JOURNAL OF BACTERIOLOGY, Feb. 1990, p. 1051–1061
0021-9193/90/021051-11$02.00/0

Vol. 172, No. 2

Evolution of Aromatic Amino Acid Biosynthesis and Application to the Fine-Tuned Phylogenetic Positioning of Enteric Bacteria†

SUHAIL AHMAD,[1] WILLIAM G. WEISBURG,[2] AND ROY A. JENSEN[1]*

Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611,[1] and Department of Microbiology, University of Illinois, Urbana, Illinois 61801[2]

Received 8 May 1989/Accepted 20 October 1989

A comprehensive phylogenetic tree for virtually the entire assemblage of enteric bacteria is presented. Character states of aromatic amino acid biosynthesis are used as criteria, and the results are compared with partial trees based upon sequencing of 16S rRNA, 5S rRNA, and tryptophan leader peptide. Three major clusters are apparent. Enterocluster 1 possesses a gene fusion (*trpG-trpD*) encoding anthranilate synthase: anthranilate 5-phosphoribosylpyrophosphate phosphoribosyltransferase of tryptophan biosynthesis. This cluster includes the genera *Escherichia*, *Shigella*, *Citrobacter*, *Salmonella*, *Klebsiella*, and *Enterobacter*. The remaining two clusters lack the *trpG-trpD* gene fusion, but differ in the presence (enterocluster 2) or absence (enterocluster 3) of the three-step overflow pathway to L-phenylalanine. Enterocluster 2 consists of the genera *Serratia* and *Erwinia*. Enterocluster 3 includes the genera *Cedecea*, *Kluyvera*, *Edwardsiella*, *Hafnia*, *Yersinia*, *Proteus*, *Providencia*, and *Morganella*. Within these three major clusters, a tentative hierarchy of subcluster ordering is formulated on the basis of all data available. This hierarchical framework is proposed as a general working basis for continued refinement of the phylogenetic relationships of enteric bacteria.

The phylogenetic history of procaryotes, once thought to be inaccessible, is being reconstructed at an impressively rapid pace (33). Nucleotide sequencing techniques have provided the technological thrust, and rRNA has been the molecule of choice because its conservative resistance to evolutionary change allows the entire phylogenetic span of the ancient procaryotes to be analyzed. Initially, phylogenetic trees were based upon 16S rRNA oligonucleotide cataloging (18, 31), but complete sequencing of 16S rRNA is now routine (31). Sequencing of 5S rRNA is also being used (21, 27). In principle, the eventual comparative sequencing of as many cistrons as possible will yield slightly different trees, and the greater the number of trees available, the greater will be the resolution of evolutionary branching.

The enteric bacteria are of special microbiological interest because of both pathogenic and nonpathogenic relationships with mammalian systems. The intensity of clinically oriented classification has produced such a degree of subdivision that the entire family of *Enterobacteriaceae* is roughly equivalent to or even less diverse than single genera (such as *Bacillus*) elsewhere. Within such a narrow phylogenetic span as the enteric bacteria, sequencing of 16S rRNA or 5S rRNA will not necessarily generate a perfect phylogenetic dendrogram. Comparative sequence analysis of cistrons that are not as evolutionarily conservative as rRNA cistrons can help to resolve the hierarchical order of cohesive groupings such as the enteric bacteria. Indeed, partial trees based upon a number of important cistrons are already emerging (20, 25, 28, 30, 35, 36). Thus far, these studies involve too few organisms to establish an overall hierarchical order. A comparison of the phylogenetic trees constructed by 16S rRNA sequencing, 5S rRNA sequencing, and amino acid sequence homology of tryptophan leader peptides is given in Fig. 1. Although exactly the same set of organisms does not appear on all four trees, discrepancies in the branching order are apparent. For example, Fig. 1B shows *Escherichia coli* to be more closely related to *Proteus mirabilis* than it is to *Salmonella typhimurium*, a result at variance with other observations (4, 8, 22, 36, 37), and Fig. 1A shows *Serratia* to be closer to *Klebsiella* than *Klebsiella* is to *Salmonella*, a result that is at variance with those shown in Fig. 1B and C.

The comparative enzymology and regulation of aromatic biosynthesis has revealed a rich diversity of biochemical character states (12). These character states have been tabulated for the convenience of the reader throughout this text (Table 1). The evolutionary history of aromatic biosynthesis and regulation is now being deduced from comparative biochemical studies (23). Such analyses have depended upon the dendrogram framework provided by 16S rRNA sequencing data because biochemical character states alone cannot provide a basis for tree construction (for reasons summarized in reference 23). We pointed out recently that fortuitous gene fusions which have generated bifunctional proteins provide perhaps the most definitive markers known of phylogenetic branch points (1). Woese has noted (33) that biochemical diversity can be used to fine-tune hierarchical order in closely related groupings, and we have provided a specific example of the utility of this approach (10).

We have carried out an in-depth comparison of aromatic amino acid biosynthesis and regulation in most of the enteric bacteria in line with a general objective to trace the evolutionary history of this biochemical pathway. Since our data, against a background of considerable additional information in the literature, also indicated a basis for fine-tuned relationships of hierarchical branching, we present a dendrogram depiciting three major enteroclusters. Tentative subcluster arrangements are also formulated.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Citrobacter freundii* ATCC 29935, *Shigella dysenteriae* ATCC 11456a, *Salmo-*

* Corresponding author.

† Florida Agricultural Experiment Station Journal Series no. R00325.

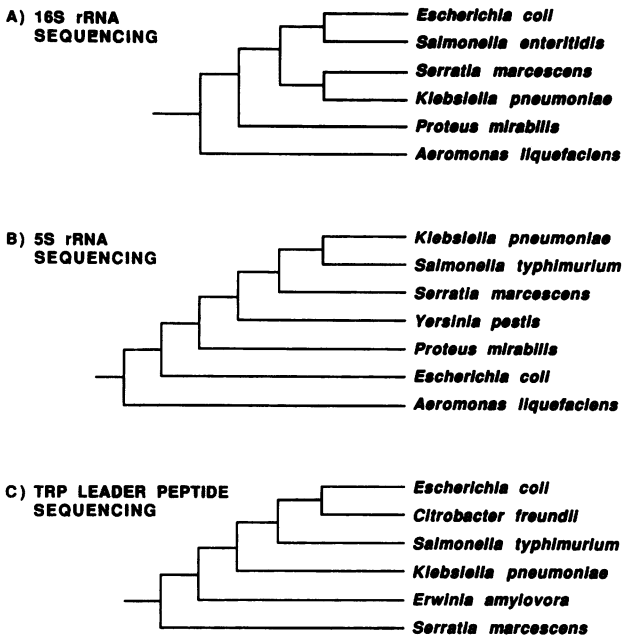

FIG. 1. Comparison of phylogenetic trees constructed for enteric bacteria by (A) 16S rRNA sequencing (W. G. Weisburg and C. R. Woese, unpublished data), (B) 5S rRNA sequencing (27), and (C) amino acid sequence homology of the *trp* leader peptide (36). The branching order, rather than the actual distances on the trees, is shown.

nella enteritidis ATCC 13076, *Enterobacter aerogenes* ATCC 13048, *Enterobacter cloacae* ATCC 13047, *Enterobacter agglomerans* ATCC 29915 (aerogenic strain), *Enterobacter agglomerans* ATCC 27155 (anaerogenic strain), *Serratia rubidaea* ATCC 27614, *Cedecae davisae* ATCC 33431, *Kluyvera ascorbata* ATCC 33433, *Hafnia alvei* ATCC 13337, *Edwardsiella tarda* ATCC 15947, *Yersinia enterocolitica* ATCC 9610, *Proteus vulgaris* ATCC 29905, *Providencia alcalifaciens* ATCC 9886, and *Morganella morganii* ATCC 25830 were obtained from the American Type Culture Collection, Rockville, Md. *Citrobacter freundii* (37°C), *Salmonella enteritidis* (37°C), *E. aerogenes* (30°C), *E. cloacae* (30°C), *E. agglomerans* (26°C), and *Serratia rubidaea* (26°C) were grown at the temperatures shown in parentheses on M9 medium as described by Winkler and Stuckman (32). *Cedecae davisae* (26°C), *Kluyvera ascorbata* (37°C), *Hafnia alvei* (30°C), *Y. enterocolitica* (30°C), *Proteus vulgaris* (37°C), and *Providencia alcalifaciens* (37°C) were grown on M9 medium supplemented with nicotinamide, *p*-aminobenzoate, D-biotin, calcium pantothenate, and thiamine (each at 1 mg/liter). *Shigella dysenteriae* (37°C), *Edwardsiella tarda* (37°C), and *M. morganii* (37°C) were grown on M9 medium supplemented with the above-mentioned vitamins and 0.1% (wt/vol) acid-hydrolyzed casein (Difco Laboratories, Detroit, Mich.). The organisms were grown to the late-exponential phase, harvested by centrifugation, washed twice with 50 mM potassium phosphate buffer (pH 7.0) containing 1 mM dithiothreitol (buffer A), and stored at −80°C until used.

