

Evolution of Vitamin B₆ (Pyridoxine) Metabolism by Gain and Loss of Genes

Tsuyoshi Tanaka, Yoshio Tateno, and Takashi Gojobori

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Japan

Vitamin B₆ (VB₆) functions as a cofactor of many diverse enzymes in amino acid metabolism. Three metabolic pathways for pyridoxal 5'-phosphate (PLP; the active form of VB₆) are known: the de novo pathway, the salvage pathway, and the fungal type pathway. Most unicellular organisms and plants biosynthesize VB₆ using one or two of these three biosynthetic pathways. However, animals such as insects and mammals do not possess any of the pathways and, thus, need to intake VB₆ in their diet to survive. It is conceivable that breakdowns of these pathways occurred in the evolutionary lineages of insects and mammals, and one of the major reasons for this would be the loss of pertinent genes. We studied the evolution of VB₆ biosynthesis from the view of the gain and loss of 10 pertinent genes in 122 species whose genome sequences were completely determined. The results revealed that each gene in the pathways was lost more than once in the entire evolutionary lineages of the 122 species. We also found the following three points regarding the evolution of PLP biosynthesis: (1) the breakdown of the PLP biosynthetic pathways occurred independently at least three times in animal lineages, (2) the de novo pathway was formed by the generation of *pdxB* in γ -proteobacteria, and (3) the order of the gene loss in VB₆ metabolism was conserved among different evolutionary lineages. These results suggest that the evolution of VB₆ metabolism was subject to gains and frequent losses of related genes in the 122 species examined. This dynamic nature of the evolutionary changes must have been responsible for the breakdowns of the pathways, resulting in profound differentiation of heterotrophy among the species.

Introduction

Vitamin B₆ (VB₆) functions as a cofactor of many enzymes. In particular, pyridoxal 5'-phosphate (PLP), which is the active form of VB₆, has multiple roles as a versatile cofactor of enzymes that are mainly involved in the metabolism of amino acid compounds (Grogan 1988; Rottmann et al. 1991; Helmreich 1992; Mihara et al. 1997; Kack et al. 1999). Moreover, VB₆ appears to play an important role against photosensitization in fungi (Bilski et al. 2000). Most unicellular organisms and plants biosynthesize PLP by themselves.

Many studies of VB₆ metabolism have been conducted in *Escherichia coli* (White and Dempsey 1970; Lam and Winkler 1990; Zhao et al. 1995; Yang, Zhao, and Winkler 1996; Man, Zhao, and Winkler 1996) and fungi such as *Cercospora nicotianae*, *Neurospora crassa*, *Aspergillus nidulans*, and *Saccharomyces cerevisiae* (Ehrenshaft et al. 1999; Osmani, May, and Osmani 1999; Bean et al. 2001; Ehrenshaft and Daub 2001). For PLP biosynthesis, three pathways have been characterized, and a total of 10 genes are involved in the pathways in bacteria and fungi (fig. 1). In the case of *E. coli*, both de novo and salvage pathways have been identified. These two pathways include enzymes encoded by eight genes in total (Mittenhuber 2001) and share only one gene, *pdxH*. In the de novo pathway, the pyridine ring of VB₆ is generated from D-erythrose-4-phosphate (E4P) and glyceraldehyde-3-phosphate. On the other hand, in the salvage pathway, PLP is synthesized without pyridine ring generation. A corresponding de novo pathway has not been discovered in fungi or plants. Instead, fungi were found to have another biosynthetic pathway, the fungal type pathway. This pathway has two genes, *SNZ* and *SNO*, whose functions are currently unknown. Therefore, species that synthesize

PLP have been reported to have at least one of the three PLP biosynthetic pathways.

Animals such as insects and mammals have to intake VB₆ compounds in their diet. In particular, humans intake VB₆ in meats and vegetables (Manore 2000). Animal lineages, therefore, have a problem regarding VB₆ metabolism, and it is conceivable that their PLP biosynthetic pathways became dysfunctional because of the loss of some of the pathway genes during their evolution, assuming that the ancestor of animals had the PLP biosynthetic pathways. In fact, *pdxA* and *pdxJ*, which are two of the above-mentioned 10 genes, have not been found in animals (Ehrenshaft et al. 1999). Moreover, *SNZ* and *SNO* have been reported to be lost in animals, except for the marine sponge *Suberites domuncula* (Seack et al. 2001). Even though the functions of *SNZ* and *SNO* are unknown, they should have an indispensable role in the biosynthetic pathway in fungi. The loss of these two genes has been considered as a cause of the inability of animal lineages to biosynthesize PLP, especially the Eumetazoan lineage.

To study the evolutionary process of VB₆ metabolism, we investigated the 10 genes involved in the three pathways for PLP biosynthesis. We were particularly interested in learning when the individual genes were gained or lost during evolution. Therefore, we focused on the gain and loss of these 10 genes in 122 species in the three domains of life, namely eubacteria, archaeobacteria, and eukaryotes (Woese, Kandler, and Wheelis 1990). We estimated the pertinent gene set of the common ancestor of the 122 species on the basis of their genealogical relationships. Next, we identified the gain and loss events for the genes by comparing the gene sets between the ancestral and extant species. On the basis of the results obtained, we report the evolutionary features of the formation and dysfunction processes of VB₆ metabolism from the view of the gains and losses of the genes.

