dehydratase that has thus far been cloned and sequenced was found to lack homology with PheA proteins (Fischer et al., 1991).

Homologue R-domains exist in mammalian aromatic amino acid hydroxylases

We noticed that BLAST analyses of various PheA proteins submitted as query entries consistently returned hits for mammalian aromatic amino acid hydroxylases. Alignment matchups were between

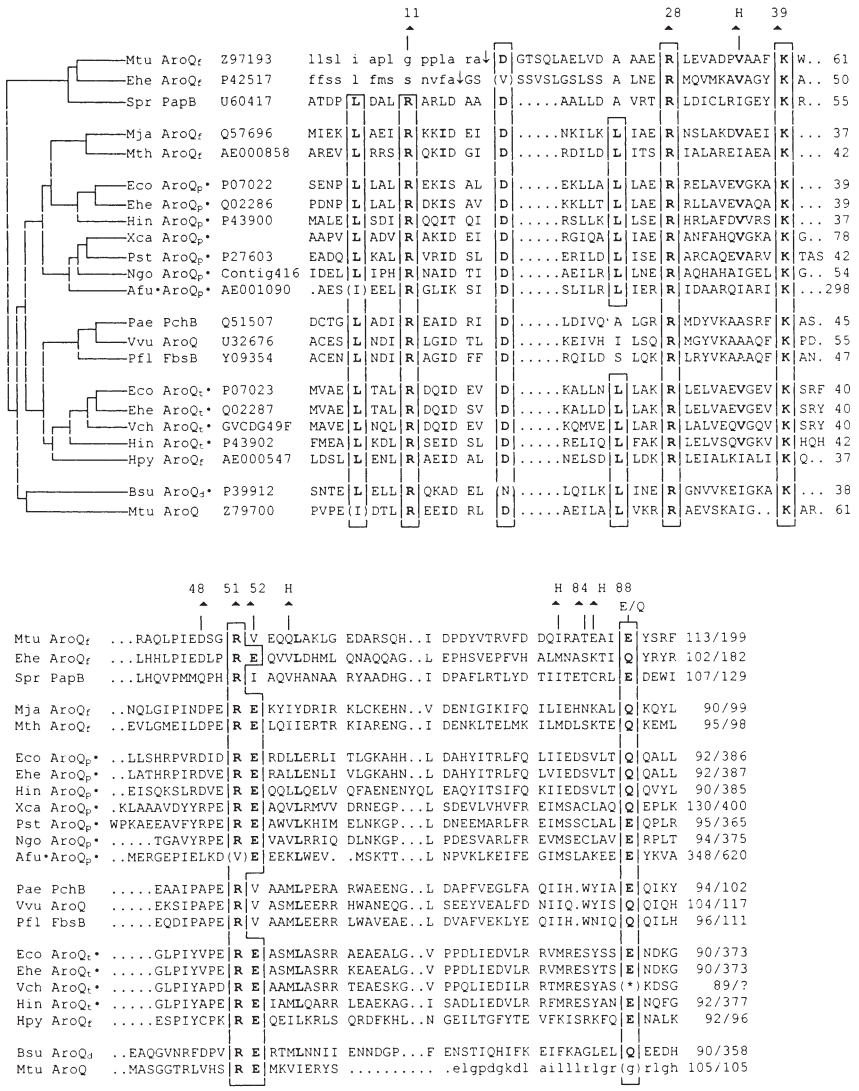


FIG. 3. Multiple alignment of the AroQ protein family. The PILEUP software program of GCG was used to obtain the multiple alignment shown. The dendrogram generated is shown at the upper left. Immediately to the right of the dendrogram are individual designations and accession or contig numbers. Ending residue numbers are shown at the right of each horizontal data block. At the lower right is given the number of the final residue presented, followed (slash) by the number of the final sequence residue. The most divergent members of the family are Ehe $AroQ_f$, Mtu $AroQ_f$ and Spr PapB, which comprise one of six groups shown by the spacing used. Lowercase letters for amino acids denote the signal sequences that are cleaved from the top two proteins. Residues located at the active site of Eco $AroQ_p$• as demonstrated by x-ray crystallography are indicated by residue numbers at the top. With respect to other indicated residues, residue 11′ is located on the other subunit in the active homodimer. Residues conserved through at least the bottom major cluster are boxed. Other highly conserved residues are in boldface. Residues marked with H at the top participate in important hydrophobic interactions in Eco $AroQ_p$. Residue numbers are shown at right. An asterisk marks the position of residue 85 in the Vch $aroQ_t$ sequence, which almost certainly is a sequencing error. The present TAA stop codon is unlikely because of its location within the coding region of other homologues; residues immediately before and after the codon are highly conserved. It is likely that this residue is E(GAA) or Q(CAA). Mtu, *Mycobacterium tuberculosis*; Ehe, *Erwinia herbicola*; Spr, *Streptomyces pristinaespiralis*; Mja, *Methanococcus jannaschii*; Mth, *Methanobacterium thermoautotrophicum*; Eco, *Escherichia coli*; Hin, *Haemophilus influenzae*; Xca, *Xanthomonas campestris*; Pst, *Pseudomonas stutzeri*; Ngo, *Neisseria gonorrhoeae*; Afu, *Archaeoglobus fulgidus*; Pae, *Pseudomonas aeruginosa*; Vvu, *Vibrio vulnificus*; Pfl, *Pseudomonas fluorescens*; Vch, *Vibrio cholerae*; Hpy, *Helicobacter pylori*; Bsu, *Bacillus subtilis*. $AroQ_f$, monofunctional chorismate mutase, $AroQ_p$•, chorismate mutase domain of P-protein; $AroQ_t$•, chorismate mutase domain of T-protein; •$AroQ_d$, chorismate mutase domain of AroA•$AroQ_d$; PchB and FbsB, isochorismate mutase; PapB, 4-amino 4-deoxychorismate mutase.