Preparation of cell extracts and enzyme assays. Cell pellets were suspended in buffer A, disrupted by sonication (three 30-s bursts of ultrasound energy at 100 W), and centrifuged at 150,000 × *g* for 1 h. The resulting supernatant was passed through a Sephadex G-25 column (1.5 by 20.0 cm) equilibrated in 10 mM potassium phosphate buffer (pH 7.0) containing 1 mM dithiothreitol (buffer B) to remove small molecules. The protein-containing fractions were pooled; this was termed the crude extract.

Chorismate mutase (CM) activity was assayed by the method of Cotton and Gibson (13). The reaction mixture, in a final volume of 0.2 ml, contained buffer A, 1 mM potassium chorismate, and a suitable amount of enzyme. Incubation at 37°C was carried out for 20 min, 0.1 ml of 1 N HCl was added, and the mixture was incubated for an additional 15 min at 37°C to convert the prephenate formed to phenylpyruvate. Phenylpyruvate was measured at 320 nm after the addition of 0.7 ml of 2.5 N NaOH. A molar extinction coefficient of 17,500 was used for calculations (13).

Prephenate dehydratase (PDT) was also assayed by the method of Cotton and Gibson (13). The reaction mixture, in

TABLE 1. Useful character states in superfamily B bacteria

Bacterium	Character states									
	DAHP synthase isozymes[a]	Presence of T protein	Presence of *trpC-trpF* fusion	Presence of *trpG-trpD* fusion	Presence of both CM-F and CDT	Absence of both CM-F and CDT	Dominant DAHP synthase isozyme[a,b]	Presence of P protein	Activation of P protein by tyrosine[c]	Cofactor specificity of CDH
Enteroclusters[d]										
1	A, B, C	+	+	+	−	−	A or C	+	−	NAD
2	A, B, C	+	+	−	+	−	A	+	−	NAD
3[e]	A, B, C	+	+	−	−	+	A or C	+	−	NAD
Oceanospirillum spp.	A, B	−	?	?	−	−	A and B	+	−	NAD
Pseudomonads										
Group V	D	−	−	−	+	−	D	+	+	NAD
Group Ia	A, E	−	−	−	−	+	A	+	++	NAD
Group Ib	A, E	−	−	−	+	−	A	+	++	NAD
Acinetobacter spp.	A, F	−	−	−	−	−	A	+	+++	NADP

[a] Abbreviations: A, DS-Tyr; B, DS-Trp; C, DS-Phe; D, DS-Cha(Trp); E, DS-Trp(Cha); F, DS-0.
[b] Refers to major isozyme expressed during growth on minimal medium.
[c] Refers to the PDT component of the P protein.
[d] The box highlights character states that discriminate the three clusters of enteric bacteria.
[e] *Alteromonas* and *Aeromonas* share all of the character states listed for enterocluster 3.

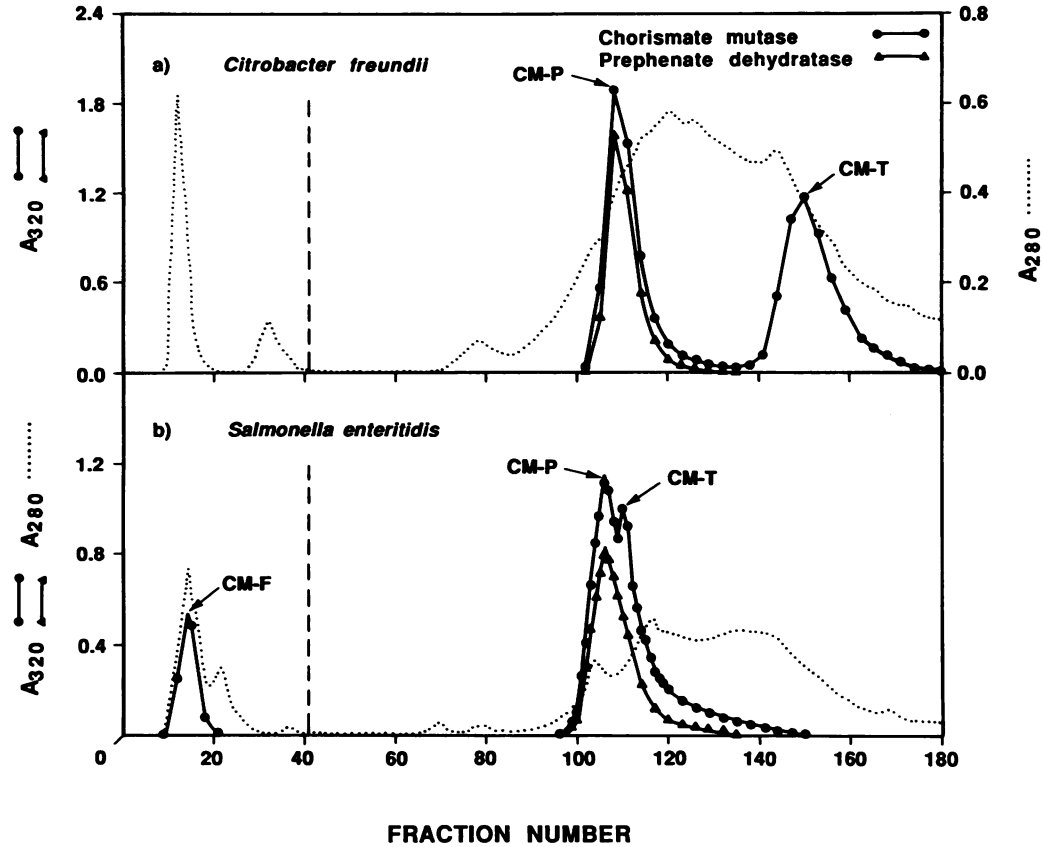

FIG. 2. Elution profiles of chorismate mutase and PDT activities following DE52 column chromatography of the crude extracts prepared from (a) *Citrobacter freundii* and (b) *Salmonella enteritidis*. The DE52 column chromatography was performed as described in Materials and Methods. The vertical dashed lines indicate the onset point of gradient elution. CM and PDT activities are expressed as phenylpyruvate A_{320} in base (an absorbance of 1.0 corresponds to 2.8 nmol of phenylpyruvate formed per min). Four times more column eluate was used for the assay of CM-F than for the assay of CM-P and CM-T in *S. enteritidis* (panel b). The distribution of proteins monitored at 280 nm is shown by dashed lines.

a final volume of 0.2 ml, contained buffer A, 1 mM potassium prephenate, and a suitable amount of enzyme. After incubation at 37°C for 20 min, 0.8 ml of 2.5 N NaOH was added and the phenylpyruvate formed was measured at 320 nm.

Prephenate dehydrogenase (PDH) and arogenate dehydrogenase activities were assayed exactly as described by Patel et al. (29) except that buffer A was used in the assay.