Materials and Methods

Genes Related to VB₆ Metabolism

There are 10 genes involved in the three PLP biosynthetic pathways: glyceraldehyde 3-P dehydrogenase A (*gapA*),

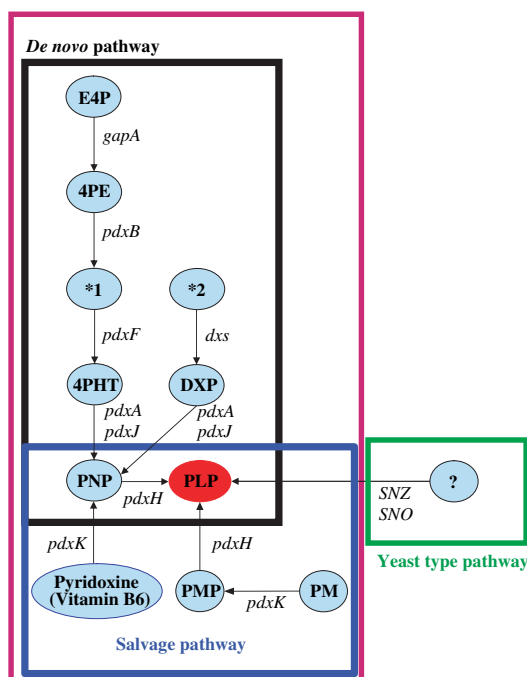
Key words: vitamin B₆, metabolic network, gene gain, gene loss, complete genome.

E-mail: tgojobor@genes.nig.ac.jp.

Mol. Biol. Evol. 22(2):243–250. 2005

doi:10.1093/molbev/msi011

Advance Access publication October 13, 2004



Pathways discovered in bacteria Pathway discovered in yeast

FIG. 1.—PLP biosynthetic pathways. Circles indicate the substrates. PLP is the active form of VB6. *1 indicates 3-hydroxy-4-phosphohydroxy- α -ketobutyrate; *2 indicates glyceraldehyde-3-phosphate.

deoxy-xylulose-P synthase (*dxs*), 4-hydroxythreonine-4-phosphate dehydrogenase (*pdxA*), erythronate-4-phosphate dehydrogenase (*pdxB*), phosphoserine aminotransferase (*pdxF*), pyridoxine-phosphate oxidase (*pdxH*), pyridoxal phosphate biosynthetic protein (*pdxJ*) and pyridoxal kinase (*pdxK*) in *E. coli*, and *SNZ* and *SNO* in yeast (Yang et al. 1998a, 1998b; Mittenhuber 2001). To identify the PLP biosynthetic pathway genes in the complete genome sequences of 122 species (nine eukaryotes, 16 archaeobacteria, and 97 eubacteria) (appendix A of Supplemental Material online), BlastP homology searches for the 10 genes were performed against each set of their predicted protein sequences (E value < 10^{-5}) (Altschul et al. 1990). As query sequences, we used the protein sequences of *gapA* (DDBJ/EMBL/GenBank accession number AAC74849), *dxs* (accession number AAC73523), *pdxA* (accession number AAC73163), *pdxB* (accession number AAC75380), *pdxF* (accession number AAC73993), *pdxH* (accession number AAC74710), *pdxJ* (accession number AAC75617), and *pdxK* (accession number AAC75471) derived from *E. coli* K-12 MG1655 and those of *SNZ* (NP_013814) and *SNO* (NP_013813) derived from *S. cerevisiae*. Next, we compared our results with the KEGG Orthology (KO) data set in the KEGG database (Bono et al. 1998; Kanehisa et al. 2002) for confirmation. The KO data set contains orthologous gene families that are categorized on the basis of experimental information, sequence homology, and gene order in the genome. If a gene were found to be related to metabolic reactions other than the PLP biosynthetic pathways, we removed it from our data set. Finally, if any of the 10 genes could not be found by the homology search against the set of predicted protein sequences of

a particular species, we conducted a TblastN search of its genome sequence (E value < 10^{-5}). If there was no homologous sequence to the query sequence in the genome, we concluded that the gene was absent from the species.

Phylogenetic Tree

To estimate the gene sets involved with the PLP biosynthetic pathways of ancestors of the 122 species, we basically employed the eukaryotic lineages of Baldauf et al. (2000) and the eubacterial and archaeobacterial lineages of Nelson et al. (2000). For missing species in the eubacterial and archaeobacterial lineages, such as proteobacteria, firmicutes, actinobacteria, chlamydia, spirochete, cyanobacteria, euryarchaeota, and crenarchaeota, we constructed their phylogenetic trees for the same gene (16s rRNA) as that used by Nelson et al. (2000). To do this, we applied their 16s rRNA sequences to the ClustalW program with 1,000 bootstrap trials (Thompson, Higgins, and Gibson 1994). We excluded the positions with gaps and corrected for multiple substitutions.

Estimation of the Gene Set of the PLP Biosynthetic Pathways in the Ancestor

We assumed that a single gene was acquired only once during the evolution of the 122 species examined in this model and ignored horizontal gene transfer and parallel evolution. The method for estimating the set of genes in the ancestor was as follows. As shown in figure 2, we used the following two states to represent whether a species had a particular gene (Gene-A): “1” was used when the species had at least one homologous gene to Gene-A, and “0” was used when the species had no homologous genes to Gene-A.