(A)

```
Cgl  PheA    8  VAYLGPA GTFTEEALYK FADAGVF... .GDGEIEQLP AKSPQEAVDA VRHG....TA QFAVVAI|EN|F VD|G|PVTPTF|D| AL..... ..... DQGSNVQIIA |E|EELDIA  95
Mtu  PheA    4  IAYLGPE GTFTEAALVR MVAAGLVPE. TGPDALQRMP VESAPAALAA VRDG....GA DYACVPI|EN|S ID|G|SVLPTL|D| SL..... ..... AIGVRLQVFA |E|TTLDVT  94
Ame  PheA    4  IAYFGPV GTFTEQAART FMAAG..... .....DELVA AETIPKALDA VRRE....A DAACVPV|EN|S VE|G|AVPATL|D| SL..... ..... AVGEPLIGVA |E|ALLPVH  84
Eco •PheA  106  IAFLGPK GSYSHLAARQ Y......AA. RHFEQFIESG CAKFADIFNQ VETG....QA DYAVVPI|EN|T SS|G|AINDVY|D| LL..... ..... .QHTSLSIVG |E|MTLTID 189
Ehe •PheA  106  IAFLGPK GSYSHLAARN Y......AS. RHFDSMVECG CLKFHDIIKQ VENG....VA DYAVMPI|EN|T SS|G|SINDVY|D| LL..... ..... .QQTSLSIVG |E|LTLPID 189
Hin •PheA  106  IAFLGKR GSYSNLAARN Y......AA. RYQKQFVELG CQSFEQVFEK VQTG....EA DFGVLPL|EN|T TS|G|AINEVY|D| LL..... ..... .QHTDLSLVG |E|LAYPIK 189
Pst •PheA   97  VAYLGPE GTFSQAAALK .......... HFGHSVISKP MAAIDEVFRE VVAG....AV NFGVVPV|EN|S TE|G|AVNHTL|D| SF..... ..... .LEHDIVICG |E|VELRIH 177
Xca •PheA  132  IGYLGPE GTFSQQAVLK .......... HFGRSAVGLP MATIEEVFQE VEAG....NA DFGVVPV|EN|S GQ|G|TIQVTL|D| MF..... ..... .LTSNLKICG |E|VELRVH 212
Ngo •PheA  105  IAYLGPQ GTFTQQAAIK .......... HFGHAAHTMA CPTIDDCFKQ VETR....QA DYLVAPV|EN|S TE|G|SVGRTL|D| LL..... ..... .AVTALQACG |E|VVLRIH 185
Bsu  PheA    3  VGYLGPA ATFTHLAV.. .......SS. CFQNGAEHVA YRTIPECIDA AVAG....EV DFAFVPL|EN|A LE|G|SVNLTI|D| YL..... ..... IHEQPLPIVG |E|MTLPIH  84
Lla  PheA    3  IAYLGPR GSFCSVVA.. .......EA. AF.KSEELYS YATILDVIEA YNEG....EC DFALVPI|EN|S TE|G|TVNMSI|D| KI..... ..... FHDSNAKVVA |E|FVLPIS  83
Ssp  PheA    6  VAHLGPR GTNTETVALV YGE..QHQE. RTGQAVEFFP CPSIGKTLEA AAQR....LV DVAIVPV|EN|S TE|G|GIAITL|D| CL..... ..... WAAENLKIQQ |E|LVLPIT  94
Mja  PheA    5  VIYTLPK GTYSEKATKK FLD..YIDG. DYKIDY.... CNSIYDVFER VDNN..... GLGVVPI|EN|S IE|G|SVSLTQ|D| LL..... ..... LQFKDIKILG |E|LALDIH  87
Ath  PheA  127  VAYQGVP GAYSEAAA.. ........G. KAYPNCDAIP CDQFDVAFQA VELW....IA DRAVLPV|EN|S LG|G|SIHRNY|D| LL..... ..... LRHR.LHIVG |E|VQIPVH 206
Afu •PheA  354  IAVLGPQ GSFSEEMALK LV........ ..GSRVPLRY CSTTDEIIKL VESG....EV DYGLVPI|EN|S VN|G|TVLPVI|D| AL..... ..... LNH.DVEVFG |E|AKLEVN 434
Sce  PheA   42  VLFLGPK GTYSHQAALQ Q........F QSTSDVEYLP AASIPQCFNQ LENDTS...I DYSVVPL|EN|S TN|G|QVVFSY|D| LLRDRMIKKA LSLPAPADTN RITPDIEVIA |E|QYVPIT 144
Spo  PheA    6  IAFLGPR GTFSHQAALL ARPDSLLCSL PSFAGKLIFA LKFILICFLA VLEALSSRQV DYAVLPI|EN|S TN|G|AVIPAY|D| LLKGR..... ..... ...DDIQAVG |E|VLVPAH 101

                                                                                                                                                C domain ←┐
Cgl  PheA  FSI S..VRPGTSL ADVKTL.... .ATHPVGYQ|Q| ...VKNWMAT TIPDAMYLSA ..SSNGAGAQ MVAE..GTAD A.....AAAP SRAAELFGLE RLVDD.VADV RGAR|TRF| VAVQAQAAVS 195
Mtu  PheA  FSI V..VKPGRNA ADVRTL.... .AAFPVAAA|Q| ...VRQWLAA HLPAADLRPA ..YSNADAAR QVAD..GLVD A.....AVTS PLAAARWGLA ALADG.VVDE SNAR|TRF| VLVGRPGPPP 194
Ame  PheA  FSV L..TRDD..V GEIRTV.... .ASHPHALA|Q| ...VRKWLED NLPGARVVAA ..GSTAAAAV AVQA..GEFD A.....AVTA PVAVEHYPLK VLATE.VADV RDAR|TRF| LLMRRPPVVL 182
Eco •PheA  HCL L..VSGTTDL STINTV.... .YSHPQPFQ|Q| ...CSKFLNR .YPHWKIEYT ..ESTSAAME KVAQAKSPHV A.....ALGS EAGGTLYGLQ VLERI.EANQ RQNF|TRF| VVLARKAINV 290
Ehe •PheA  HCV L..VNGPTDL QQIETV.... .YSHPQPFQ|Q| ...CSQFINR .FPHWKIEYT ..ESTAAAME KVAALNSPKV A.....ALGS EAGGELYQLQ VLERN.LANQ QQNH|TRF| IVLARKAIEV 290
Hin •PheA  HCV L..VNDKTDL NQIDTL.... .YSHPQVIQ|Q| ...CSQFIHS .LDRVHIEYC ..ESSSHAMQ LVASLNKPNI A.....ALGN EDGGKLYGLS VLKTN.IANQ ENNI|TRF| IVVAKEPREV 290
Pst •PheA  HHL L..VGETTKT DRITRI.... .YSHAQSLA|Q| ...CRKWLDA HYPNVERVAV ..SSNADAAK RVKSEWNS.. A.....AIAG DMAAQLYGLS KLAEK.IEDR PDNS|TRF| LIIGSQEVPP 277
Xca •PheA  QYL L...SRNGRL EDIERI.... .YAHSQSFA|Q| ...TAGWLRS HLPKVEKIAV ..SSNAEGAR RARNAEDA.. A.....AIGG ESAAHVYGLK KVIMKSIEDD DDNT|TRF| LVIGRQIFPS 312
Ngo •PheA  HNL L..RKNNGST EGIAKV.... .FSHAQALA|Q| ...CNDWLGR RLPNAERIAV ..SSNAEAAR LVAESDDGTV A.....AIAG RTAAEIYGLD MVAE.CIEDE PNNT|TRF| LVMGHHETGA 287
Bsu  PheA  QHL LVHPSRENAW KELDKI.... .YSHSHAIA|Q| ...CHKFLHR HFPSVPYEYA ..NSTGAAAK FVSDHPELNI G.....VIAN DMAASTYELK IVKRD.IQDY RDNH|TRF| VILSPDENIS 188
Lla  PheA  QNL LA.VSKE... QKIEHI.... .YSHPQALA|Q| ...TRVYLRK FYPQAQVEIT ..ESTSAAAE FVKNNPDLPA A.....AVAN SFAAKMYDLE FIAEN.IQDL AGNS|TRF| WLLG KEKQS 182
Ssp  PheA  HVL L...SRGKLL ADLTKV.... .VSHPQALG|Q| ...CQKWLGA HLPQVSLVPA ..NSTTEAIQ IIGDDP..TA A.....AIAS PRAATLFDLS VLQTD.IQDY PDNC|TRF| .....WAISS 188
Mja  PheA  HNL IGYDK..... NKIKTV.... .ISHPQALA|Q| ...CRNYIKK H..GWDVKAV ..ESTAKAVK IVAESKDETL G.....AIGS KESAEHYNLK ILDEN.IEDY KNNK|TRF| ILIGKKVKFK 184
Ath  PheA  HCL LALPGVRT.. DCVSRV.... .ISHPQALA|Q| ...TEHSLDV LTPHAAREAF ..HDTAAAAE YISANDLHDT A.....AVAS ARAAELYNLQ ILA.DGIQDD PGNV|TRF| LMLAREPIIP 308
Afu •PheA  HCL VA..KRKIEL KEIKTI.... .YSHPQAVA|Q| ...CMGFINN YLPSVAIRYT ..TSTSDAAR MLDDYS.... A.....AIMS ENAARFYRLH VLRKGIQDLK GRNI|TRF| .....YLIRR 528
Sce  PheA  HCL ISPIQLPNGI ASLGNFEEVI IHSHPQVWG|Q| VECYLRSMAE KFPQVTFIRL DCSSTSESVN QCIRSSTADC DNILHLAIAS ETAAQLHKAY IIEHS.INDK LGNT|TRF| LVLKRRENAG 263
Spo  PheA  HCI IGK..... .SLENVQKIL ..SHPQAFG|Q| ...CSKWISA NVPNAEFVSV ..SSTSQAAA LASKDITGTI .....VAISS ELCAVENQFN LLVKN.IEDD SNNR|TRF| LLLRSGGFQD 200
```

FIG. 4. *((Continued on next page))*

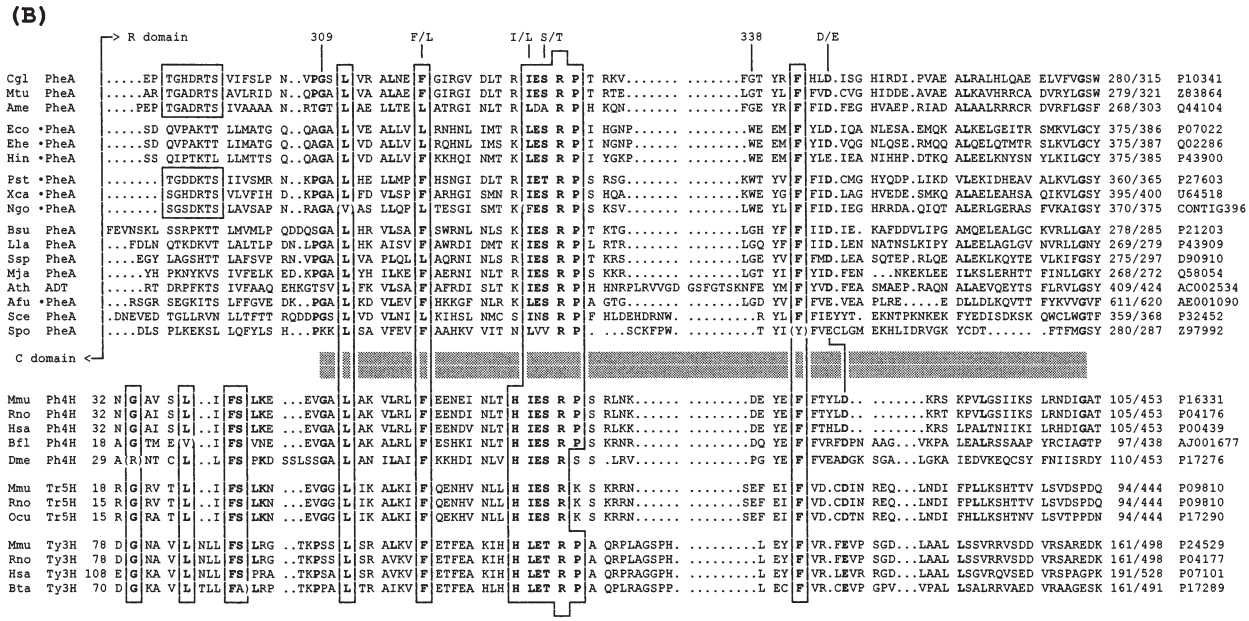


FIG. 4. (*Continued*) Multiple alignment of the PheA protein family. **(A)** Alignment of the N-terminal catalytic domains of monofunctional prephenate dehydratases (PheA) and P-protein prephenate dehydratase domains (•PheA). The PILEUP software program of GCG was used to obtain the multiple alignment shown. Ending residue numbers are shown at the right of each horizontal data block. At the right of each horizontal line is given the number of the residue immediately to the left. Invariant residues are boxed, and highly conserved residues are shown in boldface. Cgl, *Corynebacterium glutamicum*; Mtu, *Mycobacterium tuberculosis*; Ame, *Amycolatopsis methanolica*; Eco, *Escherichia coli*; Ehe, *Erwinia herbicola*; Hin, *Haemophilus influenzae*; Pst, *Pseudomonas stutzeri*; Xca, *Xanthomonas campestris*; Ngo, *Neisseria gonorrhoeae*; Bsu, *Bacillus subtilis*; Lla, *Lactococcus lactis*; Ssp, *Synechocystis* sp.; Mja, *Methanococcus jannaschii*; Ath, *Arabidopsis thaliana*; Afu, *Archaeoglobus fulgidus*; Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*. **(B)** Alignment of C-terminal inhibitor domains of PheA protein family members (top) with N-terminal inhibitor domains of eukarya hydroxylase protein family members (bottom). At the right are the numbers of the final residues presented, followed (slash) by the number of the final sequence residue and accession or contig number. The junction separating the catalytic and allosteric regions is indicated. Residues 309 and 338, shown by mutation analysis in *E. coli* to be important for feedback inhibition (Nelms et al., 1992), are marked. The motif (S/TGxDR/KTS) suggested for allosteric activation by tyrosine is boxed. Phenylalanine-4-hydroxylase (Ph4H), tryptophan 5-hydroxylase (Tr5H), and tyrosine 3-hydroxylase (Ty3H) from Mmu, *Mus musculus*; Rno, *Rattus norvegicus*; Hsa, *Homo sapiens*; Bfl, *Branchiostoma floridae*; Dme, *Drosophila melanogaster*; Ocu, *Oryctolagus cuniculus*; Bta, *Bos taurus*.

the N-terminal region of the hydroxylases (within a region known as the R-domain) and the C-terminal region of the dehydratases. When the Blocks database was searched (BLKSORT Version 7/9/97) using the dehydratase allosteric domain as a query, the Blksort Hits included portions of the hydroxylase R-domain. The alignment in Figure 4B suggests that catalytic domains of eukaryotic hydroxylases and PheA dehydratases each possess fused R-domain homologues (shaded region of multiple alignment shown in Fig. 4B). Homology is evident throughout a region limited to about 70 amino acids within the C-terminal portion of the R-domains. A central motif is IE(S/T)RP (tryptophan hydroxylases are IESRKS). The motif is flanked on the left by an $Lx_3(L/I/V)x_2(F/L)x_6(L/I/M)$ motif and on the right by a conserved phenylalanine residue. A search of the database for other proteins carrying fused R-domains gave negative results. The first 25-residue portion of PheA R-domains exhibits no relationship to the hydroxylase R-domain. The most N-terminal segments of the R-domains of tyrosine hydroxylase, tryptophan hydroxylase, and phenylalanine hydroxylase (not shown in Fig. 4) exhibit variable length and divergent sequences that undoubtedly relate to the individuality of regulation in place.