Arogenate dehydratase activity was assayed either by the measurement of phenylalanine formation by high-pressure liquid chromatography as described by Zamir et al. (40) or by the coupled assay described by Ahmad and Jensen (2). For the former assay, the reaction mixture, in a final volume of 0.2 ml, contained buffer A, 1 mM potassium arogenate, and a suitable amount of enzyme. After incubation at 37°C for 20 min, 0.05 ml of 0.5 N NaOH was added and the phenylalanine formed was estimated by high-pressure liquid chromatography as described by Lindroth and Mopper (26).

The reaction mixture for the coupled assay (2), in a final volume of 0.2 ml, contained buffer A, 1 mM potassium arogenate, 10 mM 2-ketoglutarate, 100 U of partially purified aromatic aminotransferase, and a suitable amount of enzyme. After incubation at 37°C for 20 min, 0.8 ml of 2.5 N NaOH was added and the phenylpyruvate formed was measured at 320 nm.

Protein in the crude extract was estimated by the method of Bradford (11) with bovine serum albumin as the standard protein.

DE52 column chromatography. Approximately 100 mg of crude extract protein was applied to a DEAE-cellulose (DE52) column (1.5 by 20.0 cm) equilibrated in buffer B. The column was washed with 2 bed volumes of the equilibration buffer, and then the bound proteins were eluted with 300 ml of a linear gradient of 0.0 to 0.35 M KCl in buffer B. Fractions of 2.2 ml were collected and were assayed for A_{280} and enzyme activities. All column fractions were initially screened with the maximal amount of eluate to detect possible low activities. Subsequent profiles were then produced by using appropriate amounts of eluate.

Biochemicals and chemicals. Amino acids, nitotinamide, *p*-aminobenzoate, calcium pantothenate, D-biotin, thiamine, Sephadex G-25, and dithiothreitol were obtained from Sigma Chemical Co., St. Louis, Mo. DE52 was purchased from Whatman, Inc., Clifton, N.J. Prephenate was prepared as the barium salt from culture supernatants of a tyrosine auxotroph of *Salmonella typhimurium* (16) and was converted to the potassium salt before use. Chorismate was isolated from the accumulation medium of a triple auxotroph of *Klebsiella pneumoniae* 62-1 and purified as the free acid (19). L-Arogenate was isolated from a triple auxotroph of *Neurospora crassa* ATCC 36373 by the method of

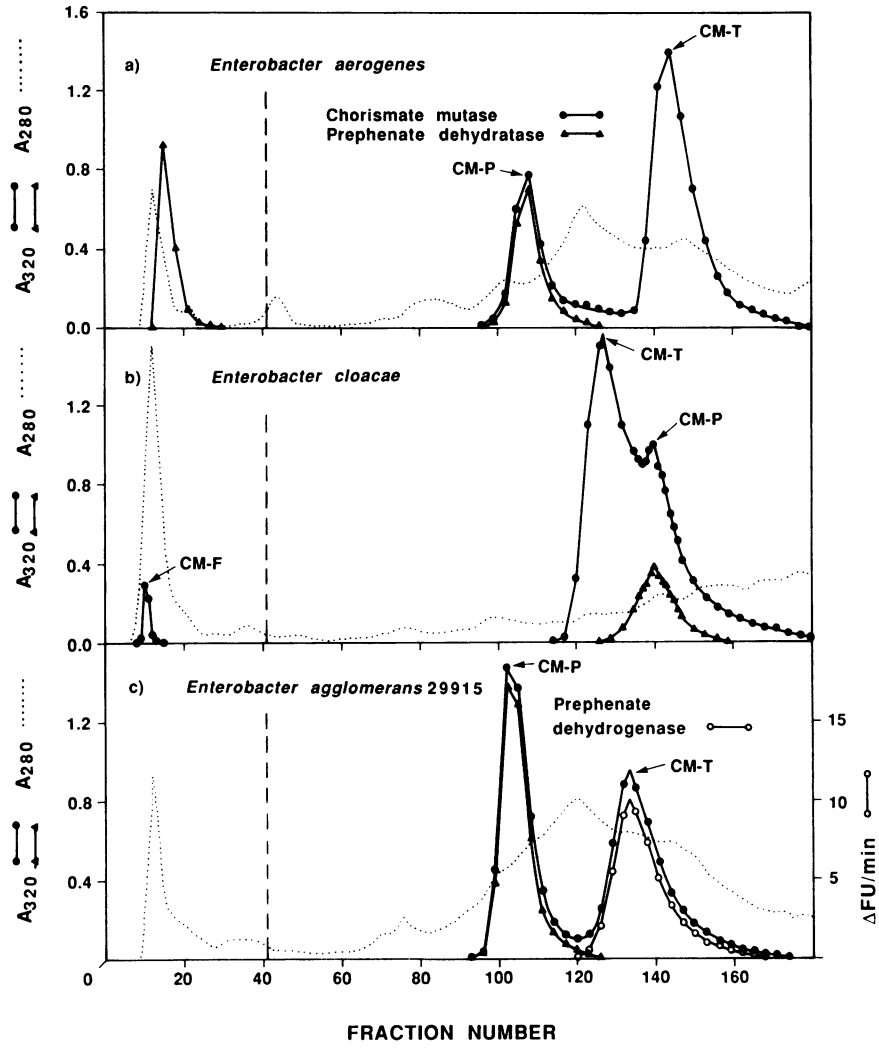

FIG. 3. Elution profiles of CM and PDT activities following DE52 column chromatography of the crude extracts prepared from (a) *Enterobacter aerogenes*, (b) *Enterobacter cloacae*, and (c) *Enterobacter agglomerans* 29915. Three times more column eluate was used for the assay of CM-F than for the assay of CM-P and CM-T in *E. cloacae* (panel b). As an exception to the conditions specified in Materials and Methods, a linear gradient of 0.0 to 0.22 M KCl was used for elution of bound proteins to obtain the data shown in panel b. The coincident elution of CM-T with CDH (assayed with prephenate as the substrate) is shown for *E. agglomerans* 29915 (panel c). The PDH activity is expressed as change in fluorescence units per minute (ΔFU/min). A change of 100 FU/min corresponds to 1.9 nmol of NADH formed per min. See the legend to Fig. 2 for an explanation of symbols and other details.

Zamir et al. (38). All other chemicals were standard reagent grade.

RESULTS

Aromatic pathway enzymes in *Citrobacter*, *Shigella*, and *Salmonella* species. Two isoenzymes of CM eluted from *Citrobacter freundii* following DE52 chromatography of the crude extract (Fig. 2a). The leading peak of activity coeluted with the PDT peak of activity, thus marking the presence of the bifunctional P protein (chorismate mutase-P:prephenate dehydratase [CM-P:PDT]), whereas the trailing peak coeluted with the cyclohexadienyl dehydrogenase (CDH) peak of activity, thus marking the presence of the bifunctional T protein (chorismate mutase-T:cyclohexadienyl dehydrogenase [CM-T:CDH]) of tyrosine biosynthesis. Although the CDH profiles are not shown here, the exact coincidence of CDH and CM-T elution profiles has been shown earlier (6, 7). Similar results were obtained from *Shigella dysenteriae* (data not shown), *Escherichia coli* (3, 13), and *Salmonella enteritidis*. However, *S. enteritidis* differed in its possession of a third species of chorismate mutase (CM-F) activity, not associated with any other activity of aromatic biosynthesis, which eluted in the wash fractions (Fig. 2b). The presence of CM-F in *Salmonella typhimurium* has previously been demonstrated (5).