If the states of two species were the same, namely (0,0) or (1,1), we assumed that the ancestor had the same state (patterns 1 and 2 in figure 2). If the state was different between the two closest species compared, namely (0,1) or (1,0), we used the states for all the other species as outgroup species. For example, let us designate a group of these species as group C. If at least one species in group C had the state of “1,” we regarded the ancestral status as “1” (pattern 3 in figure 2). If all the species in group C had the status of “0,” we regarded the ancestral state as “0” (pattern 4 in figure 2). In this way, we estimated the states for all the given genes at all internal nodes.

The Order of the Losses Among the Genes

Using the results for the estimation of the gene losses, we investigated the order of the losses among the 10 genes. If the loss of a gene occurred randomly during the evolution of the PLP biosynthetic pathways, skewness of the early loss of either one of the two genes may not be observed (null hypothesis). Therefore, we considered that the frequency of the order of gene loss between two genes followed the binominal distribution. For all pairs among the 10 genes, we examined which gene was lost first during the evolution of the 122 species and statistically tested the frequency of the order of gene loss on the basis

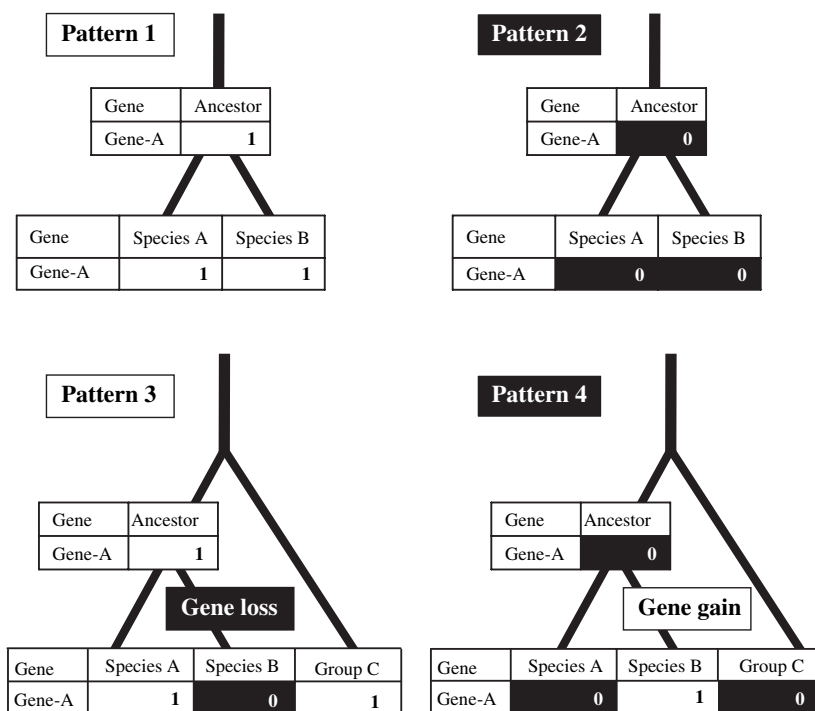


FIG. 2.—A model for estimating the ancestral gene set by comparing the gene sets for two species. In this figure, we used the following two states to represent whether the species had Gene-A: “1” was used when the species had at least one homologous gene to Gene-A, and “0” was used when the species had no homologous genes to Gene-A.

of the binominal distribution. If the losses of two genes occurred on the same branch in the phylogenetic tree, we did not count them, because we did not know which gene was lost first.

Results

Comparison of the Gene Sets Among Species

Three PLP biosynthetic pathways have been reported as described earlier: the de novo and salvage pathways in *E. coli* and the fungal-type pathway. When we examined whether the 10 genes existed in the 122 species whose genome sequences were completely or almost completely determined, we found that no species had all the genes (appendix A of Supplemental Material online). In the cases of *gapA*, *pdxA*, and *pdxK*, we found several homologous sequences among the species. When we constructed the phylogenetic trees for the 10 genes, we found that duplication of the genes existing in each gene family may have occurred not only in the particular lineage recently but also in the ancestral lineage (data not shown). This observation suggests that each of the 10 gene families, especially *gapA*, *pdxA*, and *pdxK*, evolved from the common ancestral sequence by gene duplication and gene loss. However, we could not conclude that all the genes listed were involved in VB₆ synthesis, because there was no evidence for their functions in the experimental results.

Moreover, we found that gene sets containing all seven genes for the de novo pathway (*gapA*, *dxs*, *pdxA*, *pdxB*, *pdxF*, *pdxH*, and *pdxJ*) were only found in 18 eubacteria. These bacterial species were all γ -proteobacteria,

indicating that the de novo pathway may only function in γ -proteobacteria.

