Review

Biochemical Pathways in Prokaryotes Can Be Traced Backward through Evolutionary Time[1]

Roy A. Jensen

Center for Somatic-Cell Genetics and Biochemistry, Department of Biological Sciences, State University of New York at Binghamton

For the first time, a credible prokaryotic phylogenetic tree is being assembled by Woese and others using quantitative sequence analysis of oligonucleotides in the highly conservative rRNA. This provides an evolutionary scale against which the evolutionary steps that led to the arrangement and regulation of contemporary biochemical pathways can be measured. This paper presents an emerging evolutionary picture of aromatic amino acid biosynthesis within a large superfamily assemblage of prokaryotes that is sufficiently developed to illustrate a new perspective that will be applicable to many other biochemical pathways.

The Emerging Phylogeny of Prokaryotes

Until recently, the phylogenetic past of prokaryotes was generally conceded to be forever lost. The lack of a fossil record, inevitable instances of convergent evolution, and possible scenarios of "horizontal" gene transfer by promiscuous extrachromosomal agents posed seemingly insurmountable barriers. Fortunately, horizontal gene transfer in natural populations seems limited to genes whose functions are usually dispensable in nature (Campbell 1981). The idea (Zuckerkandl and Pauling 1965) that macromolecular sequences might be sufficiently conserved to serve as "documents of evolution"—being, in essence, living fossils—has been realized in oligonucleotide cataloging (Fox et al. 1980; Stackebrandt and Woese 1981), a technique that lends itself to the computer-assisted estimation of an evolutionary similarity between any pair of taxa/organisms. This, expressed as a similarity coefficient (S_{AB}), has been discussed at length by Fox et al. (1977). At one extreme, catalogs obtained from samples of identical cell populations would yield an S_{AB} of 1.0. These catalogs provide a basis for estimating the evolutionary relationships of the organisms from which the catalogs were derived. The resulting new phylogenetic classification often differs drastically from the pragmatic classification offered by *Bergey's Manual of Determinative Bacteriology*. This is illustrated by the fact that species currently named within a single genus (e.g., *Rhodopseudomonas gelatinosa* and *Rp. capsulata*) are now not even considered to belong to the same family (Fox et al. 1980; Stackebrandt and Woese 1981).

Evolution of Biochemical Pathways

Metabolic pathways are a source of more biochemical diversity in nature than is generally appreciated. Not only may pathway enzymes differ in cofactor specificities, metal requirements, allosteric specificities, and multifunctional capabilities, but the

1. Key words: microbial phylogeny, evolution, aromatic biosynthesis, cofactor specificity. Abbreviations: DAHP = 3-deoxy-D-*arabino*-heptulosonate 7-phosphate; NAD^+ = nicotinamide adenine dinucleotide; $NADP^+$ = nicotinamide adenine dinucleotide phosphate.

Mol. Biol. Evol. 2(2):92–108. 1985.

0737-4038/85/0202-0006$02.00

enzymic-step construction of the biochemical pathways may differ. Organisms sharing the same biochemical state for characters such as those listed above are not necessarily genealogically close. Neither does the recognition of different pathway character states provide a compelling basis for conclusions about pathway evolution. On the other hand, if organisms under study have known positions on a credible phylogenetic tree, then deductions about pathway evolution are feasible. The biochemical-pathway data per se are not used to construct the phylogeny because: (1) no quantitative measure comparable to an S_{AB} value is available; (2) the various character states are not equivalent to one another—compare, for instance, such features as the presence or absence of an enzyme, enzyme sensitivity to inhibition by compound a or by compound b, or the presence or absence of a multifunctional protein (indicating occurrence or nonoccurrence of a gene fusion event); and (3) the impact of a given pathway on overall metabolism differs from organism to organism because biochemical diversity in nature is considerable (e.g., an amino acid may be a precursor of a quantitatively significant pigment in one microbe but not in others). Thus, the various character states of biochemical pathways are not readily translated into a number that expresses evolutionary distance ([1] and [2] above), and a given character state is undoubtedly not equally constrained in different organisms ([3] above). Hence, our approach is to assume the phylogeny established by oligonucleotide cataloging as the basis for reconstruction of the evolutionary history of aromatic amino acid biosynthesis. A consistent, cladistic methodology can then be employed in which the most parsimonious solutions are adopted.

Diversity of Aromatic Biosynthesis and Regulation

In all organisms studied thus far, the biosynthesis of all three aromatic amino acids begins with a common trunk of seven enzymes starting with the condensation of erythrose-4-P and phosphoenolpyruvate by DAHP synthase and ending with chorismate, the last intermediate common to all aromatic amino acids. From chorismate there are three major branches leading to *L*-phenylalanine, *L*-tyrosine, and *L*-tryptophan. This pathway possesses five enzymes that are commonly, but not always, targets of allosteric control. These enzymes are DAHP synthase, chorismate mutase (which converts chorismate to prephenate), and the enzymes catalyzing the initial irreversible step at the beginning of each of the three terminal branches.

Many means are used by microorganisms to ensure allosteric control of DAHP synthase (Jensen et al. 1967; Jensen and Nasser 1968; Jensen 1970; Jensen and Rebello 1970; Jensen and Stenmark 1970; Jensen and Twarog 1972). The discovery by Stenmark et al. (1974) of *L*-arogenate, then known as pretyrosine, in cyanobacteria was the initial hint that not only the regulation of but even the basic pathways to phenylalanine and tyrosine are highly variable in nature. Diversity (see Byng et al. 1982) of pathway construction and control has been documented by comprehensive studies of cyanobacteria (Hall et al. 1982) and of pseudomonads (Byng et al. 1980; Whitaker et al. 1981*a*, 1981*b*; Byng et al. 1983*a*, 1983*b*, 1983*c*). Five distinct rRNA homology groups of pseudomonads exist (Palleroni et al. 1973; Palleroni 1983), and we found that certain biochemical features of aromatic biosynthesis paralleled these five classes perfectly. Indeed, unknown species can be classified into these five groups and in some cases into distinct subgroups through the determination of patterns of enzyme arrangement and control (Byng et al. 1983*b*). Hence, once a

given pathway pattern is established to be conservative at some phylogenetic level, as gauged by rRNA, then the biochemical pattern can be used reliably for group placement.

The recent availability of dendrograms based on oligonucleotide catalogs has generated a new evolutionary thrust. We can now characterize biochemical pathways and their regulation in organisms that are known to have diverged from one another recently and then systematically work our way back to progressively earlier phylogenetic positions. Although our approach is to take the phylogenetic tree constructed by oligonucleotide cataloging at face value, the exact tree is undoubtedly subject to some revision. We fully expect that fine-tuning of branch placement will be in order for outlying dendrogram sections where phylogenetic distances are small. Such refinement can most conveniently and inexpensively be pursued by means of nucleic acid hybridization (Byng et al. 1983*b*). Indeed, if aromatic-pathway analysis yields patterns that lead to a questioning of the accuracy of certain branch points, additional information—including, perhaps, similar efforts with other metabolic systems—should help resolve these questions.

Current Interpretations about the Evolution of Aromatic Biosynthesis

A general outline of plausible evolutionary events that underlie aromatic-pathway biosynthesis and regulation is now feasible. Although only scattered information in a wide range of microbial groupings is available concerning aromatic-pathway biosynthesis and regulation, an extensive amount of background data exists concerning the division denoted as purple bacteria by Woese and co-workers (fig. 1).

Cofactor Specificity

Figure 1 depicts the simplest interpretation of the data thus far available about cofactor specificity of prephenate- and/or arogenate-dehydrogenase enzyme activities. In the frequent cases in which both dehydrogenase activities coexist, identical cofactor specificities have always been found. A strong rationale has been presented to argue that broad specificity for substrate (and cofactor) is a primitive property of enzymes (Jensen 1976). Such cofactor specificity for NAD^+ or $NADP^+$ is presented as the ancestral state that existed at the point of evolutionary divergence of the three superfamilies shown in figure 1.