Aromatic pathway enzymes in *Enterobacter* species. Two species of CM activity also eluted from *Enterobacter aerogenes* (Fig. 3a), one species coeluting with the PDT peak of activity (P protein) and the second coeluting with the CDH peak of activity (T protein) (dehydrogenase profile not shown). However, an additional peak of PDT activity eluted in the wash fractions. Although this enzyme did not use L-arogenate as an alternative substrate, it did utilize prephenyllactate, another structural analog of prephenate (39), as an alternative substrate. This variety of cyclohexadienyl

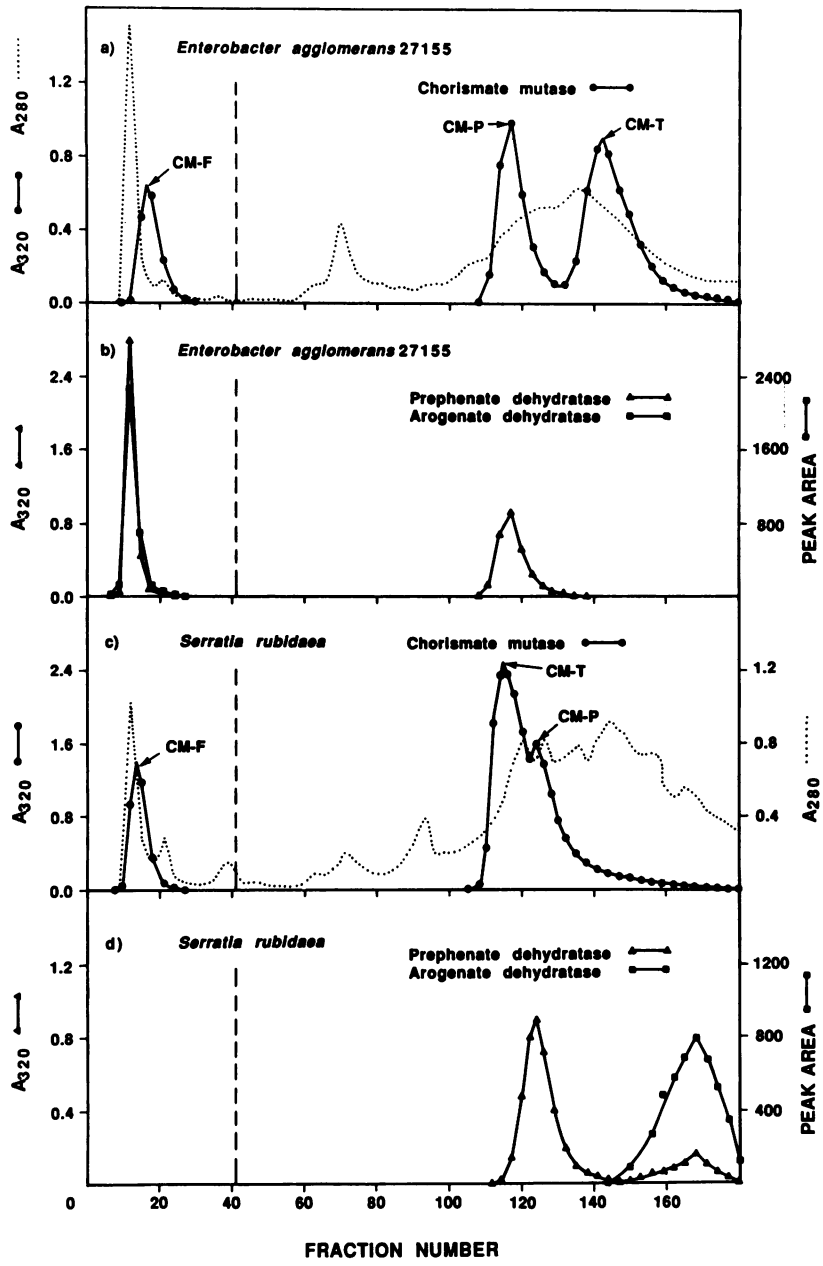

FIG. 4. Elution profiles from DE52 columns of CM activities from *Enterobacter agglomerans* 27155 (a) and *Serratia rubidaea* (c) and of PDT and arogenate dehydratase activities from *E. agglomerans* 27155 (b) and *S. rubidaea* (d) following chromatography of the crude extracts. Four times more column eluate was used for the assay of CM-F than for the assay of CM-P and CM-T in *E. agglomerans* 27155 (panel a). Arogenate dehydratase activity is expressed as peak area (a peak area of 100 corresponds to 0.14 nmol of L-phenylalanine formed per min). See the legend to Fig. 2 for an explanation of symbols and other details.

dehydratase (CDT) activity was first described for *Klebsiella pneumoniae* (5). The presence of P protein and T protein activities has also been found in *K. pneumoniae* (5, 13).

Three peaks of CM activity eluted from *Enterobacter cloacae*, the activity eluting in the wash fractions not being associated with any other activity of aromatic biosynthesis (CM-F). The activity peaks eluting in the gradient fractions proved to be CM-T and CM-P, respectively, based on their coelution with CDH and PDT peaks of activity (dehydrogenase profile not shown) (Fig. 3b).

Since *E. agglomerans* strains are classified into two major groupings (aerogenic and anaerogenic) which differ considerably from each other (8), a member from each group was included in the present study. The results obtained from *E. agglomerans* 29915 (aerogenic strain) are shown in Fig. 3c. Only two peaks of CM activity eluted, these being CM-P and CM-T on the basis of their coelution with PDT and CDH activities (CDH activity with prephenate as substrate is shown), respectively (Fig. 3c). No additional peak of mutase activity or dehydratase activity (with either prephenate or L-arogenate as substrate) was detected.

The results obtained from *E. agglomerans* 27155 (anaerogenic strain) were quite different. Three peaks of CM activity eluted (Fig. 4a). The peak of activity eluting in the wash

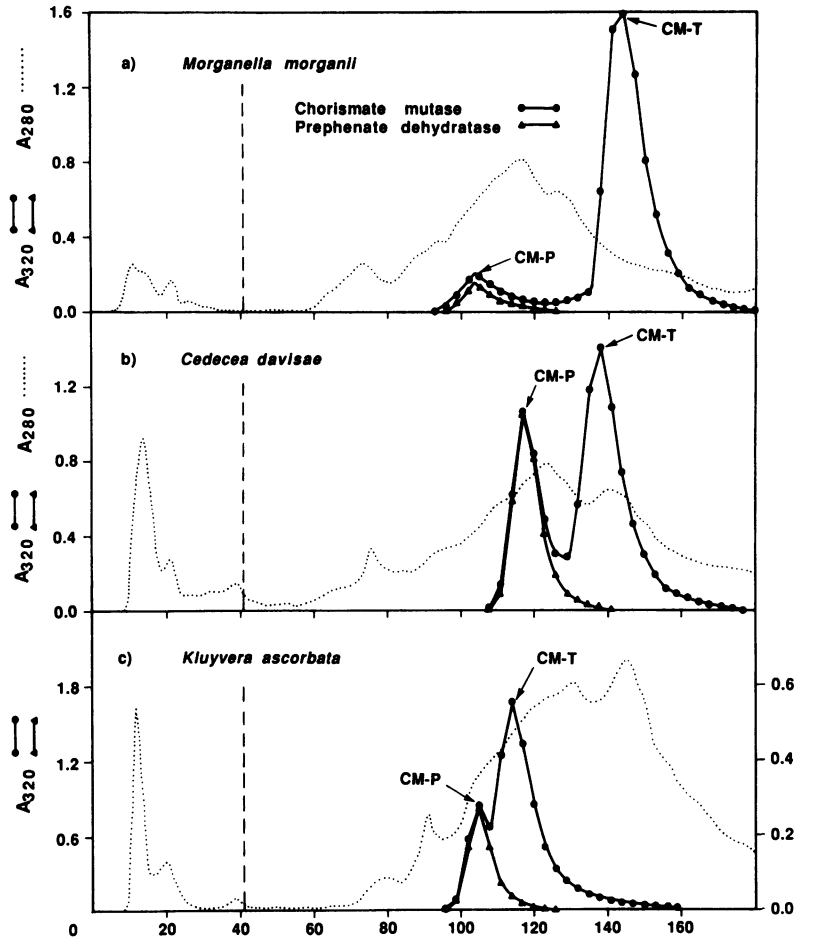

FIG. 5. Elution profiles of CM and PDT activities following DE52 column chromatography of the crude extracts prepared from (a) *Morganella morganii*, (b) *Cedecea davisae*, and (c) *Kluyvera ascorbata*. A linear gradient of 0.0 to 0.45 KCl was used to obtain the profile shown in panel a. See the legend to Fig. 2 for an explanation of symbols and other details.

fractions was not associated with any other activity of aromatic biosynthesis (CM-F), whereas the activity peaks eluting in the gradient fractions proved to be CM-P and CM-T components of the bifunctional P-protein and T protein, judging from their coincident elution with the PDT (Fig. 4b) and CDH peaks of activity (dehydrogenase profile not shown), respectively. However, an additional peak of PDT activity eluted in the wash fractions (Fig. 4b), which could utilize L-arogenate as an alternative substrate (in contrast to the PDT activity of the bifunctional P protein), thus showing the presence of CDT in this organism. The CDT enzyme from *E. agglomerans* 27155, *Serratia* species, and *Erwinia* species all could utilize prephenyllactate as an alternative substrate. The presence of both CM-F and CDT in *E. agglomerans* 27155 indicates that an operational overflow pathway for phenylalanine biosynthesis is present in this organism.