Distribution of the Genes in the Three Domains of Life

Based on the gene sets of the 122 species, we compared the gene sets among the three domains of life (fig. 3). *SNZ*, *SNO*, and *pdxF* were observed in all the domains, indicating that they existed before the divergence into the three domains. This result also indicates that the fungal type pathway composed of *SNZ* and *SNO* was formed before the divergence into the three domains. The other four genes in the de novo and salvage pathways (*gapA*, *dxs*, *pdxH*, and *pdxK*) were discovered in eukaryotes and eubacteria, indicating that part of these pathways was present in eukaryotes and eubacteria. Based on our assumption that a single gene was acquired only once during the evolution of the 122 species examined, we interpret this result to indicate that the salvage pathway composed of *pdxH* and *pdxK* was formed before the divergence into the three domains. We also note that *pdxA*, *pdxB*, and *pdxJ* were only observed in eubacteria. When we focused on the eubacteria, we found that both *pdxA* and *pdxJ* were observed not only in proteobacteria but also firmicutes, cyanobacteria, chlorobi, and aquificae, whereas *pdxB* only existed in γ -proteobacteria. Therefore, we consider that both *pdxA* and *pdxJ* were generated in the eubacterial lineage after the divergence from the other two domains and that *pdxB* was generated in the γ -proteobacterial lineage after the divergence from the other lineages. These three genes are, therefore, considered to have contributed to the formation of the de novo pathway in the eubacterial lineage. In particular, *pdxB* may be the most important

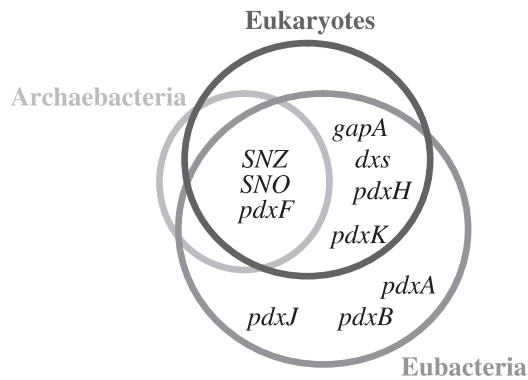


FIG. 3.—Distribution of the 10 genes related to the PLP biosynthetic pathways.

gene for the completion of this pathway because it was only generated in γ -proteobacteria.

Estimation of the Losses of the Genes

Because the gene sets of the PLP biosynthetic pathways were different among the 122 species examined, it was considered that loss as well as gain of genes had occurred for the 10 genes. We, therefore, examined how many losses of the 10 genes had occurred during the evolution of the 122 species. To estimate the occurrence of the gain or loss of a particular gene in the evolutionary lineage from the common ancestor to each extant species, we needed to estimate whether the gene had existed in the common ancestor. Using the phylogenetic tree, we estimated the ancestral gene set at each node based on the assumption that each of the genes was acquired only once during the evolution of the entire 122 species. As a result, we found that a total of 132 gene losses had occurred during the evolution of the 122 species from the common ancestor (table 1). The numbers of losses of *pdxB*, *gapA*, and *pdxJ* were 2, 3, and 8, respectively, and all smaller than the numbers of losses of the other seven genes. In the case of *gapA*, the lower number of gene losses may be explained by the following functional constraint: *GapA* functions not only in the PLP biosynthetic pathways but also in glycolysis (Seta et al. 1997), and, therefore, *gapA* is not expected to be lost because of functional constraints. Because we estimated that *pdxB* had emerged in the γ -proteobacterial lineage, we did not need to count the number of losses of the gene before the divergence of this lineage from the other lineages, and, thus, the total number of losses of this gene was smaller than those of the other genes. In the case of *pdxJ*, we estimated that this gene emerged in the eubacterial lineage and was lost during an early period in the particular lineages of the 97 eubacterial species. The total number of losses of *pdxJ* was, therefore, smaller than those of the other genes, because it was only present in the lineages for a comparatively short time.

Losses of *SNO* and *SNZ* in Animal Lineages

The *SNO* and *SNZ* genes have only been reported in the sponge *S. domuncula* among Metazoa. By homology

Table 1
Numbers of Losses of the Genes

Gene Name	Number of Species	Number of Gene Losses
<i>dxs</i>	76	10
<i>gapA</i>	103	3
<i>pdxA</i>	46	13
<i>pdxB</i>	18	2
<i>pdxF</i>	67	24
<i>pdxH</i>	52	14
<i>pdxJ</i>	42	8
<i>pdxK</i>	45	20
<i>SNO</i>	46	20
<i>SNZ</i>	48	18

searches against the genome sequences of six animals (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Ciona intestinalis*, *Caenorhabditis elegans*, and *Drosophila melanogaster*), we found that these two genes also existed in the genome of *C. intestinalis*. The positions of the two genes in the genome of *C. intestinalis*, in which they aligned head-to-head, were the same as those in *S. domuncula* and *S. cerevisiae* (fig. 4). However, the two genes were aligned head-to-tail in the genomes of 33 eubacteria and archaeobacteria in which *SNO* was adjacent to *SNZ*. Moreover, when we conducted BlastP homology searches among *SNZ* or *SNO* protein sequence families, the sequences in *C. intestinalis* and *S. domuncula* were reciprocally the best-hit sequences. Therefore, it is reasonable to conclude that *SNO* and *SNZ* in *C. intestinalis* were not transferred from bacterial species. From these results, we conclude that *SNO* and *SNZ* existed in the Eumetazoa lineage and that losses of *SNO* and *SNZ* occurred independently at least three times in *C. elegans*, *D. melanogaster*, and vertebrate lineages (fig. 5).