So far, no instance of cofactor specialization has been found within superfamily A. In superfamily C, on the other hand, specialization for $NADP^+$ occurred in an ancestor of the upper tripartite cluster (fig. 1) at an evolutionary time following divergence of that ancestor from group IV pseudomonads. In superfamily B, specialization for $NADP^+$ occurred within the *Acinetobacter* lineage, in contrast to specialization for NAD^+, which occurred in a common ancestor of *Escherichia coli*, group I, and group V pseudomonads. (Note the inappropriate contemporary naming of the two *A. calcoaceticus* strains within the same species; their oligonucleotide catalogs indicate a phylogenetic separation above the level of genus.)

Data concerning species within superfamilies A and C are too scattered to warrant interpretations of their aromatic-pathway evolution, but information about superfamily B is sufficiently comprehensive to suggest probable evolutionary progressions. Members of superfamily B thus far studied possess tyrosine-branch dehydrogenases that are specific for either NAD^+ or $NADP^+$. Do any superfamily

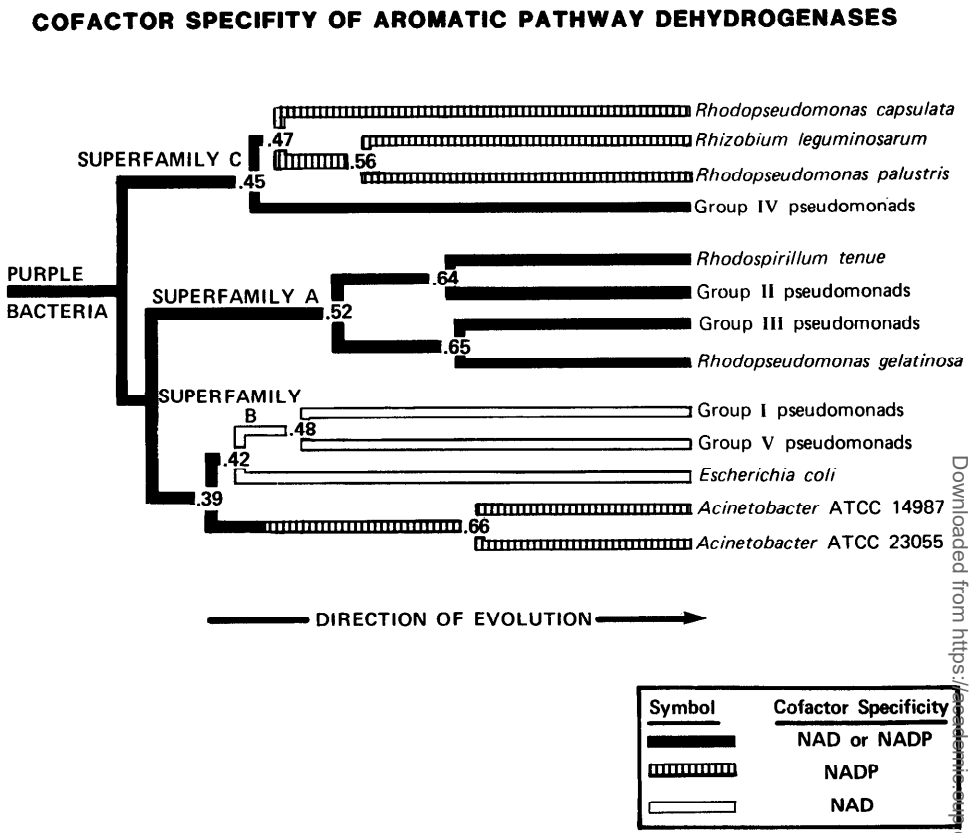

FIG. 1.—Dendrogram based on oligonucleotide cataloging of members of three superfamily components of the "purple bacteria". The latter comprise one of eight major "eubacterial" lines of prokaryote descent depicted by Fox et al. (1980) within a lineage that is divergent from lineages containing archaebacterial and eukaryotic kingdoms. Superfamilies A, B, and C correspond to purple nonsulfur-2, purple sulfur, and purple nonsulfur-1 groupings presented by Fox et al. (1980). The numbers positioned at points of divergence are S_{AB} values obtained by C. R. Woese and co-workers. The dendrogram connections of the three superfamilies are drawn according to the dendrograms given by Palleroni (1983). Although the dendrogram given in figure 4 of Fox et al. (1980) shows superfamily A to connect at the deepest phylogenetic level (S_{AB} = ~0.30), the deductions made about the evolution of cofactor specificity fit either tree equally well. *Rhizobium* strains were *Rh. leguminosarum* 3841 and *Rh.* sp. KH486 from J. E. Beringer and A. Johnston. Other strains carry the following ATCC numbers: *Rhodopseudomonas capsulata,* ATCC 11166; *Rp. palustris,* ATCC 17001; *Rhodospirillum tenue,* ATCC 25093; *Rp. gelatinosa,* ATCC 17011; and *Rp. gelatinosa,* ATCC 17013. An extensive list of species comprising pseudomonad groups I, II, III, IV, and V are given in Byng et al. (1980), Whitaker et al. (1981*a*, 1981*b*), and Byng et al. (1983*a*). The S_{AB} values given in this and succeeding figures were obtained from Fox et al. (1980), Stackebrandt and Woese (1981), Woese et al. (1982), and personal communication of unpublished data from G. Fox and C. R. Woese. The clusters of cofactor specificity thus far identified are shown by patterns defined in the box on the lower right.

B organisms retain the broad cofactor specificity consistent with the interpretation drawn in figure 1 for ancestral stem organisms of superfamily B? Because of its deep point of divergence within superfamily B (fig. 2), the *Desulfovibrio* lineage is a logical grouping for possible identification of retained breadth of cofactor specificity.

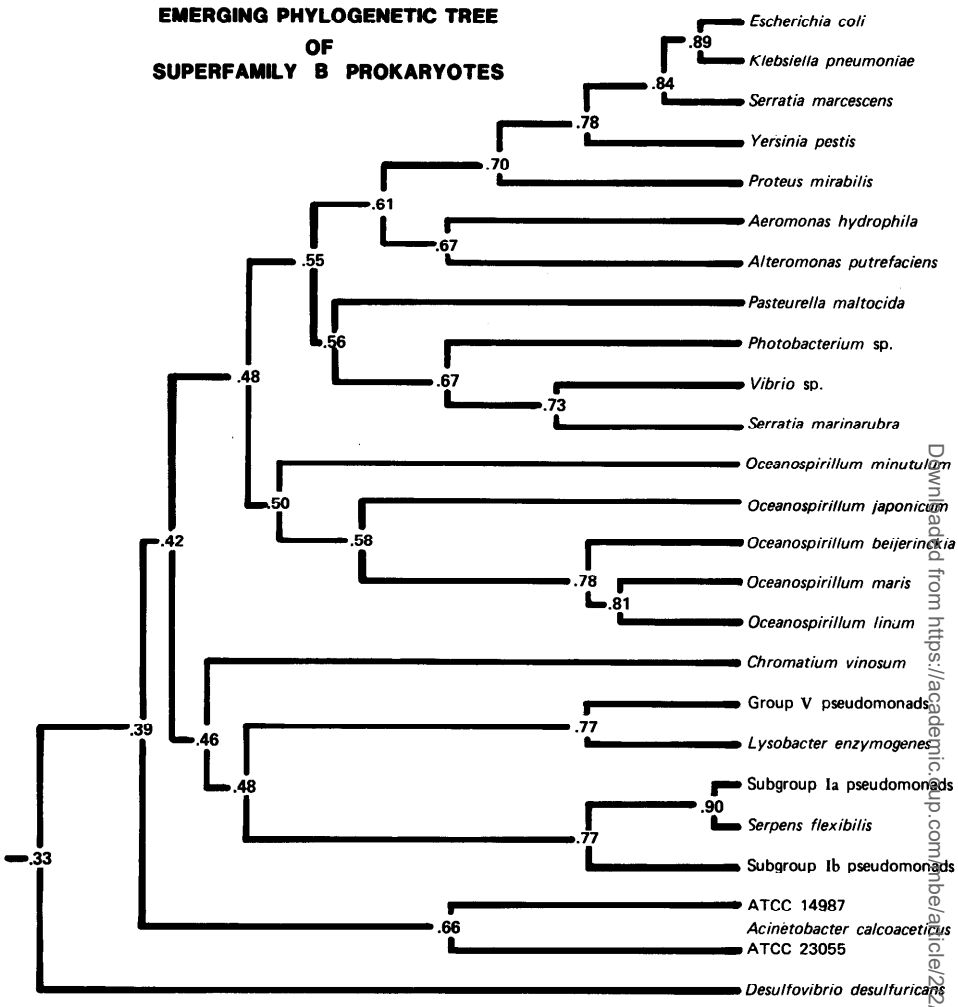

FIG. 2.—Dendrogram of superfamily B organisms based on a combination of published (Fox et al., 1980; Stackebrandt and Woese, 1981; Woese et al., 1982) and unpublished data obtained by C. R. Woese and co-workers. Numbers shown at points of bifurcation are S_{AB} values obtained by oligonucleotide cataloging.