Aromatic pathway enzymes in *Serratia* species. The results obtained from *Serratia rubidaea* were similar to the results obtained from *E. agglomerans* 27155 (three isozymes of CM exist: CM-F, CM-T and CM-P; a PDT enzyme and a CDT enzyme exist). A minor difference was that CDT eluted late in the gradient fractions (Fig. 4c and d). The presence of a complete overflow pathway in *Serratia marcescens* and in *Erwinia* species (except for *E. carotovora*, which lacks CDT) has already been reported (5).

Aromatic pathway enzymes in *Morganella*, *Cedecea*, *Kluyvera*, *Hafnia*, *Edwardsiella*, *Yersinia*, *Proteus*, and *Providencia* species. Only two peaks of CM activity eluted from *Morganella morganii*, *Cedecea davisae*, *Kluyvera ascorbata* (Fig. 5a, b, and c, respectively) and from *Hafnia alvei*, *Edwardsiella tarda*, and *Yersinia enterocolitica* (Fig. 6a, b, and c, respectively). In each case, one peak of activity coeluted with the PDT peak of activity (P protein), whereas the second activity peak coeluted with the CDH peak of activity (T protein) (dehydrogenase profile not shown). No other dehydratase activity was recovered in column fractions when either prephenate or L-arogenate was used as an alternative substrate. No arogenate dehydratase activity was detected in crude extracts prepared from these organisms. Similar results were obtained from *Proteus vulgaris* and *Providencia alcalifaciens* (data not shown) and have already been reported for *Proteus mirabilis* (5).

Regulation of mutase and dehydratase activities. For each organism studied in this work, the CM component of the bifunctional P protein (CM-P) was always inhibited by L-phenylalanine and showed no sensitivity toward other aromatic amino acids (Table 2). On the other hand, CM-T

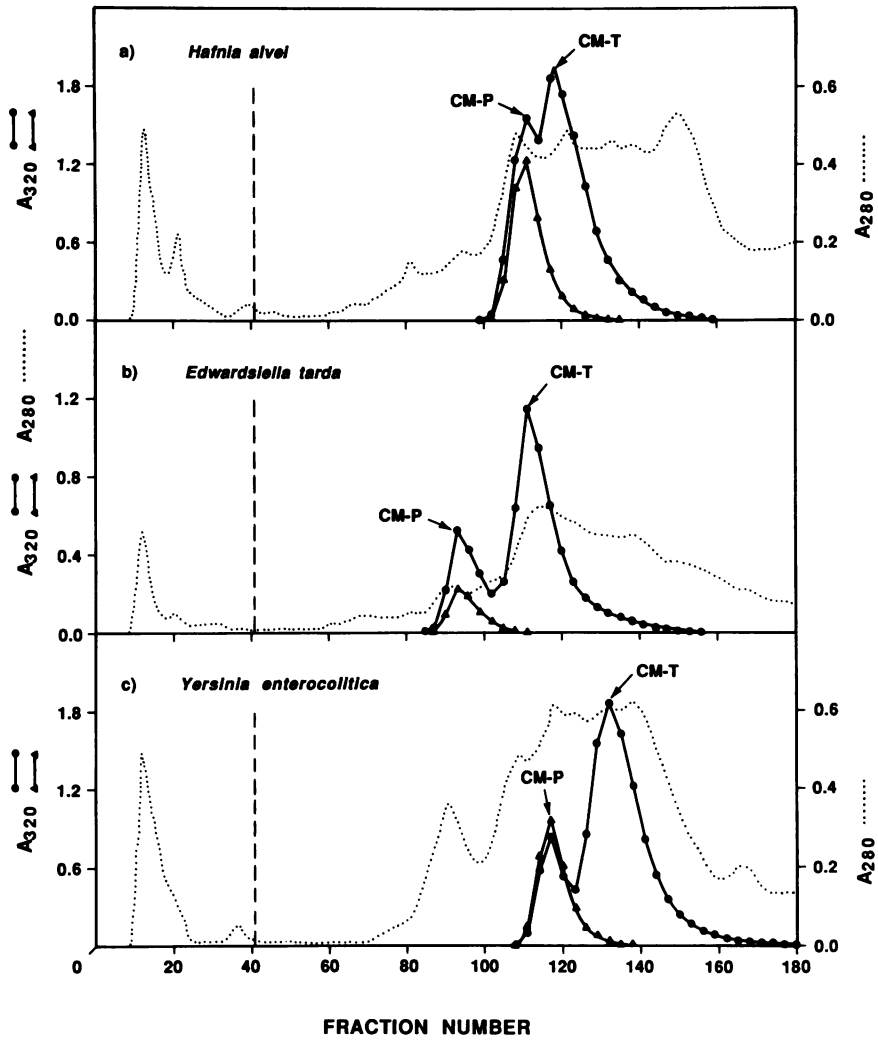

FIG. 6. Elution profiles of CM and PDT activities following DE52 column chromatography of the crude extracts prepared from (a) *Hafnia alvei*, (b) *Edwardsiella tarda*, and (c) *Yersinia enterocolitica*. See the legend to Fig. 2 for an explanation of symbols and other details.

(part of the bifunctional T protein) and CM-F (when present) were completely insensitive to any of the aromatic amino acids present alone or in combination (data not shown).

The PDT component of the bifunctional P protein (PDT-P) was tightly regulated by L-phenylalanine in all the organisms studied in this paper (Table 3). However, in *Enterobacter* species (except anaerogenic *E. agglomerans*), *M. morganii* and *Cedecea davisae*, PDT-P was additionally inhibited by L-tyrosine and L-tryptophan, the inhibition being cumulative (Table 3). This curious pattern of inhibition was first noticed in *Erwinia* species (5). The CDT activity, when present, was not inhibited by any of the aromatic amino acids, whether assayed with prephenate or L-arogenate as the substrate (data not shown).

DISCUSSION

Biochemical character states as evolutionary milestones. Phenotypic characteristics that can be assigned a plus-or-minus state for the purposes of drawing evolutionary conclusions are termed character states. The multibranched pathway for biosynthesis of aromatic amino acids has provided a wide diversity of useful character states (23). The enteric lineage shown in Fig. 7 is only a small section of superfamily B, which, in turn, is one of three subdivisions of the gram-negative purple bacteria, commonly termed the gamma subdivision (33). Of the ancestral character states shared by the enteric lineage and shown at the left of Fig. 7, only the bifunctional P protein is a constant character state present throughout the entire superfamily B (indeed, it is found beyond this family also [1]). The bifunctional T protein of tyrosine biosynthesis (6, 7; Ahmad, unpublished data), the *trpC-trpF* gene fusion (15), and the presence (9) of three regulatory isozymes of 3-deoxy-D-arabinoheptulosonic acid 7-phosphate (DAHP) synthase (DS) (DS-Tyr, DS-Phe, and DS-Trp) are character states that persist without exception throughout the enteric lineage. These stable character states clearly mark the enteric lineage, since they are absent elsewhere within superfamily B.