Correlation of Losses of Genes Among Pyridoxine Biosynthetic Pathways

We found that simultaneous loss of *SNO* and *SNZ* occurred more often than loss other pairs of genes (appendix B of Supplemental Material online) and also that the loss of *SNZ* occurred together with that of *SNO* in the same branch. As mentioned above, the two genes were often present next to each other in the genome. Moreover, the two genes only function in the fungal type pathway. These observations support that the losses of these two genes occurred at the same time.

Next, we examined the order of the losses for the 10 genes. When we compared the order of the losses between two genes, a significant bias was observed in nine particular combinations of genes: *pdxH* and *dxs*, *pdxJ* and *pdxH*, *pdxH* and *pdxF*, *pdxA* and *dxs*, *pdxJ* and *pdxA*, *pdxJ* and *dxs*, *pdxJ* and *pdxK*, *pdxJ* and *SNO*, and *pdxK* and *pdxF* (appendix B of Supplemental Material online). In every pair of these nine combinations, we observed that the loss of the latter gene occurred more frequently after the loss of the former gene. These biases were statistically significant against the binomial distribution ($P < 0.05$). From this observation, we deduced the patterns of losses of five genes, as shown in figure 6. The loss of *pdxJ* caused the loss of at least one of *pdxA*, *pdxH*, and *pdxK*. These

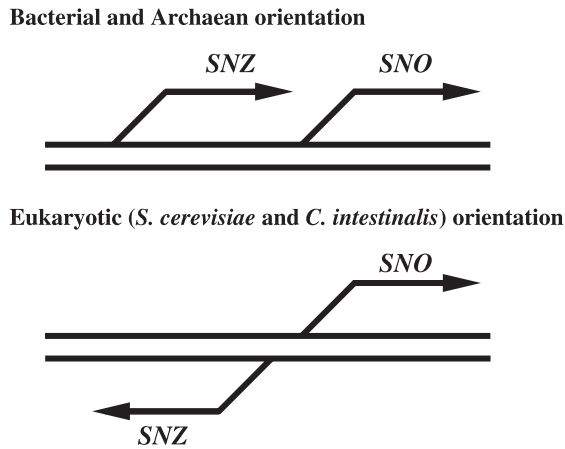


FIG. 4.—Orientations of *SNZ* and *SNO* in different genomes.

four genes encode enzymes whose reactions are connected through pyridoxine 5'-phosphate (PNP). Moreover, the losses of these four genes caused the loss of at least one of *pdxF* and *dxs*. These two genes encode enzymes whose reactions are connected to *pdxA* and *pdxJ* through 4-phosphohydroxy-L-threonine (4PHT) and 1-deoxyxylulose 5-phosphate (DXP), respectively. If we designate the two genes encoding enzymes that catalyze two consecutive reactions in the neighbor genes, we can, thus, conclude that the loss of one gene accelerates the loss of the neighbor gene.

Discussion

To our knowledge, this is the first attempt to elucidate the evolution of VB₆ metabolism by focusing on the gain and loss of the 10 genes involved in the PLP biosynthetic pathways. On the assumption that each of these genes was acquired only once during the evolution of the entire 122 species, we found that every gene in the PLP biosynthetic pathways had been lost more than once in the evolutionary lineages of the 122 species. This suggests that the breakdown of the PLP biosynthetic pathways by gene losses may have occurred in many lineages, which should be examined experimentally.

We also revealed three aspects to the evolution of the PLP biosynthetic pathways by estimating the gain and loss of the 10 genes. The first aspect is related to the evolutionary order of the generation of the three PLP biosynthetic pathways. From the distribution of the 10 genes in the 122 species examined, we found that the fungal type and salvage pathways were probably older than the *de novo* pathway on the basis of the following two results. First, *pdxK* and *pdxH* of the salvage pathway exist in eubacteria and eukaryotes, and *SNO* and *SNZ* of the fungal type pathway are found in all the three domains, namely, eubacteria, archaeobacteria, and eukaryotes. Therefore, we consider that the fungal type and salvage pathways both existed before the separation of the three domains of life. Second, *pdxB* of the *de novo* pathway only exists in γ -proteobacteria, indicating that it was generated in γ -proteobacteria after the divergence of the three domains of life.

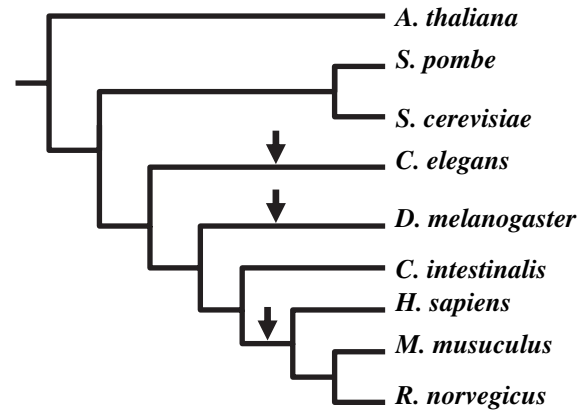


FIG. 5.—Losses of *SNO* and *SNZ* in the eukaryotic lineages. The phylogenetic tree shows the genealogical relationships among nine eukaryotes. The arrows on the branches represent the losses of *SNZ* and *SNO*.