DAHP Synthase in Superfamily B

The dendrogram of superfamily B organisms, in relationship to the number of isozymes of DAHP synthase that are expressed in the lineages given, is shown in figure 3. An ancestral state of two isozymes is postulated to have already existed at this evolutionary time. Gene duplication leading to a third isozyme in the enteric lineage seems likely, whereas one of the two ancestral isozymes has been lost in group V pseudomonads (mainly species of *Xanthomonas*).

The DAHP synthase isozymes are named according to their specificities for feedback inhibitors, that is, DAHP synthase–tyr, DAHP synthase–trp, and DAHP synthase–phe. Group V pseudomonads possess only DAHP synthase–trp. Two-

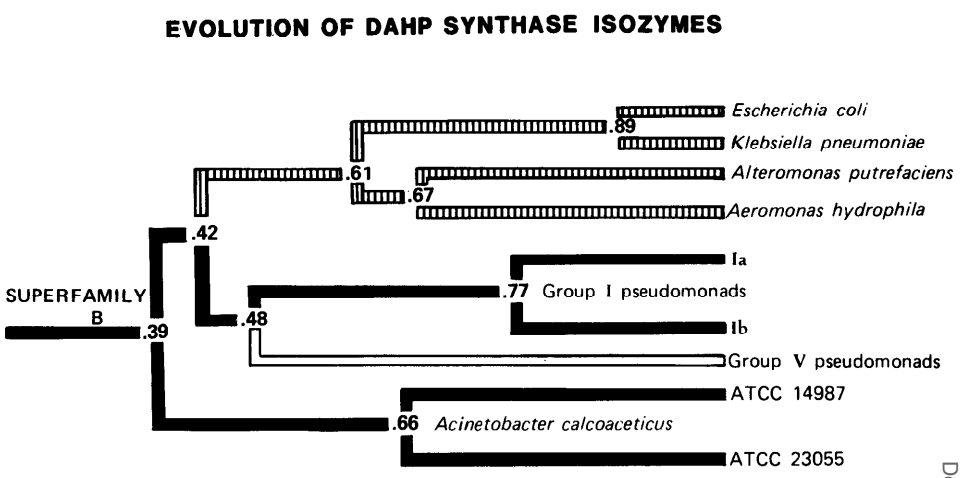

FIG. 3.—Dendrogram based on oligonucleotide cataloging of member groups within superfamily B in correlation with information available about 3-deoxy-*D-arabino*-heptulosonate 7-phosphate (DAHP) synthase. Solid branches indicate that two isozymes of DAHP synthase exist; barred branches, three isozymes; open branches, one enzyme. In *Escherichia coli* three differentially regulated isozymes of DAHP synthase are specified by cistrons *aroF*, *aroG*, and *aroH*, which map at 17 min, 37 min, and 57 min, respectively (Bachmann 1983). *Alteromonas putrefaciens* was fortuitously characterized in our pseudomonad studies (Whitaker et al. 1981*a*), having previously been misnamed as *Pseudomonas putrefaciens*. *Aeromonas hydrophila* ATCC 9071, *A. liquifaciens* ATCC 14715, and *A. formicans* ATCC 13137 all have been shown to possess three DAHP synthase isozymes (Jensen, Nasser, and Nester 1967; Jensen and Stenmark 1970). Group I and group V pseudomonads were characterized in comprehensive detail for DAHP synthase (Whitaker et al. 1981*a*). Results shown for *Acinetobacter* strains are based on unpublished data. Nucleotide sequencing of *aroG* and *aroH* in *E. coli* K-12 have affirmed the origin of DAHP synthase-phe and DAHP synthase-trp from a common ancestral gene (Davies and Davidson 1982).

isozyme organisms characteristically possess DAHP synthase–tyr and DAHP synthase–trp. DAHP synthase–phe occurs only in three-isozyme systems. Since DAHP synthase–phe seems to have been the most recently acquired isozyme in superfamily B, it would be relatively easy to trace its origin within the enteric lineage. Is it present within the *Pasteurella/Photobacterium/Vibrio* lineage? If so, is it present within the little known but phylogenetically diverse *Oceanospirillum* lineage (fig. 2)?

Otherwise present throughout superfamily B, DAHP synthase–tyr became lost from the group V pseudomonad lineage. Since *Lysobacter* diverged from this group, the presence or absence of DAHP synthase–tyr in *Lysobacter* will pinpoint the evolutionary loss of this isozyme before or after the point of divergence shown in figure 2.

DAHP synthase–trp has been detected in every superfamily B organism examined to date. A progression of DAHP synthase–trp enzyme types exists with respect to sensitivity to inhibition by chorismate. The enzyme in *E. coli* lacks sensitivity to inhibition by chorismate, in *Pseudomonas aeruginosa* it exhibits weak inhibition by chorismate, and in *X. campestris* it exhibits potent inhibition by chorismate (table 1). Feedback control of a single DAHP synthase by chorismate in *X. campestris* can be reconciled with a physiologically efficient pattern of regulation in which direct feedback inhibition of chorismate mutase and anthranilate synthase by end products would elevate chorismate levels, indirectly leading to inhibition of

Table I
Comparison of Allosteric Sensitivities of DAHP Synthase-trp from *Pseudomonas aeruginosa* with Those of the Single DAHP Synthase Enzyme of *Xanthomonas campestris*

Organism	Pseudomonad Group	K_i (*L*-tryptophan)		K_i (chorismate)	
		PEP	E-4-P	PEP	E-4-P
Pseudomonas aeruginosa[a]	I	0.04 m*M*	0.005 m*M*	1.35 m*M*	2.25 m*M*
Xanthomonas campestris[b]	V	0.38 m*M*	0.46 m*M*	0.40 m*M*	0.10 m*M*

Note.—E-4-P = erythrose-4-phosphate; PEP = phosphoenolpyruvate.
[a] Data from Whitaker et al. (1982). Results are representative of other Group I organisms.
[b] Data from Whitaker et al. (1985). Results are representative of other Group V organisms.