Oligomerization properties of R-domain proteins

Evidence in the literature suggests that various PheA R-domains can bind one or more aromatic amino acids as modulators of activity, and these frequently promote molecular mass transitions. Three purified P-proteins have been studied in detail, those from *E. coli* (Hudson and Davidson, 1984), *Alcaligenes* (now *Ralstonia*) *eutrophus* (Friedrich et al., 1976), and *Acinetobacter calcoaceticus* (Ahmad et al., 1988). The *E. coli* P-protein is a homodimer that is converted to a tetramer either by exposure to L-phenylalanine or by use of high protein concentrations. Neither L-tyrosine nor L-tryptophan affected activity or molecular mass. The *A. eutrophus* P-protein persists as a tetramer in the presence or absence of L-phenylalanine, L-tyrosine, or L-tryptophan. All three amino acids bound to the enzyme as allosteric effectors, with L-phenylalanine being a potent inhibitor and both L-tyrosine and L-tryptophan being strong activators. The *A. calcoaceticus* P-protein exists as a dimer and is converted to a tetramer in the presence of either L-phenylalanine (inhibitor) or L-tyrosine (activator). L-Tryptophan is an activator with unknown effects on molecular mass. High protein concentration did not facilitate tetramer formation, but hysteretic activation was reported.

Among the monofunctional PheA proteins, the *B. subtilis* enzyme undergoes dimer-to-octamer transitions in the presence of activating effectors (prephenate, L-methionine, or L-leucine) and octamer-to-dimer transitions in the presence of inhibiting effectors (L-phenylalanine or L-tryptophan) (Pierson and Jensen, 1974; Riepl and Glover, 1978). It was concluded that the dimer is intrinsically inactive and the octamer is the active species.

The bacterial phenylalanine hydroxylases lack an R-domain and exist as monomers (Zhao et al., 1994), in contrast to mammalian aromatic amino acid hydroxylases, which possess N-terminal R-domains and are multimers (Hufton et al., 1995). The C-domain of mammalian hydroxylases has a 20-residue C-terminal leucine zipper that stabilizes active tetramer. The C-domains of PheA proteins have no comparable motif. The core C-domain of rat phenylalanine hydroxylases (267 residues) (Dickson et al., 1994) and rat tyrosine hydroxylase (318 residues) (Lohse and Fitzpatrick, 1993) are similar in size to bacterial phenylalanine hydroxylase (267 residues), which is a monomer that lacks an R-domain. The core C-domain of Lohse and Fitzpatrick (1993) was indeed found to be monomeric. Thus, the determinants for dimer and tetramer formation appear to be located within the R-domain, whereas the far C-terminal leucine zipper of the C-domain shifts the dimer-tetramer equilibrium toward tetrameric species. The hydroxylase R-domain is generally considered to be an inhibitory domain that interferes with the active site of the C-domain of rat tyrosine hydroxylase (Hufton et al., 1995). Phosphorylation and dephosphorylation, which are known to occur at serine residues of the R-domain, have been reported to mediate dimer-tetramer interconversion in mammalian phenylalanine hydroxylases (Smith et al., 1984). The activation of rat phenylalanine hydroxylase by L-phenylalanine has also been linked to phosphorylation (Tipper and Kaufman, 1992).

The foregoing general background information suggests that R-domain homologue regions that are fused to catalytic domains of prephenate dehydratase, on the one hand, and to catalytic domains of mammalian aromatic amino acid hydroxylases, on the other hand, are inhibition domains. Effector molecules that bind to the R-domain produce conformational changes that enhance (inhibition) or relax (activation) the C-domain/R-domain interaction. Frequently, but not always, these changes are associated with molecular weight transitions. The details of effector identity and quantitative effect are highly variable from system to system.

In view of such system-specific individuality, perhaps the highly conserved IE(S/T)RP motif of the R-domain specifies an interdomain interface with the C-domain. A close examination of the C-domains of PheA and aromatic hydroxylase protein families did not reveal any common motif candidate, but the key residues sought may be too widely spaced to be recognized.

REGULATION OF *aroQ$_p$•pheA*

In enteric bacteria, such as *E. coli* (Hudson and Davidson, 1984) and *E. herbicola* (Xia et al., 1993b), *aroQ$_p$•pheA* is controlled soley by attenuation. The latter mode of attenuation depends on a ribosome-stalling mechanism operating during translation of a phenylalanine-rich leader peptide located immediately upstream of *aroQ$_p$•pheA*. In this case, alternative stem-loop structures are favored, depending on the rate of leader-peptide translation. *P. stutzeri aroQ$_p$•pheA* clearly lacks an upstream leader peptide that would be needed

to participate in a ribosome-stalling mechanism, as it, in fact, overlaps with *serC*. Translation of *serC* itself probably could not act as a leader peptide, as *serC* is not phenylalanine rich at the carboxy-terminus.

In *Xanthomonas campestris* (Gu et al., 1997), we noted the existence of alternative stem-loop structures between *serC* and $aroQ_p$•*pheA*—one a possible antiterminator structure and the other a rho-independent terminator. As no upstream leader peptide is present, an attenuation mechanism exploiting the alternative stem-loop structures would require an unknown regulatory element (e.g., an RNA-binding protein).

Yet another mechanism of regulation is suggested in both *P. stutzeri* and *P. aeruginosa*, where a strong stem-loop structure was identified that sequesters the ribosome-binding site of $aroQ_p$•*pheA* (Xie et al., 1999). This could provide a mechanism whereby *serC* translation is required to activate $aroQ_p$•*pheA* translation by unmasking the sequestered ribosome-binding site. The additional presence of an alternative stem-loop structure might reflect a mechanism whereby $aroQ_p$•*pheA* is differentially regulated at the translational level. Thus, under conditions of L-serine sufficiency where *serC* translation is minimal, $aroQ_p$•*pheA* translation may be uncoupled from *serC* translation if an appropriate secondary mRNA structure is presented in response to L-phenylalanine limitation. This would require an unknown regulatory gene. In recent years, regulatory mechanisms acting at the level of translation initiation have gained recognition as being more important than previously thought in prokaryotes (De Smit and Van Duin, 1990).

INTERCISTRONIC REGION BETWEEN $aroQ_p$•*pheA* AND $hisH_b$

Within the supraoperon boundaries, only the intergenic junction separating $aroQ_p$•*pheA* and $hisH_b$ is sufficiently long to contain promoters, attenuators, and associated regulatory elements that do not overlap coding regions. In *P. stutzeri*, a strong hairpin structure ($\Delta G = -27.0$ kcal/mol) overlaps the •*pheA* stop codon (Xie et al., 1999). This could be an attenuator structure. If so, unknown elements of regulation upstream and downstream may act to remediate the imperfect uridine-rich segment present, as reported in other systems (Reynolds et al., 1992). Downstream at the far 3′-end of the intergenic space is another hairpin structure ($\Delta G = -19.0$ kcal/mol), which sequesters the ribosome-binding site. An alternative stem-loop structure ($\Delta G = -18.0$ kcal/mol) would expose the ribosome-binding site. Perhaps unknown elements of regulation exist to dictate the stabilization of one structure or the other.

The shorter intergenic region of *P. aeruginosa* was also examined for comparable secondary structure possibilities. No hairpin resembling the far upstream stem-loop of *P. stutzeri* was found. However, a downstream hairpin structure ($\Delta G = -18.0$ kcal/mol) sequestering the ribosome-binding site was found, also accompanied by an alternative stem-loop structure ($\Delta G = -27.0$ kcal/mol).

No hairpin structures were found in the intergenic regions between $hisH_b$ and $tyrA_c$ or between $tyrA_c$ and *aroF*. Thus, the features of close intergenic spacing and the possible mechanisms in place to mask ribosome-binding sites prior to the event of upstream translation may accommodate coupled translation through much or all of the systems under some conditions. Under other conditions, translation may be uncoupled for selective expression of some genes.

THE $hisH_B$ GENE FAMILY

Relationship between imidazole acetol phosphate aminotransferase and histidine/aromatic biosynthesis

Histidine biosynthesis requires an imidazole acetol (IAP) aminotransferase to catalyze the formation of histidinol phosphate by transamination of IAP (Winkler, 1996). Although a number of microbial aminotransferases are not essential for growth because of the backup capabilities attributed to the broad specificities of the intracellular repertoire of aminotransferases (Jensen and Calhoun, 1981), the presence of a family Iβ aminotransferase is essential in all organisms studied that rely on endogenous histidine biosynthesis (Jensen and Gu, 1996).

B. subtilis IAP aminotransferase can provide a backup function in tyrosine and phenylalanine biosynthesis (Nester and Montoya, 1976), a role normally fulfilled by an aromatic aminotransferase specified by *aroJ*. Thus, $hisH_b$ mutants are auxotrophic for L-histidine, whereas *aroJ* mutants remain prototrophic for

histidine, phenylalanine, and tyrosine. Double mutants *(hisH$_b$ aroJ)* require histidine, tyrosine, and phenylalanine. As expected from the in vivo results, purified IAP aminotransferase from *B. subtilis* was shown to transaminate phenylpyruvate and *p*-hydroxyphenylpyruvate in vitro (Weigent and Nester, 1976).