Applying the second result mentioned above to the existing model for explaining the evolution of the metabolic networks, the patchwork model and *de novo* invention (Jensen 1976; Schmidt et al. 2003), we propose that the process for the formation of the *de novo* pathway in γ -proteobacteria was as follows (fig. 7). Originally, the common ancestor of the 97 eubacterial species studied had part of the *de novo* pathway involving five genes (*dxs*, *pdxA*, *pdxF*, *pdxH*, and *pdxJ*). Because *gapA* functioned not only in the PLP biosynthetic pathways but also in glycolysis (Seta et al. 1997), we think that this gene also existed in the common ancestor. As shown in figure 7, when *pdxB* was generated in the lineage of γ -proteobacteria, the reaction catalyzed by the product of *pdxB* was connected to the two metabolic reactions that were separately catalyzed by the products of *gapA* and *pdxF*. As a result, the *de novo* pathway was completed by the presence of the seven genes (*dxs*, *gapA*, *pdxA*, *pdxB*, *pdxF*, *pdxH*, and *pdxJ*) in γ -proteobacteria. We have, therefore, reached the same conclusion as Mittenhuber (2001) but via a different process. Mittenhuber postulated that the *de novo* pathway was largely restricted to γ -proteobacteria on the basis of the functions of *pdxA* and *pdxJ* and the requirement of VB₆ in the *de novo* pathway.

However, we could not answer which of the fungal type and salvage pathways was established as the first PLP biosynthetic pathway, because we could not estimate the times when *pdxK*, *pdxH*, *SNO*, and *SNZ* originated. In other words, we could not determine which of the gene sets of the two pathways was generated first.

The second aspect is related to the losses of *SNO* and *SNZ* in animal lineages. Among animals, these two genes have only been discovered in the marine sponge *S. domuncula* (Seack et al. 2001). Therefore, it is plausible that the losses of *SNZ* and *SNO* occurred only once in the Eumetazoan lineage after its divergence from the Poriferal lineage. We found that the two genes were present in the complete genome sequence of *C. intestinalis* in the present study. This species is more closely related to mammals than *D. melanogaster* and *C. elegans*, neither of which have the two genes (fig. 5). Therefore, we consider that *SNZ* and *SNO* existed in animals after the divergence

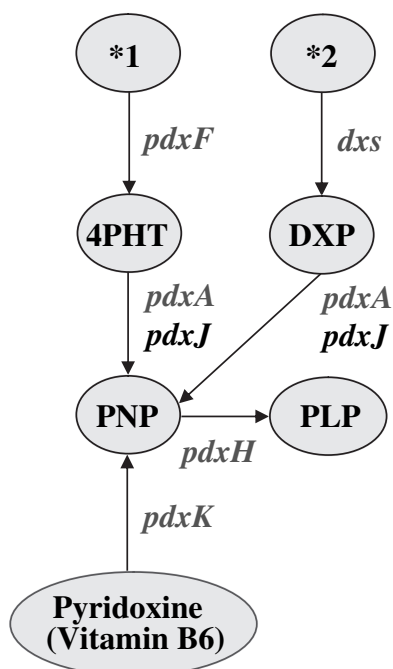
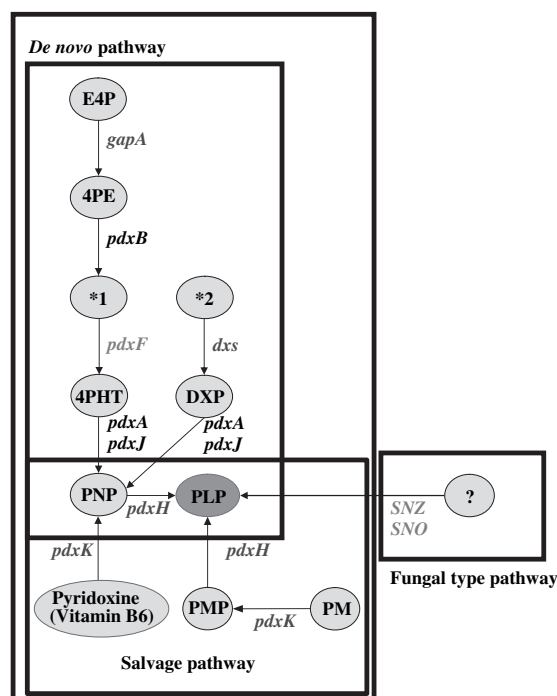


FIG. 6.—Loss of five genes in the PLP biosynthetic pathways. The pathway shown is part of the *de novo* and salvage pathways. The gene shown in black (*pdxI*) is lost first, followed by the genes shown in red (*pdxA*, *pdxH*, and *pdxK*), and then the genes shown in blue (*dxs* and *pdxF*). Circles indicate the substrates. *1 indicates 3-hydroxy-4-phosphohydroxy- α -ketobutyrate; *2 indicates glyceraldehyde-3-phosphate.

between invertebrates and vertebrates and that their losses occurred independently at least three times in the animal lineage, as shown in figure 5. We reject the possibility of horizontal gene transfer from the bacterial lineage to *C. intestinalis* not only by the homology of the two genes but also by the order and orientation of the two genes in the genome (fig. 4).