Other character states deduced to be ancestral for the enteric lineage exhibit diversity within the lineage. These are (i) the expression of DS-Tyr as the major fractional isozyme activity during growth in minimal medium [DS-Tyr(dom)], (ii) the absence of both CM-F and CDT, and (iii) the existence of unfused cistrons encoding *trpG* and *trpD*. The

TABLE 2. Inhibition of CM isozymes[a] by aromatic amino acids

Organism	% Inhibition of CM-P by:			
	L-Phenylalanine	L-Tyrosine	L-Tryptophan	Aromatic amino acids[b]
Citrobacter freundii	16	0	0	15
Salmonella enteritidis	20	0	0	21
Enterobacter aerogenes	30	0	1	30
Enterobacter cloacae	16	0	0	15
Enterobacter agglomerans 29915	11	0	0	10
Enterobacter agglomerans 27155	9	0	0	9
Serratia rubidaea	31	0	0	29
Morganella morganii	40	0	0	40
Cedecea davisae	28	2	0	28
Kluyvera ascorbata	43	0	1	43
Hafnia alvei	18	1	0	18
Edwardsiella tarda	29	0	0	30
Yersinia enterocolitica	52	3	2	52

[a] CM isozymes were recovered from DE52 column fractions. Neither CM-F (when present) nor CM-T was inhibited by any of the aromatic amino acids or combinations thereof (data not shown).
[b] The final concentration of aromatic amino acids added individually or in combination (Aro) was 0.5 mM.

TABLE 3. Inhibition of PDT-P[a] by aromatic amino acids

Organism	% Inhibition of PDT-P by:							
	Phe (0.1 mM)	Tyr (0.1 mM)	Trp (0.1 mM)	Aro[b] (0.1 mM)	Phe (0.5 mM)	Tyr (0.5 mM)	Trp (0.5 mM)	Aro (0.5 mM)
Citrobacter freundii	35				86	0	0	84
Salmonella enteritidis	15	1	0	15	68	1	4	68
Enterobacter aerogenes	72	22	38	78	74	44	65	78
Enterobacter cloacae	62	9	11	71	87	32	47	90
Enterobacter agglomerans 29915	71	31	30	85	93	60	66	94
Enterobacter agglomerans 27155	12	0	0	12	54	1	0	55
Serratia rubidaea	43	2	1	41	89	4	3	87
Morganella morganii	70	18	50	78	81	33	75	85
Cedecea davisae	56	31	41	62	87	47	64	89
Kluyvera ascorbata	51	0		51	86	0	0	85
Hafnia alvei	51	0		50	91	0	4	90
Edwardsiella tarda	84	0	0	84	98	0	0	96
Yersinia enterocolitica	70	0	0	68	94	2	3	93

[a] PDT-P activities were recovered from DE52 column fractions. CDT activity was not inhibited by any aromatic amino acid with either prephenate or L-arogenate as the substrate (data not shown).
[b] Aro, Combination of L-phenylalanine, L-tyrosine, and L-tryptophan.

presence of CM-F and CDT (overflow pathway to phenylalanine) is generally characteristic of superfamily B (23), and cryptic inactivation of these enzymes appears to have occurred shortly after the divergence of the enteric lineage from the rest of superfamily B (4).

Bifunctional proteins as nested markers of phylogenetic clusters. Gene fusions that produce bifunctional proteins are relatively infrequent evolutionary events, which appear to be exceedingly stable. They are ideal markers of phylogenetic clustering at different hierarchical levels. Thus, the bifunctional P protein marks superfamilies A and B (1), the bifunctional T protein and the *trpC-trpF* fusion mark the enteric lineage, and the bifunctional *trpG-trpD* fusion marks enterocluster 1 (Fig. 7). It will be interesting to determine the hierarchical distribution of other known gene fusions (e.g., two in the pathway of histidine biosynthesis and the two aspartokinase/homoserine dehydrogenase isozymes in *Escherichia coli*).

Enterocluster 1. Members of enterocluster 1 share a common bifunctional anthranilate synthase:anthranilate 5-phosphoribosylpyrophosphate phosphoribosyltransferase (AS:PRT) gene product of the *trpG-trpD* fusion (8, 37). Two subgroups are proposed: one contains *Escherichia*, *Shigella*, *Citrobacter* and *Salmonella*; the other contains *Klebsiella* and *Enterobacter*. *Escherichia*, *Shigella*, and *Citrobacter* all lack both CM-F and CDT, and all express DS-Phe as the dominant isozyme (the major isozyme expressed as defined in reference 24) when grown on minimal medium (9). In contrast, *Salmonella* possesses an active CM-F and expresses DS-Tyr as the dominant isozyme (9, 24). *K. pneumoniae* and *E. aerogenes* have in common a type of CDT that does not utilize L-arogenate as an alternative substrate to prephenate but will utilize prephenyllactate (5, 39). It is interesting that strains now belonging to *K. pneumoniae* and *E. aerogenes* were previously placed together in the genus *Aerobacter* (14). However, this genus name has now been discontinued, and organisms previously listed under *Aerobacter* have been reclassified as either *Klebsiella* or *Enterobacter*. *Enterobacter cloacae* and aerogenic strains of *Enterobacter agglomerans* lack CDT altogether. Although the dendrogram shows *E. cloacae* and *Salmonella* species to have the same character states, *E. cloacae* was placed as shown, since all three *Enterobacter* species in enterocluster 1 share an unusual P-protein property (sensitivity of the dehydratase component to inhibition by L-tyrosine or L-tryptophan, in addition to the expected inhibition by L-phenylalanine).

Enterocluster 2. *Serratia* and *Erwinia* are the sole occupants of enterocluster 2. *Enterobacter agglomerans* ATCC 27155 represents anaerogenic species that, in contrast to aerogenic species such as ATCC 29915, lack the *trpG-trpD* fusion (8). The latter has been proposed for renaming within the genus *Erwinia* (see reference 8 and citations therein). The overflow pathway enzymes, CM-F and CDT, are uniformly present among *Serratia* and *Erwinia* species. *Erwinia carotovora*, as the sole exception, lacks CDT. All *Erwinia* species shown, including *Erwinia carotovora*, share in common the property of the P-protein dehydratase whereby L-tyrosine and L-tryptophan are inhibitory, in addition to the expected feedback inhibition by L-phenylalanine (5). Members of enterocluster 2 are readily distinguished from members of enterocluster 1 by the absence of the *trpG-trpD* fusion (8, 37), and they are distinguished from members of enterocluster 3 by the presence of the overflow pathway enzymes, CM-F and CDT.

Enterocluster 3. Enterocluster 3 possesses aromatic pathway character states that are synonymous with the deduced character states of the last common ancestor of the enteric lineage (Fig. 7). The character states of *Aeromonas* and *Alteromonas* are also identical to the deduced ancestral