DAHP synthase. This is a variation of a control pattern (Sequential Feedback Inhibition) that is well known in *Bacillus* species. Since in the latter case DAHP synthase and chorismate mutase activities coexist as a multifunctional protein (Byng and Jensen 1983), perhaps the allosteric binding sites for chorismate and *L*-tryptophan of the xanthomonad DAHP synthase arose by fusion of a gene for an unregulated DAHP synthase with the gene specifying an *L*-tryptophan-inhibited anthranilate synthase, after gene duplication of the anthranilate synthase cistron (Jensen 1976). If so, DAHP synthase–trp of both *P. aeruginosa* and *X. campestris* have already lost (author's unpublished data) intrinsic activity for anthranilate synthase (aminase); however, it is possible that data supporting such a gene-fusion event might yet be obtained from a study of *Acinetobacter* or *Desulfovibrio.*

Of interest is whether DAHP synthase–trp or DAHP synthase–tyr originated first. If DAHP synthase–trp was the ancestral enzyme, the physiologically important allosteric effector was probably chorismate. The subsequent evolutionary acquisition of DAHP synthase–tyr may have correlated with progressively increasing and decreasing sensitivities of DAHP synthase–trp to feedback inhibition by *L*-tryptophan and chorismate, respectively. If this scenario is correct, then DAHP synthase–trp of group V pseudomonads probably resembles the ancient evolutionary state, that is, that characterized by sensitive regulation by chorismate. On the other hand, if DAHP synthase–tyr originated first, the major sensitivity of DAHP synthase–trp to chorismate in group V pseudomonads may reflect an evolved regulatory adjustment to the loss of DAHP synthase–tyr. It should be informative to determine whether *Desulfovibrio* possesses one or two molecular species of DAHP synthase.

Metabolic Segment for *L*-Phenylalanine

The P-protein, having been found in every superfamily B organism studied to date, is clearly of ancient origin (fig. 4). This multifunctional protein carries both chorismate mutase and prephenate dehydratase activities and was first described in *Escherichia coli* and *Klebsiella pneumoniae* (Cotton and Gibson 1965). In *Acinetobacter calcoaceticus* prephenate dehydratase activity is heavily dependent on the presence of *L*-tyrosine (author's unpublished data). In group I pseudomonads prephenate dehydratase activity is stimulated (K_m effect) by *L*-tyrosine (Byng et al. 1983*c*). In *E. coli* prephenate dehydratase activity is not activated by *L*-tyrosine at

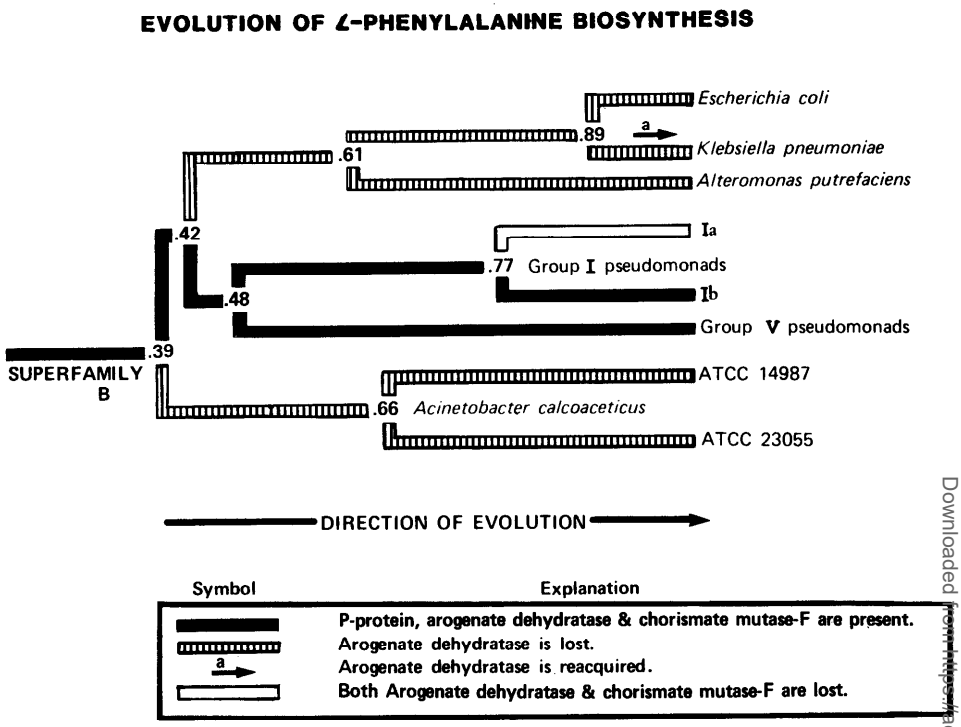

FIG. 4.—Dendrogram based on oligonucleotide cataloging of member groups within superfamily B in correlation with information available about phenylalanine biosynthesis and regulation. Arogenate dehydratase is reacquired only in a mutant strain of *Klebsiella* that has been subjected to multiple rounds of mutagenesis (see text for details).

all, even being inhibited by very high concentrations of *L*-tyrosine (Cotton and Gibson 1965). A comparison of the P-protein in *E. coli* with that in *P. aeruginosa* and *A. calcoaceticus* reveals similar enzymological characteristics. In *E. coli* the P-protein is active in the dimeric state, and the presence of feedback inhibitor (*L*-phenylalanine) promotes formation of an inactive tetramer (Baldwin et al. 1981). In *Acinetobacter* an active state requires *L*-tyrosine, and *L*-phenylalanine produces an inactive state of twofold greater molecular weight (author's unpublished data). It seems likely that these states correspond to dimer and tetramer moieties, respectively. *Pseudomonas aeruginosa* is intermediate in the sense that *L*-tyrosine activates prephenate dehydratase activity much less dramatically. Molecular weight transitions have not been studied in *P. aeruginosa* yet. An interesting continuum of diversity is thus emerging with respect to the effect of *L*-tyrosine on the prephenate dehydratase activity of the P-protein, and it should be revealing to characterize other P-proteins, selected with reference to dendrogram position, along these lines.

Figure 4 pictures an ancestral state in which the P-protein coexists with arogenate dehydratase and a monofunctional species of chorismate mutase, chorismate mutase–F (Byng et al. 1983*c*). At an evolutionary time preceding an S_{AB} value of 0.61 within the enteric lineage, arogenate dehydratase was lost. Although chorismate mutase–F per se is absent from *E. coli* and *K. pneumoniae*, we presume that chorismate mutase–F was the progenitor of the T-protein chorismate mutase prior

to the gene fusion with the gene encoding prephenate dehydrogenase that formed the multifunctional T-protein. Thus, *Alteromonas putrefaciens,* which predates evolution of the T-protein (fig. 5), still possesses chorismate mutase–F. *Acinetobacter* species, which lost arogenate dehydratase, also retained chorismate mutase–F as would be expected in the absence of an evolved T-protein.

Although wild-type *K. pneumoniae* ATCC 25304 lacks activity for arogenate dehydratase as does *E. coli,* a well-known derivative carrying multiple aromatic-pathway mutations has regained activity for arogenate dehydratase. *Klebsiella pneumoniae* 62-1, ATCC 25306 (Cotton and Gibson 1965) has been subjected to multiple mutagenesis protocols in order to obtain a triple auxotroph that has been widely used for accumulation of chorismic acid. If the active arogenate dehydratase of *K. pneumoniae* 62-1 is an atavistic enzyme resulting from reactivation of a cryptic gene that was initially silenced by a point mutation, then the enzyme would be expected to resemble the arogenate dehydratase of pseudomonad subgroup Ib species (Whitaker et al. 1981*b*). In conformity with this expectation, the *K. pneumoniae* enzyme will accept prephenate in addition to *L*-arogenate as substrate and is not feedback inhibited by *L*-phenylalanine. This arogenate dehydratase is synonymous with the "prephenate dehydratase A" noted by Cotton and Gibson (1965). Thus, wild-type *K. pneumoniae* appears to possess a pseudogene. Although only a few examples of such cryptic genes have been demonstrated (e.g., Lawther