The third aspect is related to the evolutionary order of the gene loss. The losses of five genes occurred in the order shown in figure 6 during the evolution of the PLP biosynthetic pathways. Historically, five models have been proposed to explain the formation of metabolic pathways: the retrograde model, the patchwork model, *de novo* invention, specialization of a multifunctional enzyme, and pathway duplication (Horowitz 1945; Jensen 1976; Schmidt et al. 2003). However, these models only consider the gene gain. Because there are other reports that gene losses have often occurred in the metabolic pathways in bacterial lineages (Tatusov et al. 1996; Shigenobu et al. 2000), it is not sufficient to only consider the gain of genes for the evolution of metabolic pathways in bacterial lineages.

Therefore, we propose a new model based on our results that explains the evolution of metabolic pathways by gene loss. Once the loss of a gene has occurred in a metabolic pathway, the neighboring gene is more easily lost than other genes in the pathway. This can be explained by functional constraints. The breakdown of a metabolic pathway by gene loss will decrease the functional constraints on the other genes of the pathway. Our model



Pathways discovered in bacteria Pathway discovered in yeast

FIG. 7.—Formation of the PLP biosynthetic pathways. The genes shown in red and blue were the primary genes for VB6 biosynthesis. The genes shown in black were generated in eubacteria. Circles indicate the substrates. PLP is the active form of VB6.

suggests that the functional constraint on the proximal genes to the lost gene decreases more extensively than that on the distant genes. Of course, it is possible that the functional constraint is affected by other pathways. For example, if a gene is also involved in another metabolic pathway, as in the case of *gapA*, its functional constraint may not be changed.

Our approach to estimating the gain and loss of genes is affected by at least two points. The first point is the frequency of the gain and loss of genes. In this study, we considered that gene gain occurred only once, even though gene loss could have occurred more than once in the evolution of the 122 species examined. As a result, we concluded that there were seven genes in the common ancestor of the 122 species examined and that a total of 132 gene losses took place during the evolution of the 122 species. However, when we performed an estimation based on the parsimony method, there were only three genes in the common ancestor and the number of gene losses was underestimated because of overestimation of the gene gain (data not shown). These results indicate that the prediction of the gene set in the ancestor and the gain and loss of genes are clearly affected by the initial assumption.

However, we can emphasize the low possibility of gene gain for the following reasons. The cause of gene gain is mainly horizontal gene transfer or parallel evolution. Therefore, if there is a difference in the gene sets among closely related species, the number of gene losses is expected to be larger than that of gene gains. If horizontal gene transfer and parallel evolution occurred,

then gene loss would decrease and gene gain would increase. When we have evidence for horizontal gene transfer and parallel evolution of the 10 genes in this study, it will be possible to estimate the times of the gain and loss of the genes more accurately. In fact, we examined the probability of horizontal gene transfer for the 10 genes between eubacteria and archaeobacteria using the method of Nakamura et al. (2004) and concluded that the probability was negligible.

The second point is that our results are affected by the topology of the phylogenetic tree. If the topology is changed, the estimation of the evolutionary times of the gain and loss of genes are changed accordingly. As a result, it is possible to miscount the total numbers of gains and losses of the genes. However, our results showed that the sets of genes were different among the 122 species examined (appendix A of Supplemental Material online). Because the gain and loss of genes cause the differentiation of the sets of genes in the species, our conclusion that the evolutionary process of VB6 metabolism has been quite dynamic regarding the events of gain and loss of genes, under some constraints, is not altered, even when the topology of the phylogenetic tree changes.

Studies using comparative analysis have often shown differences in gene sets involved in metabolic pathways among species (Huynen, Dandekar, and Bork 1999). By estimating the gain and loss of genes, we are able to learn not only the differences in a metabolic pathway among the species examined but also in which lineage the change in the metabolic pathway occurred during evolution. This means that we will be able to understand the evolutionary processes of the metabolic networks by evaluating the gains and losses of genes. In some metabolic pathways, dysfunctions in particular lineages have been reported (Smirnov 2001; Meganathan 2001). By applying our approach to these metabolic pathways, we will be able to elucidate the dysfunctions in these pathways by the gain and loss of genes. It is also possible to further extend our approach to other metabolic networks in the KEGG (Kanehisa et al. 2002) and EcoCyc (Karp et al. 2002) databases, to more clearly understand the evolutionary processes of the metabolic pathways they contain.

Acknowledgments

We thank the associate editor and two anonymous reviewers for their constructive suggestions.