Multiple alignment (Fig. 5) clearly shows that *P. stutzeri* HisH$_b$ belongs to the Iβ subfamily of family I aminotransferases. This subfamily consists entirely of enzymes capable of catalyzing the IAP aminotransferase reaction of L-histidine biosynthesis (Jensen and Gu, 1996). They are not, however, necessarily restricted to this reaction. Subfamily Iβ appears to split into two homology groupings, which correlate with broad substrate specificity (HisH$_b$) and narrow substrate specificity (HisH$_n$). *E. coli* HisH$_n$ exemplifies a case of narrow substrate specificity (Martin et al., 1971). On the other hand, HisH$_b$ enzymes from *Z. mobilis* (Gu et al., 1995) and *B. subtilis* (Nester and Montoya, 1976) illustrate cases of a broadened substrate specificity to accommodate the aromatic ring. It is interesting that *P. aeruginosa*, *Haemophilus influenzae*, and *M. tuberculosis* each possesses two *hisH* paralogs, one of which clusters with the known broad-specificity enzymes (denoted HisH$_b$). As *P. aeruginosa* possesses a *hisH$_n$* paralog, *P. stutzeri* probably also possesses a *hisH$_n$* paralog.

An evolutionary scenario was advanced suggesting that an ancestral *hisH$_b$* gene encoded a broad-specificity enzyme that was competent for both histidine and aromatic amino acid biosynthesis (Gu et al., 1995). Gene duplication produced a gene copy *(hisH$_n$)* that became biochemically specialized for histidine biosynthesis and became incorporated into the histidine operon. The remaining *hisH$_b$*, although still producing a gene product competent for catalysis of the histidine pathway reaction, became specialized for aromatic biosynthesis, as is suggested by its persistent genetic linkage with *tyrA* homologues. If this is so, *E. coli* must have lost *hisH$_b$*, the latter likely having been replaced by *tyrB* (which seems to be unique to enteric bacteria). On the other hand, *B. subtilis* either lost *hisH$_n$* or exists in a lineage whose phylogenetic divergence preceded the hypothetical gene duplication event that generated *hisH$_n$*.

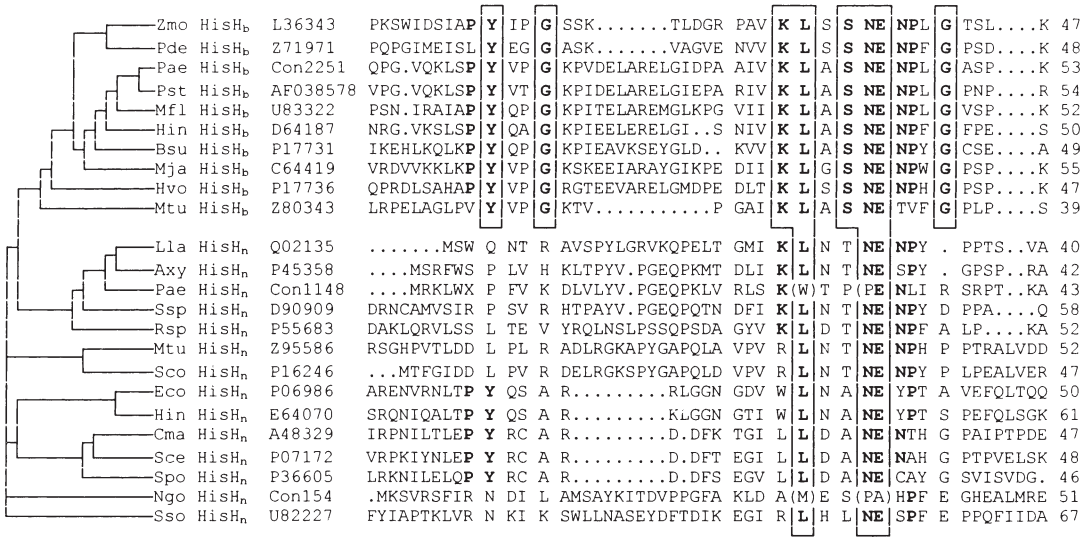
Not consistent with the foregoing scenario is the observation that the HisH$_b$ protein class exhibits more similarity to one subgroup of HisH$_n$ proteins than the latter exhibits to the remaining HisH$_n$ subgroups. Because the application of several tree-building algorithms gave results similar to the PILEUP dendrogram, one has to concede the possibility of the opposite scenario, namely, that *hisH$_n$* was the ancestral gene and *hisH$_b$* emerged more recently. The observation that HisH$_b$ possesses more invariant residues than does HisH$_n$ is also consistent with this possibility.

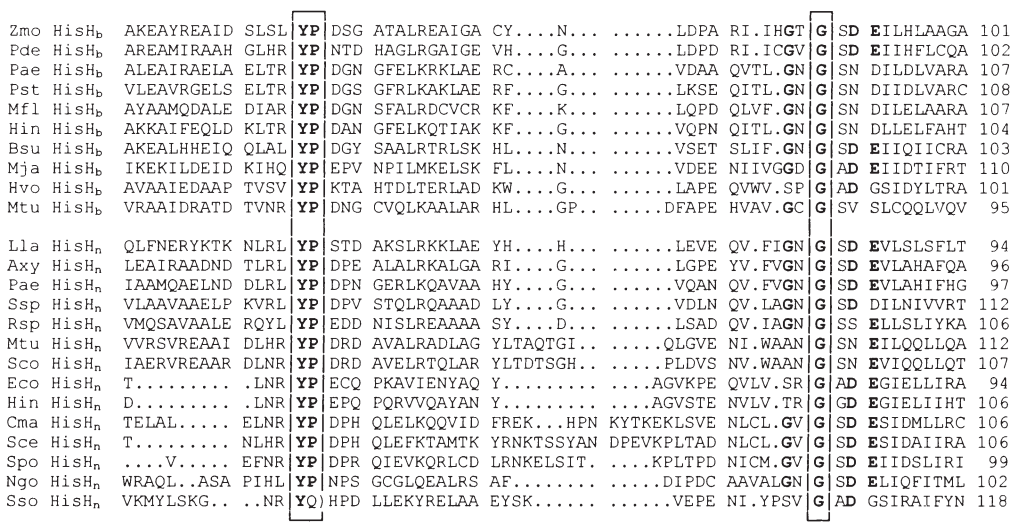
OTHER GENE FAMILIES REPRESENTED BY GENES OF THE *PSEUDOMONAS* SUPRAOPERON

The tyrA *gene family*

Cyclohexadienyl dehydrogenase (TyrA$_c$) has been purified and characterized extensively from *P. stutzeri*. An extensive comparative analysis of this protein family has also been carried out (G. Xie and R.A. Jensen, unpublished observations).

FIG. 5. (*next 3 pages*) Multiple alignment of deduced amino acid sequences of HisH$_b$ and HisH$_n$ proteins. The PILEUP software program of GCG was used to obtain the multiple alignment shown. The dendrogram generated is shown at the upper left. Immediately to the right of the dendrogram are the individual designations, as well as accession or contig numbers. Ending residue numbers are shown at the right of each horizontal data block. At the lower right of the last page of the sequence is given the number of the final residue presented, followed (slash) by the number of the final sequence residue. Residues that are highly conserved are printed in boldface type. Highly conserved residues are boxed in at the level of HisH$_b$ or at the level of the entire HisH homology group. The four invariant residues conserved at the superfamily level of aminotransferase homologues are marked with asterisks. Dot characters indicate gaps introduced to optimize the alignment. Zmo, *Zymomonas mobilis*; Pde, *Paracoccus denitrificans*; Pae, *Pseudomonas aeruginosa*; Pst, *Pseudomonas stutzeri*; Mfl, *Methylobacillus flagellatum*; Hin, *Haemophilus influenzae*; Bsu, *Bacillus subtilis*; Mja, *Methanococcus jannaschii*; Hvo, *Halobacterium volcanii*; Mtu, *Mycobacterium tuberculosis*; Lla, *Lactococcus lactis*; Axy, *Acetobacter xylinum*; Ssp, *Synechocystis* sp.; Rsp, *Rhizobium* sp.; Sco, *Streptomyces coelicolor*; Eco, *Escherichia coli*; Cma, *Candida maltosa*; Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*; Ngo, *Neisseria gonorrhoeae*; Sso, *Sulfolobus solfataricus*; HisH$_b$, broad-specificity imidazole acetolphosphate aminotransferase; HisH$_n$, narrow-specificity imidazole acetolphosphate aminotransferase.





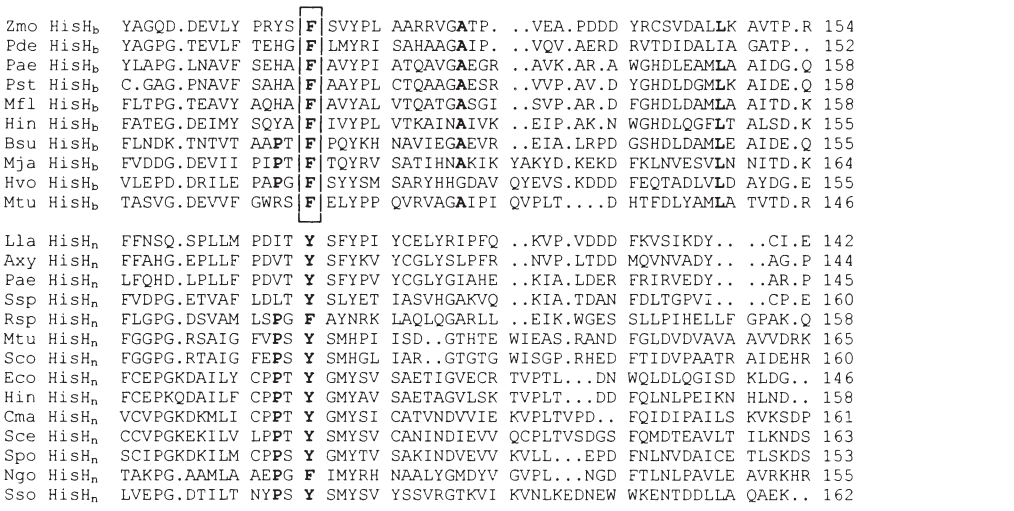


FIG. 5. (*Continued on next page*)