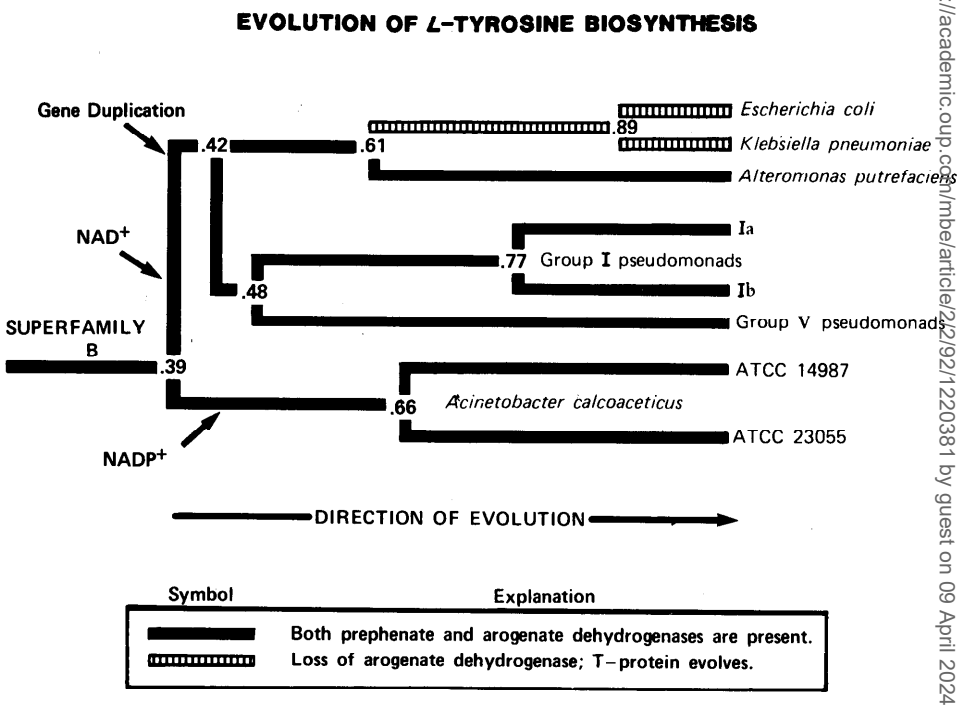

FIG. 5.—Dendrogram based on oligonucleotide cataloging of member groups within superfamily B in correlation with information available about tyrosine biosynthesis and regulation. At an S_{AB} value of 0.39, two lineages separated, one having a dehydrogenase specific for $NADP^+$ and the other having a dehydrogenase that evolved specificity for NAD^+. The probability that duplication of the cistron encoding the primitive dehydrogenase occurred prior to the 0.42 bifurcation within the NAD^+ lineage is shown.

et al. 1981), their possible significance in evolution has been discussed recently (Hall et al. 1983).

One might predict that all enteric bacteria having an S_{AB} relationship of 0.61 or higher with respect to *E. coli* are likely to carry the cryptic gene for arogenate dehydratase. This could be tested by cloning the gene from an organism such as *P. aeruginosa* (pseudomonad subgroup Ib) and using this as a molecular probe for detection of the cryptic gene. It also seems likely that restoration of an active arogenate dehydratase could be achieved by selective pressure using an appropriate mutant background in an organism such as *E. coli.*

Subgroup Ia pseudomonads lack both arogenate dehydratase and chorismate mutase–F (Byng et al. 1983*c*). The absence of chorismate mutase–F would appear to be a formidable barrier to the future evolution of a T-protein arrangement. Extensive genetic and physiological studies of *P. aeruginosa,* a subgroup Ib species possessing both of these activities, have established the roles of these activities in an unregulated, overflow pathway to *L*-phenylalanine (Fiske et al. 1983). The operation of this flow route seems to be intimately tied to the level of pathway flux permitted by any given carbon source.

Metabolic Segment for *L*-Tyrosine

Most members of superfamily B possess both prephenate dehydrogenase activity and arogenate dehydrogenase activity, as shown in figure 5. In the *Acinetobacter* lineage a single enzyme of broad substrate specificity appears to account for both activities. Thus, (1) the two activities copurified during enzyme fractionation, (2) identical Km values for $NADP^+$ were calculated, and (3) identical K_i values of 9 μM for *L*-tyrosine inhibition were measured.

The second lineage diverging at an S_{AB} value of 0.39 exhibits dehydrogenase specialization for NAD^+ (see fig. 1). Since these dehydrogenases have been separated in group V pseudomonads (Whitaker et al. 1985) and in group I pseudomonads (Patel et al. 1978; author's unpublished data), it is likely that gene duplication of the cistron encoding an ancestral dehydrogenase of broad substrate specificity followed shortly *after* the event of cofactor specialization. Consistent with the presence of two enzyme species, K_m values for NAD^+ differ depending on whether prephenate or *L*-arogenate is used as substrate, and K_i values for *L*-tyrosine inhibition differ between prephenate dehydrogenase and arogenate dehydrogenase. In *X. campestris* (Whitaker et al. 1985) one dehydrogenase is specific for prephenate, whereas the second dehydrogenase is reactive with either prephenate or arogenate. The latter dehydrogenase is presumably like the ancestral enzyme, whereas the former dehydrogenase has evolved specialized substrate specificity.

Within the enteric lineage, at a time following divergence from *A. putrefaciens,* arogenate dehydrogenase activity was lost. This evolutionary event was approximately coincident with the fusion of the cistrons encoding prephenate dehydrogenase and chorismate mutase–F, thereby creating the bifunctional T-protein. It is a striking contrast that the bifunctional T-protein of tyrosine biosynthesis is of recent origin within superfamily B, whereas the bifunctional P-protein of phenylalanine biosynthesis is very ancient. A systematic analysis of the organisms between *E. coli* and *A. putrefaciens* on the dendrogram of figure 2 should pinpoint the evolutionary origin of the T-protein. The foregoing scenario predicts that the presence of the T-protein will inevitably correlate with the absence of chorismate mutase–F.

Thus far, no aromatic-pathway dehydrogenase activity has been found to be insensitive to feedback inhibition by *L*-tyrosine. However, quantitative differences have been found to be consistent characteristics of different dendrogram sections; for example, pseudomonad group I (characterized by hypersensitivity of arogenate dehydrogenase to inhibition) can be reliably distinguished from pseudomonad group V (characterized by hypersensitivity of prephenate dehydrogenase to inhibition) on this criterion alone (Byng et al. 1980).

The Ancestral Superfamily B Pathway

The separate conclusions assembled in figures 3–5 concerning the evolutionary events involved in particular segments of aromatic amino acid biosynthesis can be combined to yield a view of the ancestral pathway of aromatic biosynthesis. The biosynthetic pathway that existed in the common ancestor of superfamily B organisms at an evolutionary time defined by an S_{AB} value of 0.39 can be deduced as shown in figure 6. This ancestor possessed two isozymes of DAHP synthase, one sensitive to feedback inhibition by *L*-tyrosine and the other by *L*-tryptophan. The bifunctional P-protein, probably sensitive to *L*-phenylalanine-mediated feedback inhibition, had already evolved—being coexistent with an unregulated *L*-arogenate flow route to *L*-phenylalanine. Dual flow routes existed to *L*-tyrosine, but each of the two dehydrogenase reactions was catalyzed by a common enzyme. This broadly specific dehydrogenase also used either pyridine nucleotide cofactor and was probably already sensitive to feedback inhibition by *L*-tyrosine. No contemporary organism has yet been found that would fit the projected description of the superfamily B ancestor; it is most reminiscent of group I pseudomonads such as *Pseudomonas aeruginosa* (especially for phenylalanine synthesis) and of *Acinetobacter* species (especially for tyrosine synthesis).

Note in figure 6 that whereas gene duplication of the ancestral dehydratase (b) must have preceded acquisition of feedback inhibition, the evolution of feedback sensitivity in the ancestral dehydrogenase (c) apparently preceded gene duplication. Thus, all contemporary arogenate dehydratases so far studied within superfamily B (group I pseudomonads, group V pseudomonads, and *Klebsiella pneumoniae*) are unregulated (in contrast to prephenate dehydratases). On the other hand, both the activities of arogenate dehydrogenase and prephenate dehydrogenase within the $NADP^+$ lineage of superfamily B have been found to be sensitive to feedback inhibition in every case evaluated thus far.