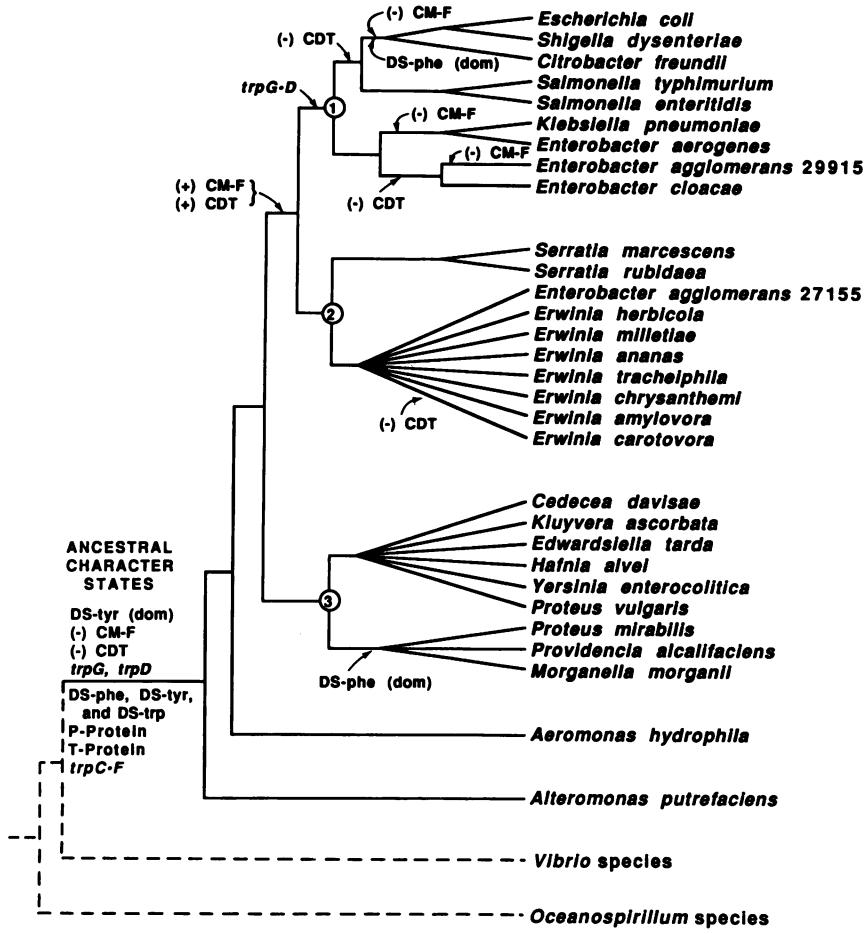

FIG. 7. Schematic representation of the phylogenetic relationships among enteric bacteria. The dendrogram was based upon 16S rRNA oligonucleotide cataloging data (18, 31) as a starting point, biochemical character states then being used for refinement of the dendrogram. Enteroclusters 1, 2, and 3 are indicated by circled numbers. Ancestral character states shown on the left are three gene fusions: *trpC-trpF* and those producing the bifunctional P protein (CM:PDT) and the bifunctional T protein (CM:CDH). The existence of three regulatory isozymes of DAHP synthase (DS-Tyr, DS-Phe, and DS-Trp) is another character state, DS-Phe being unique to the enteric lineage within superfamily B. The common ancestor expressed DS-Tyr as the dominant (24) isozyme DS-Tyr(dom), expressed *trpG* and *trpD* as individual gene products, and had lost the ability to express the CM-F and CDT components of the overflow pathway to L-phenylalanine (17). Loss or gain of a given character state is indicated by (−) or (+), respectively. The fusion of *trpG* and *trpD* is indicated as *trpG-D*. The expression of DS-Phe as the dominant isozyme is indicated by DS-Phe(dom). The outlying branchpoint positions of the genera *Vibrio* and *Oceanospirillum* are shown by dashed lines.

character states. Since *Aeromonas* and *Alteromonas* differ in a large number of character states from members of the nearest divergent lineage within superfamily B (*Oceanospirillum*), the three enteroclusters and the genera *Aeromonas* and *Alteromonas* comprise an assemblage known as the enteric lineage (5–7). Enterocluster 3 includes two subclusters. One contains *Cedecea*, *Kluyvera*, *Edwardsiella*, *Hafnia*, *Yersinia*, and *Proteus vulgaris*; these all possess DS-Tyr as the dominant isozyme of DAHP synthase expressed during growth on minimal medium (9). The second subcluster contains *Providencia*, *Morganella*, and *Proteus mirabilis*; these all possess DS-Phe as the dominant isozyme of DAHP synthase expressed during growth on minimal medium (9). The fact that *Proteus vulgaris* (the type species for *Proteus*) does not cluster with *Proteus mirabilis* reflects the heterogeneity among species currently placed together within the genus *Proteus*. It is noteworthy that *Providencia alcalifaciens* and *Morganella morganii* were initially placed within the genus *Proteus* by phenotypic analysis, being especially similar to *Proteus mirabilis* (14).

Comparison of the proposed phylogenetic tree with existing dendrograms. Although a comprehensive tree containing most of the enteric bacteria has not been previously advanced, incomplete dendrograms that include some of the better-known enteric bacteria have been constructed on the basis of various criteria and methodology (22, 27, 28, 34, 35) (Fig. 1). Our dendrogram is generally consistent with the order of branching obtained in these studies. Our initial dendrogram framework was based upon published (18) and unpublished oligonucleotide sequencing data obtained for 16S rRNA by C. R. Woese, G. Fox, W. G. Weisburg, and co-workers.

The tree given in Fig. 7 identifies three enteroclusters which we propose as valid, definitive phylogenetic subgroups. We have found that analysis of *trpG-trpD*, CM-F, and CDT is sufficient to unambigously place a given enteric bacterium within one of the three enteroclusters. It should be noted here that *Vibrio* species comprise a lineage that diverged from the ancestral trunk before the divergence of the enteric lineage but after the divergence of *Oceanospiril-*

lum species (27, 34); no comparative enzymological data on aromatic biosynthesis are yet available for *Vibrio* species, unfortunately. It is therefore likely that some of the character states that appeared in the ancestor of the enteric lineage within superfamily B (such on DS-Phe and T protein) may also be found in this lineage. Thus, we expect that examination of *Vibrio* species will help pinpoint the exact point of origin of these character states.

Dynamic evolutionary gain/loss of CM-F and CDT. CM-F and CDT are generally present within superfamily B as stable character states. They were initially described in *Pseudomonas aeruginosa* as part of a second overflow phenylalanine pathway (17). The physiological significance of these dual pathways is still enigmatic. The ability to express CM-F and CDT was lost within the common ancestor of the enteric lineage, since CM-F and CDT activities are absent in *Alteromonas* spp., *Aeromonas* spp., and enterocluster 3. Because both reappeared in a common ancestor of enteroclusters 1 and 2, it seems probable that loss of gene expression rather than gene deletion occurred. The cryptic genes were then available for genetic changes restoring expression. It is feasible to probe for the predicted cryptic genes in *Alteromonas* spp., *Aeromonas* spp., and enterocluster 3.

Within enterocluster 1, CM-F and CDT have been very unstable character states. Each member shown in Fig. 7 lacks CM-F or CDT or both. It is intriguing that in *Escherichia coli* unidentified reading frames flank the *pheA* and the *aroF tyrA* operons (22). Transcription of each unidentified reading frame leads to small polypeptides devoid of any known function. These may correspond to CM-F and CDT genes, as suggested in the scheme proposed by Ahmad and Jensen (4) for the evolution of aromatic amino acid biosynthesis in the purple bacteria. If this is true, *E. coli* may possess cryptic genes for CM-F and CDT whose transcription is prematurely terminated to produce nonfunctional polypeptides.

Perspective. The enteric bacteria comprise a cohesive grouping of closely related organisms. Because of their special relationship to humans, more hierarchical subdivision has been carried out with these bacteria than with any other procaryote groupings. As previously discussed, sequencing of even generally ideal molecules such as 16S rRNA genes will not resolve perfect trees at very fine-tuned genealogical levels. We propose on the basis of all the combined information available that the three enteroclusters define lineages having the order of phylogenetic branching shown in Fig. 7. Further branching at an even more fine-tuned level is presented as a tentative basis for confirmation or alteration as additional information involving essential genes and their gene products becomes available.

ACKNOWLEDGMENTS

These studies were supported by grant DMB-8615314 from the National Science Foundation.

We are indebted to Carl Woese, Department of Microbiology, University of Illinois, Urbana, who not only provided the 16S rRNA sequencing results shown in this paper but who has also been a most considerate source of unpublished data essential for our evolutionary studies.