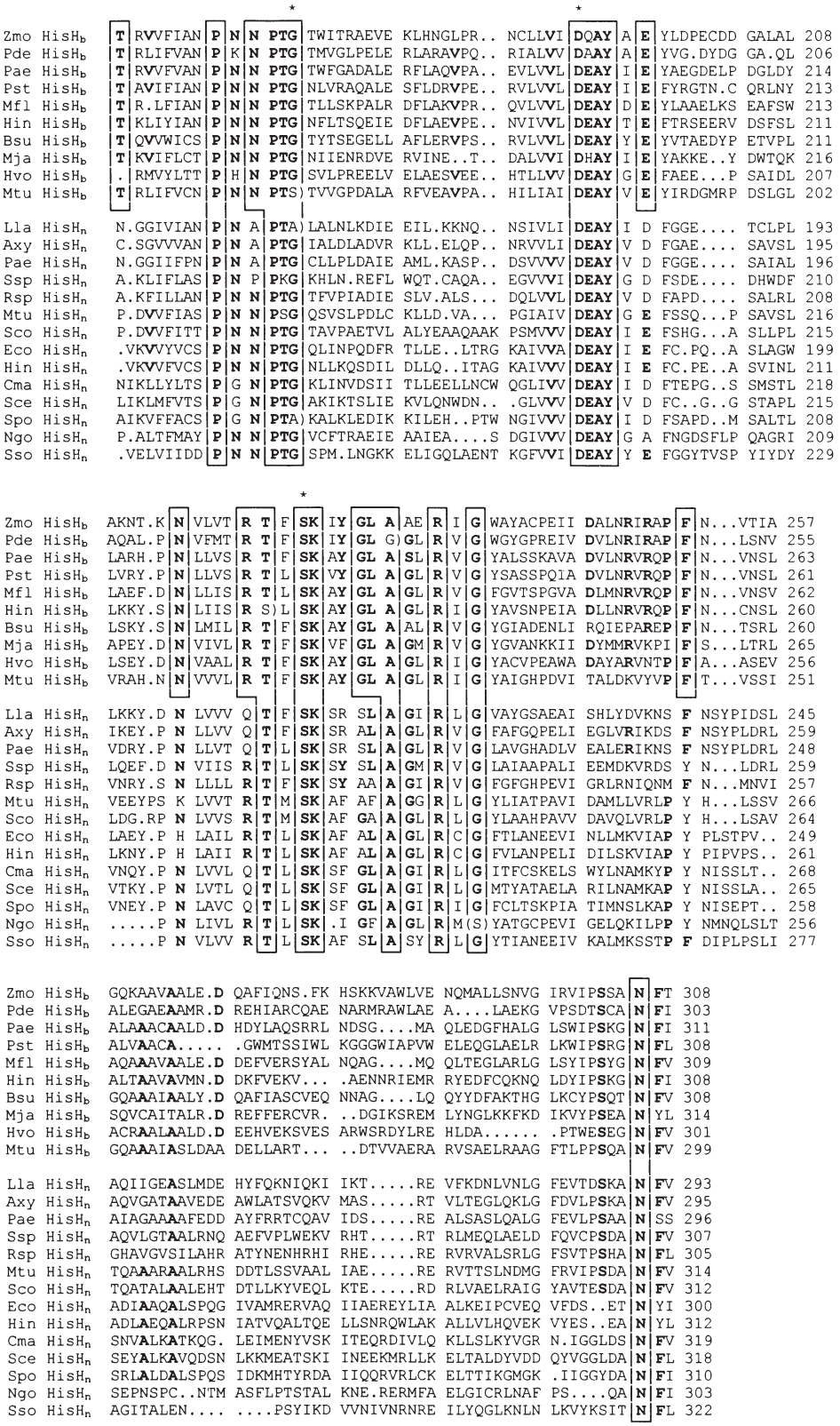


FIG. 5. (*Continued on next page*)

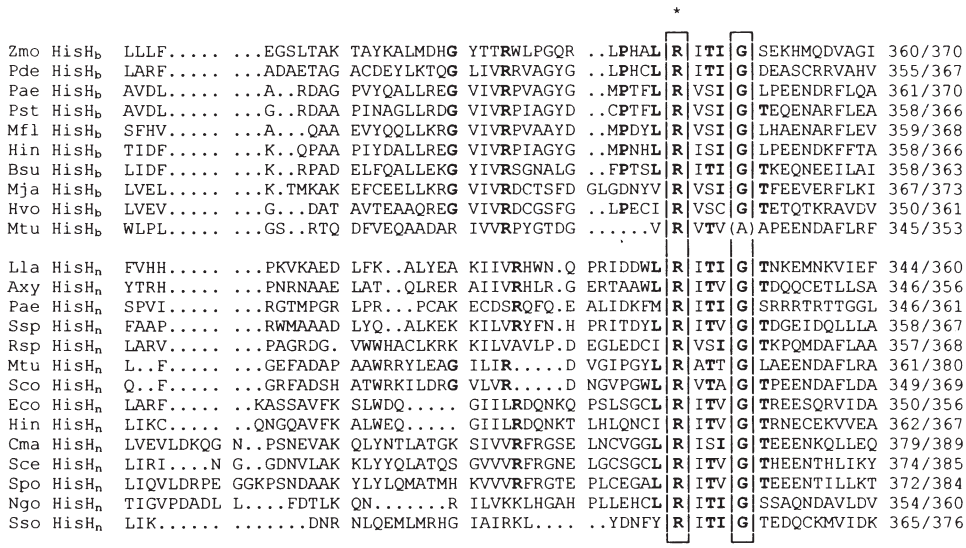


FIG. 5. (*Continued*)

The aroF *gene family*

The gene encoding EPSP synthase (AroF) has been studied from many microbial and plant organisms because it is a sensitive target for the highly effective herbicidal and antimicrobial agent, glyphosate. No comparative analysis of this highly conserved gene is given in this report because of the quantity and quality of information already in the literature.

cmk *and* rpsA *gene families*

Cytidylate kinase (Cmk) has been recognized at the molecular-genetic level only recently and has been characterized most extensively in *B. subtilis* (Schultz et al., 1997) and *E. coli* (Fricke et al., 1995). *rpsA* encodes the 30S ribosomal protein S1. Comprehensive analyses of these gene families are not included here because the overview emphasis is restricted to genes that contribute directly or indirectly to aromatic amino acid biosynthesis.

EMERGING PATTERNS OF GENE ORDER CONSERVATION IN EUBACTERIA

The presence of some of the same linked genes (i.e., *hisH$_b$-tyrA$_c$-aroF*) in *P. aeruginosa/P. stutzeri* as found in *B. subtilis*, an established (Henner and Yanofsky, 1993) supraoperon system, suggested to us that these genes might be conserved across surprisingly wide phylogenetic boundaries. We have analyzed the database to search for evidence of the occurrence of these genes together in other organisms. Figure 6 reveals three gene clusters that generally persist throughout both the gram-positive and gram-negative assemblages of bacteria: *aroE-aroB*, *hisH$_b$-tyrA-aroF*, and *cmk-rpsA*. For *Paracoccus* and *Zymomonas*, one might predict that future sequencing will show *tyrA* to be immediately followed by *aroF*. *Burkholderia* can be expected to exhibit a gene organization similar to that of *Bordetella*, but some gene shuffling comparable to that seen in comparison of *Escherichia* and *Yersinia* would not be surprising. In *Burkholderia*, *tyrA* and *aroF* overlap in different reading frames by 88 nucleotides.

Given the arrangements shown in Figure 6, a reasonable working hypothesis is that the ancestral gene arrangement for the organisms shown included linkage of *aroE-aroB*, *hisH$_b$-tyrA-aroF*, and *cmk-rpsA*. Linkage of these genes exists in contemporary lineages of both gram-positive and gram-negative bacteria.

Gram-positive bacteria

In the gram-positive grouping, *Staphylococcus* maintains *aroG*, *aroB*, and *aroF* as a closely linked gene triad in the same relative order as the corresponding genes in the *B. subtilis* supraoperon, but the entire nine

TABLE 2. SEQUENCE QUERIES USED

GenBank ID	*Swiss prot. ID*	*Description*
114181	P07639	*Escherichia coli* AroB
728898	P24167	*Escherichia coli* AroE
2506201	P07638	*Escherichia coli* AroF
114183	P12008	*Escherichia coli* AroG
464976	Q04983	*Zymomonas mobilis* $TyrA_c$
130048	P21203	*Bacillus subtilis* PheA
2506180	P23721	*Escherichia coli* SerC
2506790	P23863	*Escherichia coli* Cmk
2507321	P02349	*Escherichia coli* RpsA

genes that separate *aroB* and *aroF* in *B. subtilis* are absent from this region in *Staphylococcus*. It seems likely that *aroH-trpABDCEF* arose in *Bacillus* as an insertion between *aroB* and $hisH_b$-*tyrA-aroF*, as the *Bacillus* arrangement is thus far unique. In *Lactococcus tyrA* is followed by a putative terminator, and *aroF* is translationally coupled with *aroE*, which, in turn, is separated from *pheA* by a single nucleotide (Griffin and Gasson, 1995). *Streptococcus* and *Enterococcus* exhibit a similar gene arrangement. Neither *aroE* nor *pheA* is linked to *aroF* in *B. subtilis* or *Staphylococcus aureus*. *Mycobacterium* lacks linkage of *tyrA* and *aroF*, as well as of *cmk* and *rpsA*. *Clostridium* exhibits some gene shuffling, including the presence of *aroD* between *aroG* and *aroE*, the insertion of *aroB* between *tyrA* and *aroF* instead of between *aroE* and $aroC_{II}$, and the insertion of *lytB* between *cmk* and *rpsA*. The *cmk-lytB-rpsA* gene order is also present in the deeply branching *Thermotoga* lineage, and in fact, the gene arrangement seen in *Thermotoga aroG-aroE-aroB-*$aroC_{II}$ seems to be generally characteristic of gram-positive bacteria.