If the ancestral superfamily B pathway is regarded as an intermediate evolutionary state, a still earlier ancestral pathway can be estimated, as illustrated at the top of figure 6. The inference of this pathway is speculative, but when similar estimations become available for superfamilies A and C, a plausible ancestral pathway at the evolutionary stem of divergence of all three superfamilies will be deducible. The sole point of regulation shown in figure 6 (*top*) is at DAHP synthase. In organisms where allosteric regulation of aromatic biosynthesis is relatively simple (e.g., in cyanobacteria), early pathway control is commonly the sole or major regulation in force (Jensen and Hall 1982). If, as initial evolutionary events, *L*-tyrosine and *L*-tryptophan came to be utilized as end-product signals for early pathway control, then a selective pressure may have been created favoring preferential channeling of chorismate to *L*-phenylalanine. This would ensure that excess tyrosine

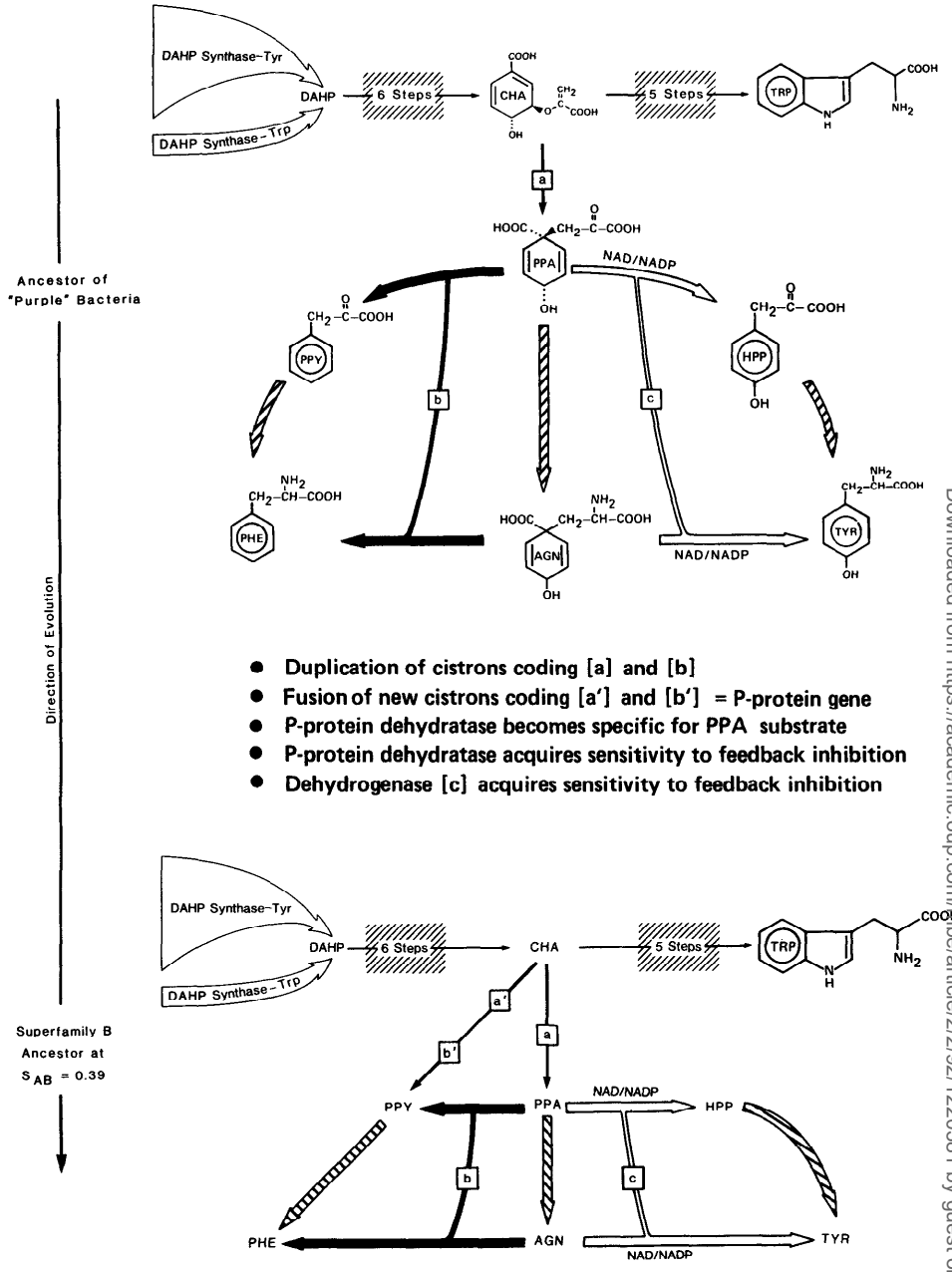

FIG. 6.—Aromatic biosynthesis and regulation deduced to exist in the ancestral superfamily B organism (bottom diagram) and an earlier hypothetical ancestor, showing the structures of chorismate (CHA), prephenate, (PPA), phenylpyruvate (PPY), 4-hydroxyphenylpyruvate (HPP), *L*-arogenate (AGN), *L*-phenylalanine (PHE), *L*-tyrosine (TYR), and *L*-tryptophan (TRP). Enzyme denotations: a = chorismate mutase, b = dehydratase exhibiting broad specificity for prephenate or *L*-arogenate, and c = dehydrogenase exhibiting broad specificity for prephenate or *L*-arogenate and able to utilize either NAD^+ or $NADP^+$. The striped arrows indicate aminotransferase reactions that were catalyzed by a single enzyme. (Many contemporary aminotransferase enzymes are still broadly specific for the three substrates shown (Jensen and Hall 1982). Evolutionary steps intervening between the two evolutionary states are specified in the middle section between the two diagrams.

and tryptophan could not produce early pathway regulation that might provide starvation for phenylalanine. This may account for the ubiquity of the anciently evolved P-protein, in contrast to the rarity of the recently evolved T-protein.

The events during the evolutionary transition to the superfamily B ancestor that are pictured in figure 6—that is, duplication of the ancestral cistrons encoding chorismate mutase and the dehydratase, followed by gene fusion to account for the bifunctional enzyme—are those that would explain the evolution of the P-protein. The dehydratase component of the P-protein became specific for prephenate, and sensitivity to feedback inhibition evolved. The ancestral dehydrogenase, still broadly specific for either cofactor and either of the cyclohexadienyl substrates, probably evolved sensitivity to *L*-tyrosine-mediated feedback inhibition at this time.

Prospects for an Expanded Perspective on Aromatic-Pathway Evolution

The Major *L*-Tryptophan Branch

Obviously the tryptophan pathway must have coevolved in a very specific relationship with the phenylalanine and tyrosine pathways. Considerable biochemical and genetic diversity of tryptophan biosynthesis is indeed known to exist in nature (Crawford 1975). Of those organisms within superfamily B that have been studied thus far, extensive differences in tryptophan-pathway regulation and enzyme organization have been demonstrated, for example, in *Escherichia coli* vis-à-vis *Pseudomonas aeruginosa* (a group Ib pseudomonad). Among the enteric organisms, *E. coli* and *Klebsiella pneumoniae* possess an organized complex of anthranilate synthase and phosphoribosyl transferase, whereas *Serratia, Proteus,* and *Aeromonas* species lack this enzyme complex. These results are in pleasing correspondence with the hierarchical evolutionary relationships of these organisms, as shown in figure 2. Inclusion of tryptophan-pathway studies in parallel with the study of phenylalanine and tyrosine biosynthesis would expand the base of information regarding a larger biochemical unit.

NAD^+ Biosynthesis

A larger evolutionary insight with respect to aromatic biosynthesis will be possible through studies of vitamin-like compounds that originate from the aromatic pathway. Thus, *Xanthomonas pruni* (a group V pseudomonad) uses the aerobic tryptophan catabolic pathway for NAD^+ biosynthesis, in contrast to *E. coli,* which uses the anaerobic dihydroxyacetone phosphate-aspartate pathway (Foster and Moat 1980). Hence, NAD^+ biosynthesis is directly linked to aromatic biosynthesis in one superfamily B organism but not in another.