LITERATURE CITED

1. **Ahmad, S., and R. A. Jensen.** 1986. The evolutionary history of two bifunctional proteins that emerged in the purple bacteria. Trends Biochem. Sci. **11**:108–112.
2. **Ahmad, S., and R. A. Jensen.** 1987. A simple spectrophotometric assay for arogenate dehydratase. Anal. Biochem. **163**:107–111.
3. **Ahmad, S., and R. A. Jensen.** 1987. The prephenate dehydrogenase component of the bifunctional T-protein in enteric bacteria can utilize L-arogenate. FEBS Lett. **216**:133–139.
4. **Ahmad, S., and R. A. Jensen.** 1988. New prospects for deducing the evolutionary history of metabolic pathways in prokaryotes: aromatic biosynthesis as a case-in-point. Origins Life. **18**:41–57.
5. **Ahmad, S., and R. A. Jensen.** 1988. Phylogenetic distribution of components of the overflow pathway to L-phenylalanine in enteric bacteria. Curr. Microbiol. **16**:295–302.
6. **Ahmad, S., and R. A. Jensen.** 1988. The phylogenetic origin of the bifunctional tyrosine-pathway protein in the enteric lineage of bacteria. Mol. Biol. Evol. **5**:282–297.
7. **Ahmad, S., and R. A. Jensen.** 1988. The stable evolutionary fixation of a bifunctional tyrosine-pathway protein in enteric bacteria. FEMS Microbiol. Lett. **52**:109–116.
8. **Ahmad, S., and R. A. Jensen.** 1989. Utility of a bifunctional tryptophan-pathway enzyme for the classification of the Herbicola-Agglomerans complex of bacteria. Int. J. Syst. Bacteriol. **39**:100–104.
9. **Ahmad, S., and R. A. Jensen.** 1989. The stable phylogenetic distribution of the recently evolved L-phenylalanine-inhibited isozyme of 3-deoxy-D-*arabino* heptulosonate 7-phosphate synthase in enteric bacteria. Curr. Microbiol. **18**:341–349.
10. **Berry, A., and R. A. Jensen.** 1988. Biochemical evidence for phylogenetic branching patterns: biochemical-pathway characteristics can indicate points of divergence among prokaryotes. BioScience **38**:99–103.
11. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem. **72**:248–254.
12. **Byng, G. S., J. F. Kane, and R. A. Jensen.** 1982. Diversity in routing and regulation of complex biochemical pathways as indicators of microbial relatedness. Crit. Rev. Microbiol. **8**: 227–252.
13. **Cotton, R. G. H., and F. Gibson.** 1965. The biosynthesis of phenylalanine and tyrosine: enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. Biochim. Biophys. Acta **100**:76–88.
14. **Cowan, S. T.** 1974. Family I. *Enterobacteriaceae* Rahn 1937, 281 *Nom. gen. cons. opin.* 15, Jud. Comm. 1958, 73, p. 290–293. *In* R. E. Buchanan and N. E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore.
15. **Crawford, I. P.** 1980. Gene fusions in the tryptophan pathway: tryptophan synthase and phosphoribosyl-anthranilate isomerase: indoleglycerolphosphate synthase, p. 151–173. *In* H. Bisswanger and E. Schminke-Ott (ed.), Multifunctional proteins. John Wiley & Sons, Inc., New York.
16. **Dayan, J., and D. B. Sprinson.** 1970. Preparation of prephenic acid. Methods Enzymol. **17A**:559–561.
17. **Fiske, M. J., R. J. Whitaker, and R. A. Jensen.** 1983. Hidden overflow pathway to L-phenylalanine in *Pseudomonas aeruginosa*. J. Bacteriol. **154**:623–631.
18. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. A. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Leuhrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209**: 457–463.
19. **Gibson, F.** 1970. Preparation of chorismic acid. Methods Enzymol. **17A**:362–364.
20. **Harnes, E., J.-H. Hsu, C. S. Subramanyam, and H. E. Umbarger.** 1985. Comparison of the regulatory regions of *ilv* GEDA operons from several enteric organisms. J. Bacteriol. **164**: 207–216.
21. **Hori, H., and S. Osawa.** 1979. Evolutionary change in 5S rRNA secondary structure and a phylogenetic tree of 54 5S rRNA species. Proc. Natl. Acad. Sci. USA **76**:381–385.
22. **Hudson, G. S., and B. E. Davidson.** 1984. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of

Escherichia coli K12. J. Mol. Biol. **180**:1023–1051.
23. **Jensen, R. A.** 1985. Biochemical pathways in prokaryotes can be traced backward through evolutionary time. Mol. Biol. Evol. **2**:92–108.
24. **Jensen, R. A., and D. S. Nasser.** 1968. Comparative regulation of isoenzymic DAHP synthetases in microorganisms. J. Bacteriol. **95**:188–196.
25. **Keener, S. L., K. P. McNamee, and K. McEntee.** 1984. Cloning and characterization of *rec*A genes from *Proteus vulgaris*, *Erwinia carotovora*, *Shigella flexneri*, and *Escherichia coli* B/r. J. Bacteriol. **160**:153–160.
26. **Lindroth, P., and K. Mopper.** 1979. High performance liquid chromatography determinations of subpicomole amounts of amino acids by precolumn fluorescence derivatization with *o*-pthalaldehyde. Anal. Chem. **51**:1667–1674.
27. **MacDonell, M. T., D. G. Swartz, B. A. Ortiz-Conde, G. A. Last, and R. R. Colwell.** 1986. Ribosomal RNA phylogenies for the vibrio-enteric group of eubacteria. Microbiol. Sci. **3**:172–178.
28. **Nichols, B. P., G. F. Miozzari, M. Cleemput, G. N. Bennett, and C. Yanofsky.** 1980. Nucleotide sequences of *trp*G regions of *Escherichia coli*, *Shigella dysenteriae*, *Salmonella typhimurium* and *Serratia marcescens*. J. Mol. Biol. **142**:503–517.
29. **Patel, N., D. L. Pierson, and R. A. Jensen.** 1977. Dual enzymatic routes to L-tyrosine and L-phenylalanine via pretyrosine in *Pseudomonas aeruginosa*. J. Biol. Chem. **252**:5839–5846.
30. **Shiaun, D., and A. Campbell.** 1988. Transcriptional regulation and gene arrangement of *Escherichia coli*, *Citrobacter freundii* and *Salmonella typhimurium* biotin operons. Gene. **67**:203–211.
31. **Stackebrandt, E., and C. R. Woese.** 1981. The evolution of prokaryotes, p. 1–32. *In* M. J. Carlile, J. F. Collins, and B. E. B. Moseley (ed.), Molecular and cellular aspects of microbial evolution. 32nd Symposium of the Society for General Microbiology. Cambridge University Press, Cambridge.
32. **Winkler, U. K., and M. Stuckman.** 1979. Glycogen, hyaluronate and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. J. Bacteriol. **138**:663–670.
33. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.
34. **Woese, C. R., W. G. Weisburg, C. M. Hahn, B. J. Paster, L. B. Zablen, B. J. Lewis, T. J. Macke, W. Ludwig, and E. Stackebrandt.** 1985. The phylogeny of purple bacteria: the gamma subdivision. Syst. Appl. Microbiol. **6**:25–33.
35. **Yamagata, H., K. Nakamura, and M. Inouye.** 1981. Comparison of the lipoprotein gene among the *Enterobacteriaceae*. DNA sequence of *Erwinia amylovora* lipoprotein gene. J. Biol. Chem. **256**:2194–2198.
36. **Yanofsky, C.** 1984. Comparison of regulatory and structural regions of genes of tryptophan metabolism. Mol. Biol. Evol. **1**:143–161.
37. **Zalkin, H.** 1980. Anthranilate synthase: relationship between bifunctional and monofunctional enzymes, p. 123–149. *In* H. Bisswanger, and E. Schmincke-Ott (ed.), Multifunctional proteins. John Wiley & Sons, Inc., New York.
38. **Zamir, L. O., R. A. Jensen, B. H. Arison, A. W. Douglas, G. Albers-Schonberg, and J. R. Bowen.** 1980. Structure of arogenate (pretyrosine), an amino acid intermediate of aromatic biosynthesis. J. Am. Chem. Soc. **102**:4499–4504.
39. **Zamir, L. O., R. Tiberio, K. A. Devor, F. Sauriol, S. Ahmad, and R. A. Jensen.** 1988. Structure of D-prephenyllactate: a carboxycyclohexadienyl metabolite from *Neurospora crassa*. J. Biol. Chem. **263**:17284–17290.
40. **Zamir, L. O., R. Tiberio, M. Fiske, A. Berry, and R. A. Jensen.** 1985. Enzymatic and non-enzymatic dehydration reactions of L-arogenate. Biochemistry **24**:1607–1612.