Formation of a new supraoperon combination in gram-negative bacteria

Members of the β and γ subdivisions of Proteobacteria possess the bifunctional $AroQ_p$•PheA (Ahmad and Jensen, 1988a). In the *Pseudomonas* lineage, the gene fusion event that presumably created $aroQ_p$•*pheA*

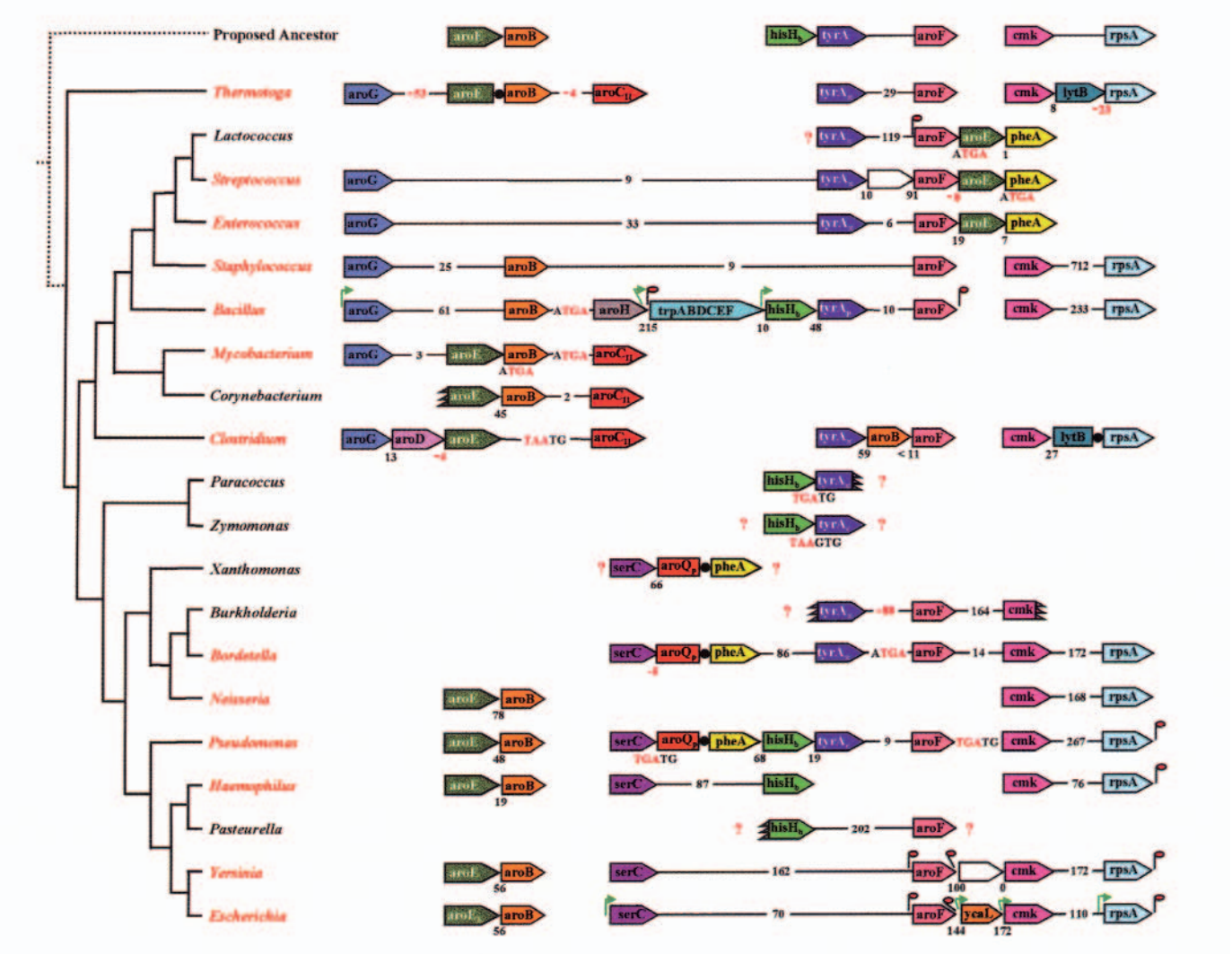
FIG. 6. Conserved gene organization in the domain Bacteria. Organisms having the gene organizations shown are placed on a dendrogram (at left) derived from 16S rRNA sequence comparisons. Organisms whose entire genome has been sequenced are shown in orange. Species names belonging to the genera shown are *Thermotoga maritima, Lactococcus lactis, Streptococcus pneumoniae, Enterococcus faecalis, Staphylococcus aureus, Bacillus subtilis, Mycobacterium tuberculosis, Corynebacterium pseudotuberculosis, Clostridium acetobutylicum, Paracoccus denitrificans, Zymomonas mobilis, Xanthomonas campestris, Burkholderia pseudomallei, Bordetella pertussis, Neisseria gonorrhoeae, Pseudomonas aeruginosa, Haemophilus influenzae, Pasteurella multocida, Yersinia enterocolitica*, and *Escherichia coli*. Homologous genes are color-coded. An open box indicates an unidentified open reading frame. Genes connected by a bar are adjacent. Incomplete gene sequences are indicated with ragged edges. Flanking regions marked with question marks might contain genes of interest, but these regions have not yet been sequenced. Intergenic distances are not shown proportionally in order to facilitate a visual comparison, but intervening base pair numbers are indicated. A bullet between genes indicates a gene fusion (see Table 1). Translational coupling via overlapping stop and start codons is indicated by showing these codons (stop codon in red). More extensive gene overlap is shown in red (e.g., −8 indicates that 8 nt are shared by flanking genes). The intergenic space between *aroB* and *aroF* in *Clostridium* (<11) was estimated by multiple alignment comparisons because a start codon for *aroF* was uncertain. Transcriptional start points (green arrows) and transcriptional terminators (red flags) are shown for the well-documented *B. subtilis* and *E. coli* systems. The locations of other strong rho-independent terminator structures are shown, but no attempt has been made to identify promoter regions. Following the preliminary indications of conserved gene order in the developing supraoperons of *B. subtilis* and *P. aeruginosa*, the sequences listed in Table 2 were used as queries against the Unfinished Microbial Genomes Blast Database at NCBI. The results were processed with the LOCATE Program. This program (available on request from T. Brettin) uses the output from **tblastn** to locate sequences within a contig that are similar to the query. It uses a blast e-value threshold of $p < 0.001$. The output is easily scanned by eye. For clarity of visual presentation, some aspects of gene linkage are abbreviated. Thus, in *Thermotoga*, the full gene linkage is *tyrA-aroF-aroD-aroG-aroE*•*aroB-*$aroC_{II}$. In *Enterococcus* and *Streptococcus*, *aroB* (not shown) precedes *aroG*. In *Clostridium*, the full gene linkage is *tyrA-aroB-aroF-aroG-aroD-aroE-*$aroC_{II}$.

is herein hypothesized to have established its linkage upstream with *serC* and its linkage downstream with *hisH$_b$-tyrA-aroF-cmk-rpsA* to yield the *serC. . . . rpsA* supraoperon organization of genes. It should be most instructive to examine the full gene organization in members of the α division of Proteobacteria, for example, *Z. mobilis*, where *aroQ* and *pheA* are unfused.

Helicobacter pylori, a member of the Proteobacteria whose genome has been completely sequenced, does not exhibit any portion of the *serC. . . . .rpsA* supraoperon organization found in other Proteobacteria. Genome constriction has led to loss of *pheA* and *hisH$_b$* (indeed, the entire histidine pathway). Such losses would, of course, disrupt the prior gene organization. An uncertain homolog of SerC is highly divergent from the SerC protein family and probably uses hydroxypyruvate rather than phosphohydroxypyruvate as substrate. *Helicobacter* may use cyclohexadienyl dehydratase (PheC) for phenylalanine biosynthesis, as it possesses several weak homologues of PheC, a nonhomologue of the prephenate-specific PheA proteins (Zhao et al., 1993).

Disruption of supraoperon organization

Gene duplications and gene fusions occur by mechanisms that can be expected to alter gene order. As outlined, the *aroQ$_p$•pheA* gene fusion (event A in Fig. 7) may have joined a newly organized *serC-aroQ$_p$•pheA* grouping to a previously existing *hisH$_b$-tyrA-aroF-cmk-rpsA* linkage.



The *Haemophilus*-enteric lineage represents a relatively small cluster within the γ assemblage of Proteobacteria where yet a second gene fusion may have disrupted the existing supraoperon organization. This relatively recent gene fusion arose in an ancestral position (B in Fig. 7) (Ahmad and Jensen, 1988b). The $aroQ_t$ domain (chorismate mutase) of the newly evolved bifunctional $aroQ_t$•*tyrA* presumably arose from $aroQ_p$• by gene duplication followed by fusion to the previously monofunctional *tyrA* (Xia and Jensen, 1992).

In enteric bacteria, a dynamic series of additional evolutionary events altered aromatic amino acid biosynthesis and regulation. Two additional paralogs of the gene encoding DAHP synthase were generated to give the three-isoenzyme assemblage found only in enteric bacteria (Ahmad et al., 1987). Each evolved a differential sensitivity to feedback inhibition by one of the three aromatic amino acids. One of these genes, denoted $aroA_y$, was joined in operonic linkage with $aroQ_t$•*tyrA* (event C in Fig. 7). This latter group was moved adjacent to $aroQ_p$•*pheA*, but in convergent orientation on the opposite strand (Fig. 7). Perhaps the latter event directly resulted in the disruption of $hisH_b$, which is absent in *E. coli*. In *H. influenzae,* $hisH_b$ is near but not adjacent to $aroQ_p$•*pheA* on the opposite strand (Fig. 7). To compensate for loss of $hisH_b$ for aromatic aminotransferase function, *E. coli* generated *tyrB*, a close paralog of *aspC* (Jensen and Gu, 1996).

Reconstruction of the exact changes in gene organization that occurred in transition from supraoperon organization exemplified by *P. aeruginosa* to that of *H. influenzae* to that of *E. coli* is an intriguing prospect to anticipate. Too many evolutionary events separate these organisms to devise a credible scenario at present. However, when comparable genomic information becomes available for a suitably spaced phylogenetic progression of organisms between the current ones, there are realistic prospects that the individual steps of gene reordering taken during evolution can be deduced.

THE *aroE-aroB* SUPRAOPERON

In gram-negative bacteria, the *aroE-aroB* linkage is consistently observed, and this extends to some of the gram-positive bacteria. Interestingly, *aroG* and *aroB* comprise the proximal genes of both *Staphylococcus* and the aforementioned *B. subtilis* supraoperon, but *aroE* is absent between *aroG* and *aroB*.

Shikimate kinase homologues

Shikimate kinase is encoded by two paralog genes in *E. coli*, and these have been denoted *aroK* and *aroL* in the literature. The gene adjacent to *aroB* is *aroK* (our designation, $aroE_k$). The gene product of $AroE_K$ has been postulated to have some function other than shikimate kinase in vivo, as it has a 100-fold lower affinity for shikimate than does $AroE_L$ (Pittard, 1996). Indeed, the loss of $AroE_K$ has been found to confer resistance to mecillinam and has been postulated to function in the regulation of cell division, perhaps by phosphorylating a cell division protein (Vinella et al., 1996). Thus, the *aroE-aroB* tandem appears to function in multiple pathways (a mixed-function entity). Surprisingly, *H. influenzae* does not possess the *aroL* paralog, and therefore in *H. influenzae*, $aroE_K$ is probably essential as a source of shikimate kinase in addition to the additional role just discussed. The high sequence divergence of this relatively small protein hinders any dogmatic conclusions about which of the two *E. coli* paralogs might correspond to a given *aroE* gene in another organism.