PABA Synthase

Different protein-protein arrangements of the two aromatic-pathway amidotransferases (anthranilate synthase and PABA synthase) are known to exist within superfamily B (see Byng et al. 1982). Both enzymes use an identical pair of substrates and have been hypothesized (Jensen 1976) to have originated from common ancestral genes encoding the aminase and glutamine-binding subunits. Recent results in the study of *E. coli* have indeed established homologous proteins: the nucleotide sequences of *pabA* and *trp(G)D*, which encode the aminase subunits, were 53% identical at the nucleotide level (Kaplan and Nichols 1983); the nucleotide sequences of *pabB* and *trpE*, which encode glutamine-binding subunits, were 40% homologous at the nucleotide level (Goncharoff and Nichols 1984).

Escherichia coli exhibits a unique arrangement in which the glutamine-binding protein for anthranilate synthase (amidotransferase) is fused to the second tryptophan-pathway enzyme. In *P. aeruginosa* (a group Ib pseudomonad), separate glutamine-binding subunits apparently exist for association with the aminase subunits of anthranilate synthase and PABA synthase. In *Acinetobacter calcoaceticus,* a single glutamine-binding subunit is shared by both amidotransferase activities. The latter system would conform most closely to an ancestral state in which a single pair of subunits catalyzed two reactions, with anthranilate likely being the major product and PABA being a minor (vitamin) product. A feasible evolutionary progression would be the following: (1) gene duplication of the ancestral aminase, (2) increased specialization of the two reactions, (3) gene duplication of the glutamine-binding subunit, and (4) gene fusion. Step (2) yields *A. calcoaceticus;* step (3) yields *P. aeruginosa,* and step (4) yields *E. coli.*

Vitamin K

Emerging data concerning the vitamin K pathway (Bentley and Meganathan 1982), another aromatic-pathway branch, suggests possibilities for still further biochemical diversity with respect to aromatic biosynthesis.

Connecting Metabolic Networks

Although biochemical pathways are commonly treated as separate entities, they are in fact part of a metabolic whole. The use of a pathway for aromatic amino acid biosynthesis may vary greatly in different organisms because of effects of secondary metabolism or catabolism in connecting pathways. If *L*-tryptophan is the precursor of great amounts of pigment (e.g., violacein in *Chromobacterium*), then the relative output of tryptophan will be much greater than in an organism like *Escherichia coli.* Microorganisms are highly diverse with respect to the use of aromatic amino acids for pigments, antibiotic peptides, and other secondary metabolites. It seems inescapable that a complete evolutionary appreciation of a given pathway must ultimately deal with connecting biochemical networks.

The distribution in nature of a particular metabolic ability is of interest. Until recently the important mammalian enzyme phenylalanine hydroxylase, which converts phenylalanine directly to tyrosine, was known only in *Pseudomonas sp.* ATCC 11299a (Guroff and Ito 1963), a strain which we (Berry et al. 1985) have shown to be *P. acidovorans.* This enzyme has recently been found in *P. facilis* (Friedrich and Schlegel 1972), *Alcaligenes eutrophus* (Friedrich and Schlegel 1972), and *C. violaceum* (Letendre et al. 1974), all of which are members of superfamily A. The full range of its distribution remains to be determined.

Other Biochemical Pathways

The branched pathway for biosynthesis of the aspartate family of amino acids is as complex in construction as the aromatic amino acid pathway and therefore exemplifies another system offering comparable potential for probing evolutionary relationships. Comparative studies done prior to the availability of phylogenetic trees (Cohen et al. 1969) indicate a degree of biochemical diversity approaching that found in the aromatic amino acid pathway. Entirely parallel evolutionary studies of two or more such systems in oligonucleotide-cataloged organisms would provide the beginnings of information that would allow interpretations of evolutionary

events that cross specific pathway boundaries. For example, the pyridine nucleotide cofactor domain of dehydrogenases is generally considered to be of ancient origin. Would the cofactor specificity for homoserine dehydrogenase in superfamily B organisms parallel that shown (fig. 1) for aromatic-pathway dehydrogenases? If so, what about dehydrogenase activities operating in other pathways?

New prospects for deducing evolutionary histories of biochemical pathways are by no means limited to branched pathways. Variation in any biochemical pathway is amenable to evolutionary conclusions when considered in the context of an acceptable phylogeny. An excellent example is the evolution of cell wall biosynthesis. A significant amount of biochemical data has been obtained to characterize many distinctive peptidoglycan types in microorganisms, and these results were obtained in oligonucleotide-cataloged strains. An example of the emerging evolutionary picture within one large microbial grouping is given in figure 3 of Stackebrandt and Woese (1981). Apparently, little information is yet available concerning the enzymological diversity that accounts for the observed diversity of peptidoglycan makeup.

Regulatory interactions are exerted between different biochemical pathways at the levels of both feedback inhibition (Jensen 1969) and transcriptional control (Bogosian and Somerville 1983). Such interlocking relationships have been difficult to study because they are subtle and not readily anticipated. Such relationships can now be probed in closely related microorganisms, since oligonucleotide-cataloging data provide a rational basis for strain selection. Hence, it should now be possible to study biochemical evolution at a level extending beyond the boundaries of a given pathway. Indeed, to the extent that the evolution of more and more particular pathways are deciphered in prokaryotes, the greater the prospects for the global evolutionary insights that will approach an ultimate understanding of biochemical evolution at the cell level.

Acknowledgments

I wish to acknowledge the intellectual input and benchwork contributions of R. Whitaker and A. Berry and to note especially the development of many ideas by G. S. Byng.

LITERATURE CITED

BACHMANN, B. J. 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.

BALDWIN, G. S., G. H. MCKENZIE, and B. E. DAVIDSON. 1981. The self-association of chorismate mutase/prephenate dehydratase from *Escherichia coli* K-12. Arch. Biochem. Biophys. **211**:76–85.

BENTLEY, R., and R. MEGANATHAN. 1982. Biosynthesis of vitamin K (menaquinone) in bacteria. Microbiol. Rev. **46**:241–280.

BERRY, A., J. L. JOHNSON, and R. A. JENSEN. 1985. Phenylalanine hydroxylase and isozymes of 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthase in relationship to the phylogenetic position of *Pseudomonas acidovorans* (*Ps.* sp. ATCC 11299a). Arch. Microbiol. (accepted).

BOGOSIAN, G., and R. SOMERVILLE. 1983. *Trp* repressor protein is capable of intruding into other amino acid biosynthetic systems. Mol. Gen. Genet. **191**:51–58.

BYNG, G. S., A. BERRY, and R. A. JENSEN. 1983*a*. A pair of regulatory enzymes for 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthase is conserved within group I pseudomonads. J. Bacteriol. **156**:429–433.

BYNG, G. S., and R. A. JENSEN. 1983. Impact of isozymes upon partitioning of carbon flow and regulation of aromatic biosynthesis in prokaryotes. Pp. 115–140 *in* M. C. RATAZZI, J. G. SCANDALIOS, and G. S. WHITT, eds. Isozymes. Vol. **8**. Liss, New York.

BYNG, G. S., J. L. JOHNSON, R. J. WHITAKER, R. L. GHERNA, and R. A. JENSEN. 1983*b*. The evolutionary pattern of aromatic amino acid biosynthesis and the emerging phylogeny of pseudomonad bacteria. J. Mol. Evol. **19**:272–282.

BYNG, G. S., J. F. KANE, and R. A. JENSEN. 1982. Diversity in the routing and regulation of complex biochemical pathways as indicators of microbial relatedness. Crit. Rev. Microbiol. **9**:227–252.

BYNG, G. S., R. J. WHITAKER, R. L. GHERNA, and R. A. JENSEN. 1980. Variable enzymological patterning in tyrosine biosynthesis as a means of determining natural relatedness among the *Pseudomonadaceae*. J. Bacteriol. **144**:247–257.

BYNG, G. S., R. J. WHITAKER, and R. A. JENSEN. 1983*c*. Evolution of *L*-phenylalanine biosynthesis in rRNA homology group I of pseudomonads. Arch. Microbiol. **136**:163–168.

CAMPBELL, A. 1981. Evolutionary significance of accessory DNA elements in bacteria. Annu. Rev. Microbiol. **35**:55–83.

COHEN, G. N., R. Y. STANIER, and G. LEBRAS. 1969. Regulation of the biosynthesis of amino acids of the aspartate family in coliform bacteria and pseudomonads. J. Bacteriol. **99**:791–801.

COTTON, R. G. H., and F. GIBSON. 1965. The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationship to prephenate dehydratase and prephenate dehydrogenase. Biochim. Biophys. Acta **100**:76–88.

CRAWFORD, I. P. 1975. Gene rearrangements in the evolution of the tryptophan synthetic pathway. Microbiol. Rev. **39**:87–120.

DAVIES, W. D., and B. E. DAVIDSON. 1982. The nucleotide sequence of *aroG*, the gene for 3-deoxy-*D*-arabinoheptulosonate-7-phosphate synthetase (phe) in *Escherichia coli* K-12. Nucleic Acids Res. **10**:4045–4058.

FISKE, M. J., R. J. WHITAKER, and R. A. JENSEN. 1983. A hidden overflow pathway to *L*-phenylalanine in *Pseudomonas aeruginosa*. J. Bacteriol. **154**:623–631.

FOSTER, W., and A. G. MOAT. 1980. Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. Microbiol. Rev. **44**:83–105.

FOX, G. E., K. R. PECHMAN, and C. R. WOESE. 1977. Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics. Int. J. Syst. Bacteriol. **27**:44–57.

FOX, G. E., E. STACKEBRANDT, R. B. HESPELL, J. GIBSON, J. MANILOFF, T. A. DYER, R. S. WOLFE, W. E. BALCH, R. S. TANNER, L. J. MAGRUM, L. B. ZABLEN, R. BLAKEMORE, R. GUPTA, L. BONEN, B. J. LEWIS, D. A. STAHL, K. R. LUEHRSEN, K. N. CHEN, and C. R. WOESE. 1980. The phylogeny of prokaryotes. Science **209**:457–463.

FRIEDRICH, B., and H. G. SCHLEGEL. 1972. Die hydroxylierung von Phenylalanin durch *Hydrogenomonas eutropha* H16. Arch. Mikrobiol. **83**:17–31.

GONCHAROFF, P., and B. P. NICHOLS. 1984. Nucleotide sequence of *Escherichia coli pabB* indicates a common evolutionary origin of *p*-aminobenzoate synthetase and anthranilate synthetase. J. Bacteriol. **159**:57–62.

GUROFF, G., and T. ITO. 1963. Induced, soluble phenylalanine hydroxylase from *Pseudomonas* sp. grown on phenylalanine or tyrosine. Biochim. Biophys. Acta 77:159–161.

HALL, B. G., S. YOKOYAMA, and D. H. CALHOUN. 1983. Role of cryptic genes in microbial evolution. Mol. Biol. Evol. **1**:109–124.

HALL, G. C., M. B. FLICK, R. L. GHERNA, and R. A. JENSEN. 1982. Biochemical diversity for biosynthesis of aromatic amino acids among the cyanobacteria. J. Bacteriol. **149**:65–78.

JENSEN, R. A. 1969. Metabolic interlock: regulatory interactions exerted between biochemical pathways. J. Biol. Chem. **244**:2816–2823.

———. 1970. Taxonomic implications of temperature dependence of the allosteric inhibition of 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthetase in *Bacillus*. J. Bacteriol. **102**:489–497.

———. 1976. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. **30**:409–425.

JENSEN, R. A., and G. C. HALL. 1982. Endo-oriented control of pyramidally arranged metabolic branchpoints. Trends Biochem. Sciences **7**:177–180.

JENSEN, R. A., and D. S. NASSER. 1968. Comparative regulation of isoenzymic DAHP synthetases in microorganisms. J. Bacteriol. **95**:188–196.

JENSEN, R. A., D. S. NASSER, and E. W. NESTER. 1967. The comparative control of a branch-point enzyme in microorganisms. J. Bacteriol. **94**:1582–1593.

JENSEN, R. A., and J. L. REBELLO. 1970. Comparative allostery of microbial enzymes at metabolic branchpoints: evolutionary implications. Pp. 105–121 *in* C. J. CORUM, ed. Developments in Industrial Microbiology. Vol. **11**. Plenum, New York.

JENSEN, R. A., and S. L. STENMARK. 1970. Comparative allostery of 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthetase as a molecular basis for classification. J. Bacteriol. **101**:763–769.

JENSEN, R. A., and R. TWAROG. 1972. Allostery of 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthetase in *Clostridium:* another conserved generic characteristic. J. Bacteriol. **111**:641–648.

KAPLAN, J. B., and B. P. NICHOLS. 1983. Nucleotide sequence of *Escherichia coli pabA* and its evolutionary relationship to *trp(G)D*. J. Mol. Biol. **168**:451–468.

LAWTHER, R., D. CALHOUN, C. ADAMS, C. HAUSER, J. GRAY, and G. HATFIELD. 1981. Molecular basis of valine resistance in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **78**:922–925.

LETENDRE, G. H., G. DICKENS, and G. GUROFF. 1974. The tryptophan hydroxylase of *Chromobacterium violaceum.* J. Biol. Chem. **249**:7186–7191.

PALLERONI, N. J. 1983. The taxonomy of bacteria. Bioscience **33**:370–377.

PALLERONI, N. J., R. KUNISAWA, R. CONTOPOULOU, and M. DOUDOROFF. 1973. Nucleic acid homologies in the genus *Pseudomonas.* Int. J. Syst. Bacteriol. **23**:333–339.

PATEL, N., S. STENMARK-COX, and R. A. JENSEN. 1978. Enzymological basis of reluctant auxotrophy for phenylalanine and tyrosine in *Pseudomonas aeruginosa.* J. Biol. Chem. **253**:2972–2978.

STACKEBRANDT, E., and C. R. WOESE. 1981. The evolution of prokaryotes. Pp. 1–32 *in* M. J. CARLILE, J. F. COLLINS, and B. E. B. MOSELEY, eds. Molecular and cellular aspects of microbial evolution. 32d Symposium of the Society for General Microbiology. Cambridge University Press, London.

STENMARK, S. L., D. L. PIERSON, G. I. GLOVER, and R. A. JENSEN. 1974. Blue-green bacteria synthesize *L*-tyrosine by the pretyrosine pathway. Nature **247**:290–292.

WHITAKER, R. J., A. BERRY, G. S. BYNG, M. J. FISKE, and R. A. JENSEN. 1985. Clues from *Xanthomonas campestris* about the evolution of aromatic biosynthesis and its regulation. J. Mol. Evol. (accepted).

WHITAKER, R. J., G. S. BYNG, R. L. GHERNA, and R. A. JENSEN. 1981*a*. Comparative allostery of 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthase as an indicator of taxonomic relatedness in pseudomonad genera. J. Bacteriol. **145**:752–759.

———. 1981*b*. Diverse enzymological patterns of phenylalanine biosynthesis in pseudomonads are conserved in parallel with deoxyribonucleic acid homology groupings. J. Bacteriol. **147**:526–534.

WHITAKER, R. J., M. J. FISKE, and R. A. JENSEN. 1982. *Pseudomonas aeruginosa* possesses two novel regulatory isozymes of 3-deoxy-*D-arabino*-heptulosonate 7-phosphate synthase. J. Biol. Chem. **257**:12789–12794.

WOESE, C. R., P. BLANZ, R. B. HESPELL, and C. M. HAHN. 1982. Phylogenetic relationships among various helical bacteria. Curr. Microbiol. **7**:119–124.

ZUCKERKANDL, E., and L. PAULING. 1965. Molecules as documents of evolutionary history. J. Theor. Biol. **8**:357–366.

MASATOSHI NEI, reviewing editor

Received April 27, 1984; revision received September 10, 1984.