At first glance, *Helicobacter* appears to lack *aroE* upstream of *aroB* (HP0283). However, only 4 nt separate *aroB* from ORF HP0282 upstream. Recalling that *aroE* has been functionally implicated in cell division, it is suggestive that BLAST yields a weak hit for a kinesin-related protein (which is related to cell division) in the carboxy-portion of HP0282. Furthermore, *E. coli* AroE aligns with only two single gaps with the amino-terminal portion of HP0282. *Helicobacter aroB* is, in fact, the second gene in an apparent cell division operon that includes a homologue of *ftsH* (HP02286). A total of seven genes are either spaced within a few nucleotides or translationally coupled. Undoubtedly, the gene designated *aroK* (HP0157) in *H. pylori* is the gene encoding the functional shikimate kinase and may be the counterpart of the *E. coli* $aroE_L$ paralog.

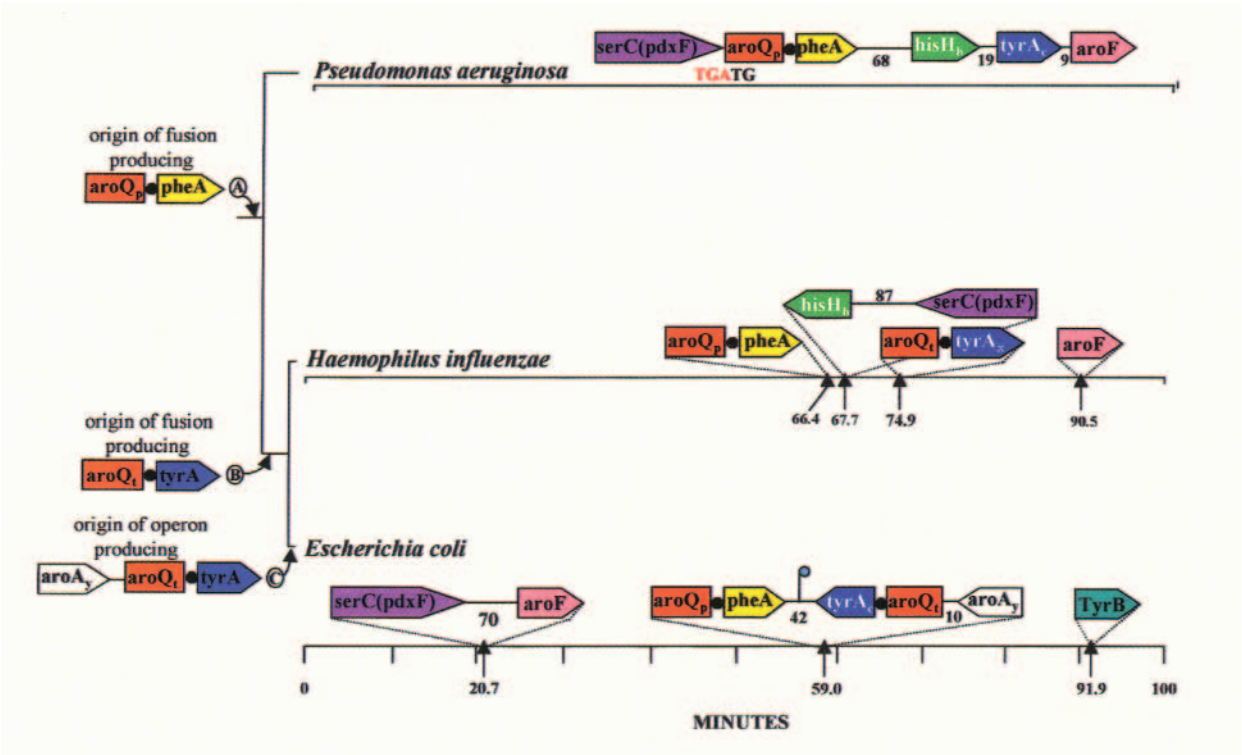


FIG. 7. Comparison of the positional placement and directional orientation of the *serC*, *aroQ$_p$*•*pheA*, *hisH$_b$*, *aroQ$_t$*•*tyrA*, and *aroF* genes on the 100-minute maps of *H. influenzae* and *E. coli*. Gene homologues are color-coded. In *E. coli*, the position of *tyrB* (which has replaced *hisH$_b$* for aromatic biosynthesis) is also shown. Genes placed on the (+) strand or on the (−) strand in the database are shown in the rightward-pointing or leftward-pointing orientation, respectively. The flag shown between •*pheA* and •*tyrA$_c$* of *E. coli* is a bidirectional terminator.

Expansion of the aroE-aroB *supraoperon in* E. coli

In *E. coli, aroE* and *aroB* are the proximal members of the large mixed-function supraoperon *aroE-aroB-urf-dam-rpe-gph-trpS* (see Introduction). Organisms in the database as close to *E. coli* as *Haemophilus* do not possess this expanded supraoperon gene organization. It will be interesting to see which relatives of *E. coli* at what hierarchical level of phylogeny exhibit the expanded supraoperon.

PERSPECTIVE

Bacterial gene order exhibits a tendency toward randomization that is perhaps surprising (Mushegian and Koonin, 1996; Watanabe et al., 1997). It has been pointed out that the infrequent instances of strong conservation of gene strings often involve demonstrated or suspected physical association of the cognate gene products (Dondekar et al., 1998; Mushegian and Koonin, 1996). In this connection, it is intriguing that *aroB* and *aroF*, which flank the distal ends of the large *B. subtilis* supraoperon, correspond to two catalytic domains within the pentafunctional AROM protein of *Aspergillus nidulans*, which have been shown to require physical interaction with one another to maintain catalytic activity (Moore and Hawkins, 1993). Overall, it appears that the multifunctional supraoperon organizations present in *B. subtilis* and *P. stutzeri/P. aeruginosa* reflect strongly conserved gene organizations across relatively wide phylogenetic distances. At the same time, there has been considerable shuffling of gene order within this overall scaffold of gene organization. The generally high degree of conservation observed may reflect global relationships of regulation that govern complex metabolic ties. Although such global relationships may be in a state of flux, some of them might have been captured and preserved in cases where gene neighbors evolved gene products that became mutually dependent on a state of physical interaction.

ACKNOWLEDGMENTS

This article is Florida Agriculture Experiment Station journal series No. R-06889. We acknowledge partial support by STD-NIH funding to G. Myers at the Los Alamos National Laboratory facilities.

REFERENCES

AHMAD, S., and JENSEN, R.A. (1988a). New prospects for deducing the evolutionary history of metabolic pathways in prokaryotes: Aromatic biosynthesis as a case-in-point. Origins Life Evol Biosphere **18,** 41–57.

AHMAD, S., and JENSEN, R.A. (1988b). The phylogenetic origin of the bifunctional tyrosine-pathway protein in the enteric lineage of bacteria. Mol Biol Evol **5,** 282–297.

AHMAD, S., JOHNSON, J.L., and JENSEN, R.A. (1987). The recent evolutionary origin of the phenylalanine-sensitive isozyme of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase in the enteric lineage of bacteria. J Mol Evol **25,** 159–167.

AHMAD, S., WILSON, A.-T., and JENSEN, R.A. (1988). Chorismate mutase: prephenate dehydratase from *Acinetobacter calcoaceticus*: Purification, properties and immunological cross-reactivity. Eur J Biochem **176,** 60–79.

ALIFANO, P., FANI, R., LIO, P., LAZCANO, A., BAZZICALUPO, M., CARLOMAGNO, M.S., and BRUNI, C.B. (1996). Histidine biosynthetic pathway and genes: Structure, regulation and evolution. Microbiol Rev **60,** 44–69.

BLANC, V., GIL, P., BAMASJACQUES, N., LORENZON, S., ZAGOREC, M., SCHLEUNIGER, J., et al. (1997). Identification and analysis of genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I. Mol Microbiol **23,** 191–202.

CRAWFORD, I.P. (1989). Evolution of a biosynthetic pathway: The tryptophan paradigm. Annu Rev Microbiol **43,** 567–600.

DE SMIT, M.H., and VAN DUIN, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. Proc Natl Acad Sci USA **87,** 7668–7672.

DICKSON, P.W., JENNINGS, I.G., and COTTON, R.G. (1994). Delineation of the catalytic core of phenylalanine hydroxylase and identification of glutamate 286 as a critical residue for pterin function. J Biol Chem **269,** 20369–20375.

DONDEKAR, T., SNEL, B., HUYNEN, M., and BORK, P. (1998). Conservation of gene order: A fingerprint of physically interacting proteins. Trends Biochem Sci **23,** 324–328.

DUNCAN, K., and COGGINS, J.R. (1986). The *serC-aroA* operon of *Escherichia coli*. Biochem J **234,** 49–57.

FISCHER, R.S., ZHAO, G., and JENSEN, R.A. (1991). Cloning, sequencing, and expression of the P-protein gene *(pheA)* of *Pseudomonas stutzeri* in *Escherichia coli*: Implications for evolutionary relationships in phenylalanine biosynthesis. J Gen Microbiol **137,** 1293–1301.

FRICKE, J., NEUHARD, J., KELLN, R.A., and PEDERSEN, S. (1995). The *cmk* gene encoding cytidine monophosphate kinase is located in the *rspA* operon and is required for normal replication rate in *Escherichia coli*. J Bacteriol **177,** 517–523.

FRIEDRICH, B., FRIEDRICH, C.G., and SCHLEGEL, H.G. (1976). Purification and properties of chorismate mutase-prephenate dehydratase and prephenate dehydrogenase from *Alcaligenes eutrophus*. J Bacteriol **126,** 712–722.

GRIFFIN, H.G., and GASSON, M.J. (1995). Genetic aspects of aromatic biosynthesis in *Lactococcus lactis*. Mol Gen Genet **245,** 119–127.

GU, W., WILLIAMS, D.S., ALDRICH, H.C., XIE, G., GABRIEL, D.W., and JENSEN, R.A. (1997). The AroQ and PheA domains of the bifunctional P-protein from *Xanthomonas campestris* in a context of genomic comparison. Microbial Comp Genomics **2,** 141–158.

GU, W., ZHAO, G.S., EDDY, C., and JENSEN, R.A. (1995). Imidazole acetol phosphate aminotransferase in *Zymomonas mobilis*: Molecular genetic, biochemical, and evolutionary analyses. J Bacteriol **177,** 1576–1584.

HENNER, D., and YANOFSKY, C. (1993). *Bacillus subtilis* and other gram-positive bacteria. In *Biochemistry, Physiology, and Molecular Genetics*. A.L. Sonenshein, J. Hoch, and R. Losick, eds. (American Society of Microbiology, Washington, DC), 269–280.

HILL, R.E., and SPENSER, I.D. (1996). Biosynthesis of vitamin B_6. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd ed. F.C. Neidhardt, R. Curtiss, II, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, et al., eds. (American Society of Microbiology, Washington, DC), **1,** 695–703.

HUDSON, G.S., and DAVIDSON, B.E. (1984). Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K-12. J Mol Biol **180,** 1023–1051.

HUFTON, S.E., JENNINGS, I.G., and COTTON, R.G. (1995). Structure and function of the aromatic amino acid hydroxylases. Biochem J **311,** 353–366.

JENSEN, R.A., and CALHOUN, D.H. (1981). Intracellular roles of microbial aminotransferases: Overlap enzymes across different biochemical pathways. Crit Rev Microbiol **8,** 229–266.

JENSEN, R.A., and GU, W. (1996). Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. J Bacteriol **178,** 2161–2171.

JUNG, E., ZAMIR, L.O., and JENSEN, R.A. (1986). Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. Proc Natl Acad Sci USA **83,** 7231–7235.

LAM, H.M., and WINKLER, M.E. (1990). Metabolic relationships between pyridoxine (vitamin B_6) and serine biosynthesis in *Escherichia coli* K-12. J Bacteriol **172,** 6518–6528.

LOHSE, D.L., and FITZPATRICK, P.F. (1993). Identification of the intersubunit binding region in rat tyrosine hydroxylase. Biochem Biophys Res Commun **197,** 1543–1548.

LYNGSTADASS, A., LØBNER-OLESEN, A., and BOYE, E. (1995). Characterization of three genes in the *dam*-containing operon of *Escherichia coli*. Mol Gen Genet **247,** 546–554.

MAN, T.K., PEASE, A.J., and WINKLER, M.E. (1997). Maximization of transcription of the *serC(pdxF)-aroA* multifunctional operon by antagonistic effects of the cyclic AMP (cAMP) receptor protein-cAMP complex and Lrp global regulators of *Escherichia coli* K-12. Int J Syst Bacteriol **47,** 132–143.

MARTIN, R.G., BERBERICH, M.A., AMES, B.N., DAVIS, W.M., GOLDBERGER, R.F., and YOURNO, J.D. (1971). Enzymes and intermediates of histidine biosynthesis in *Salmonella typhimurium*. Methods Enzymol **17B,** 3–51.

MEHTA, P.K., and CHRISTEN, P. (1993). Homology of pyridoxal-5′-phosphate-dependent aminotransferases with the *cobC* (cobalamin synthesis), *nifS* (nitrogen fixation), *pabC* (*p*-aminobenzoate synthesis) and *malY* (abolishing endogenous induction of the maltose system) gene products. Eur J Biochem **211,** 373–376.

MEHTA, P.K., HALE, T.I., and CHRISTEN, P. (1993). Aminotransferases: Demonstration of homology and division into evolutionary subgroups. Eur J Biochem **214,** 549–561.

MERRIMAN, T.R., MERRIMAN, M.E., and LAMONT, I.L. (1995). Nucleotide sequence of *pvdD*, a pyoverdine biosynthetic gene from *Pseudomonas aeruginosa*: PvdD has similarity to peptide synthetases. J Bacteriol **177,** 252–258.

MOORE, J.D., and HAWKINS, A.R. (1993). Overproduction of, and interaction within, bifunctional domains from the amino- and carboxy-termini of the pentafunctional AROM protein of *Aspergillus nidulans*. Mol Gen Genet **240,** 92–102.

MUSHEGIAN, A.R., and KOONIN, E.V. (1996). Gene order is not conserved in bacterial evolution. Trends Genet **12,** 289–290.

NELMS, J., EDWARDS, R.M., WARWICK, J., and FOTHERINGHAM, I. (1992). Novel mutations in the *pheA* gene of variants of chorismate mutase/prephenate dehydratase. Appl Environ Microbiol **5,** 2592–2598.

NESTER, E.W., and MONTOYA, A.L. (1976). An enzyme common to histidine and aromatic amino acid biosynthesis in *Bacillus subtilis*. J Bacteriol **126,** 699–705.

PIERSON, D.L., and JENSEN, R.A. (1974). Metabolic interlock: Control of an interconvertible prephenate dehydratase by hydrophobic amino acids in *Bacillus subtilis*. J Mol Biol **90,** 563–580.

PITTARD, J.P. (1996). Biosynthesis of the aromatic amino acids. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd ed. F.C. Neidhardt, J. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, eds., Vol. 1 (American Society of Microbiology, Washington, DC), 368–394.

REYNOLDS, R., BERMUDEZ-CRUZ, R.M., and CHAMBERLIN, M.J. (1992). Parameters affecting transcription termination by *E. coli* RNA polymerase. J Mol Biol **224,** 31–51.

RIEPL, R.G., and GLOVER, G.I. (1978). Purification of prephenate dehydratase from *Bacillus subtilis*. Arch Biochem Biophys **191,** 192–197.

SCHULTZ, C.P., YLISASTIGUI-PONS, L., SERINA, L., SAKAMOTO, H., MANTSCH, H.H., NEUHARD, J., et al. (1997). Structural and catalytic properties of CMP kinase from *Bacillus subtilis*: A comparative analysis with the homologous enzyme from *Escherichia coli*. Arch Biochem Biophys **340,** 144–153.

SERINO, L., REIMANN, C., BAUR, H., BEYELER, M., VISCA, P., and HAAS, D. (1995). Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*. Mol Gen Genet **249,** 217–228.

SMITH, S.C., KEMP, B.E., MCADAMT, W.J., MERCERT, J.F.B., and COTTON, R.G.H. (1984). Two apparent molecular weight forms of human and monkey phenylalanine hydroxylase are due to phosphorylation. J Biol Chem **259,** 11284–11289.

SUBRAMANIAC, P.S., XIE, G., XIA, T., and JENSEN, R.A. (1998). Substrate ambiguity of 3-deoxy-D-*manno*-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* in the context of its membership in a protein family containing a subset of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthases. J Bacteriol **180,** 119–127.

TIPPER, J., and KAUFMAN, S. (1992). Phenylalanine-induced phosphorylation and activation of rat hepatic phenylalanine hydroxylase in vivo. J Biol Chem **267,** 889–896.

TSUI, H.C., PEASE, A.J., KOCHLER, T., and WINKLER, M.E. (1994a). Detection and quantitation of RNA tran-

scribed from bacterial chromosomes. In *Methods in Molecular Genetics: Molecular Microbiology*. K. Adolph, ed. (Academic Press, New York), 179–204.

TSUI, H.C., ZHAO, G., FENG, G., LEUNG, H.C., and WINKLER, M.E. (1994b). The *mutL* repair gene of *Escherichia coli* K-12 forms a superoperon with a gene encoding a new cell-wall amidase. Mol Microbiol **11,** 189–202.

VAN DER ZEL, A., LAM, H.M., and WINKLER M.E. (1989). Extensive homology between the *Escherichia coli* K-12 SerC(PdxF) aminotransferase and a protein encoded by a progesterone induced mRNA in rabbit and human endometria. Nucleic Acids Res **17,** 8379.

VINELLA, D., GAGNY, B., JOSELEAU-PETIT, D., D'ARI, R., and CASHEL, M. (1996). Mecillinam resistance in *Escherichia coli* is conferred by loss of a second activity of the AroK protein. J Bacteriol **178,** 3818–3828.

WATANABE, H., ITOH, T., and GOJOBORI, T. (1997). Genome plasticity as a paradigm of eubacteria evolution. J Mol Biol **44** (Suppl. 1), 557–564.

WEIGENT, D.A., and NESTER, E.W. (1976). Regulation of histidinol phosphate aminotransferase synthesis by tryptophan in *Bacillus subtilis*. J Bacteriol **128,** 202–211.

WENDENBAUM, S., DEMANGE, P., DELL, A., MEYER, J.M., and ABDALLAH, M.A. (1983). The structure of pyoverdine *Pa*, the siderophore of *Pseudomonas aeruginosa*. Tetrahedron Lett **24,** 4877–4880.

WINKLER, M.E. (1996). Biosynthesis of histidine. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd ed. F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaecter, and H.E. Umbarger, eds., Vol. 1 (American Society of Microbiology, Washington, DC), 395–411.

XIA, T., and JENSEN, R.A. (1992). Monofunctional chorismate mutase from *Serratia rubideae*: A paradigm system for the three-isozyme gene family of enteric bacteria. Arch Biochem Biophys **294,** 147–153.

XIA, T., SONG, J., ZHAO, G., ALDRICH, H., and JENSEN, R.A. (1993a). The *aroQ*-encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in *Erwinia herbicola*. J Bacteriol **175,** 4729–4737.

XIA, T., ZHAO, G., and JENSEN, R.A. (1992). Loss of allosteric control but retention of the bifunctional catalytic competence of a fusion protein formed by excision of 260 base pairs from the 3′ terminus of *pheA* from *Erwinia herbicola*. Appl Environ Microbiol **58,** 2792–2798.

XIA, T.H., ZHAO, G.S., and JENSEN, R.A. (1993b). The *pheA/tyrA/aroF* region from *Erwinia herbicola*: An emerging comparative basis for analysis of gene organization and regulation in enteric bacteria. J Mol Evol **36,** 107–120.

XIE, G., BONNER, C.A., and JENSEN, R.A. (1999). A probable mixed-function supraoperon in *Pseudomonas* exhibits gene organization features of both intergenomic conservation and gene shuffling. J Mol Evol **49,** 108–121.

ZAMIR, L.O., NIKOLAKAKIS, A., BONNER, C.A., and JENSEN, R.A. (1993). Evidence for enzymatic formation of isoprephenate from isochorismate. Bioorg Med Chem Lett **7,** 1441–1446.

ZHAO, G., XIA, T., ALDRICH, H., and JENSEN, R.A. (1993). Cyclohexadienyl dehydratase from *Pseudomonas aeruginosa* is a periplasmic protein. J Gen Microbiol **139,** 807–813.

ZHAO, G.S., XIA, T.H., SONG, J., and JENSEN, R.A. (1994). *Pseudomonas aeruginosa* homologues of mammalian phenylalanine hydroxylase and 4α-carbinolamine dehydratase/DCoH as part of a three-component gene cluster. Proc Natl Acad Sci USA **91,** 1366–1370.

Address reprint requests to:
Roy A. Jensen
Department of Microbiology and Cell Science
Building 981
University of Florida
P.O. Box 110700
Gainesville, FL 32611-700

E-mail: rjensen@micro.ifas.ufl.edu