Literature Cited

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of Eukaryotes based on combined protein data. *Science* **290**:972–977.
- Bean, L. E., W. H. Dvorachek Jr, E. L. Braun, A. Errett, G. S. Saenz, M. D. Giles, M. Werner-Washburne, M. A. Nelson, and D. O. Natvig. 2001. Analysis of the *pdx-1* (*snz-1/sno-1*) region of the *Neurospora crassa* genome: correlation of pyridoxine-requiring phenotypes with mutations in two structural genes. *Genetics* **157**:1067–1075.
- Bilski, P., M. Y. Li, M. Ehrenshaft, M. E. Daub, and C. F. Chignell. 2000. VB6 (pyridoxine) and its derivatives are efficient singlet oxygen quenchers and potential fungal antioxidants. *Photochem. Photobiol.* **71**:129–134.
- Bono, H., S. Goto, W. Fujibuchi, H. Ogata, and M. Kanehisa. 1998. Systematic prediction of orthologous units of genes in the complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **9**:32–40.
- Ehrenshaft, M., P. Bilski, M. Y. Li, C. F. Chignell, and M. E. Daub. 1999. A highly conserved sequence is a novel gene involved in de novo VB6 biosynthesis. *Proc. Natl. Acad. Sci. USA* **96**:9374–938.
- Ehrenshaft, M., and M. E. Daub. 2001. Isolation of PDX2, a second novel gene in the pyridoxine biosynthesis pathway of eukaryotes, archaeobacteria, and a subset of eubacteria. *J. Bacteriol.* **183**:3383–3390.
- Grogan, D. W. 1988. Temperature-sensitive murein synthesis in an *Escherichia coli* *pdx* mutant and the role of alanine racemase. *Arch. Microbiol.* **150**:363–367.
- Helmreich, E. J. 1992. How pyridoxal 5'-phosphate could function in glycogen phosphorylase catalysis. *Biofactors* **3**:159–172.
- Horowitz, N. H. 1945. On the evolution of biochemical synthesis. *Proc. Natl. Acad. Sci. USA* **31**:152–157.
- Huynen, M. A., T. Dandekar, and P. Bork. 1999. Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.* **7**:281–291.
- Jensen, R. A. 1976. Enzyme recruitment on evolution of new function. *Annu. Rev. Microbiol.* **30**:409–425.
- Kack, H., J. Sandmark, K. Gibson, G. Schneider, and Y. Lindqvist. 1999. Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes. *J. Mol. Biol.* **291**:857–876.
- Kanehisa, M., S. Goto, S. Kawashima, and A. Nakaya. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res.* **30**:42–46.
- Karp, P. D., M. Riley, M. Saier, I. T. Paulsen, J. Collado-Vides, S. M. Paley, A. Pellegrini-Toole, C. Bonavides, and S. Gama-Castro. 2002. The EcoCyc database. *Nucleic Acids Res.* **30**:56–58.
- Lam, H. M., and M. E. Winkler. 1990. Metabolic relationships between pyridoxine (VB6) and serine biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.* **172**:6518–6528.
- Man, T. K., G. Zhao, and M. E. Winkler. 1996. Isolation of a *pdxJ* point mutation that bypasses the requirement for the *PdxH* oxidase in pyridoxal 5'-phosphate coenzyme biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.* **178**:2445–2449.
- Manore, M. M. 2000. Effect of physical activity on thiamine, riboflavin, and VB6 requirements. *Am. J. Clin. Nutr.* **72**:598S–606S.
- Meganathan, R. 2001. Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms. *Vitam. Horm.* **61**:173–218.
- Mihara, H., T. Kurihara, T. Yoshimura, K. Soda, and N. Esaki. 1997. Cysteine sulfinatase desulfinase, a NIFS-like protein of *Escherichia coli* with selenocysteine lyase and cysteine desulfurase activities: gene cloning, purification, and characterization of a novel pyridoxal enzyme. *J. Biol. Chem.* **272**:22417–22424.
- Mittenhuber, G. 2001. Phylogenetic analyses and comparative genomics of VB6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *J. Mol. Microbiol. Biotechnol.* **3**:1–20.
- Nakamura, Y., T. Itoh, H. Matsuda, and T. Gojobori. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* **36**:760–766.
- Nelson, K. E., I. T. Paulsen, J. F. Heidelberg, and C. M. Fraser. 2000. Status of genome projects for nonpathogenic bacteria and archaea. *Nat. Biotechnol.* **18**:1049–1054.

- Osmani, A. H., G. S. May, and S. A. Osmani. 1999. The extremely conserved *pyroA* gene of *Aspergillus nidulans* is required for pyridoxine synthesis and is required indirectly for resistance to photosensitizers. *J. Biol. Chem.* **274**:23565–23569.
- Rottmann, W. H., G. F. Peter, P. W. Oeller, J. A. Keller, N. F. Shen, B. P. Nagy, L. P. Taylor, A. D. Campbell, and A. Theologis. 1991. 1-aminocyclopropane-1-carboxylate synthase in tomato is encoded by a multigene family whose transcription is induced during fruit and floral senescence. *J. Mol. Biol.* **222**:937–961.
- Schmidt, S., S. Sunyaev, P. Bork, and T. Dandekar. 2003. Metabolites: a helping hand for pathway evolution? *Trends Biochem. Sci.* **28**:336–341.
- Seack, J., S. Perovic, V. Gamulin, H. C. Schroder, P. Beutelmann, I. M. Muller, and W. E. Muller. 2001. Identification of highly conserved genes: SNZ and SNO in the marine sponge *Suberites domuncula*: their gene structure and promoter activity in mammalian cells(1). *Biochim. Biophys. Acta.* **1520**:21–34.
- Seta, F. D., S. Boschi-Muller, M. L. Vignais, and G. Branlant. 1997. Characterization of *Escherichia coli* strains with *gapA* and *gapB* genes deleted. *J. Bacteriol.* **179**:5218–5221.
- Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**:81–86.
- Smirnoff, N. 2001. L-ascorbic acid biosynthesis. *Vitam. Horm.* **61**:241–266.
- Tatusov, R. L., A. R. Mushegian, P. Bork, N. P. Brown, W. S. Hayes, M. Borodovsky, K. E. Rudd, and E. V. Koonin. 1996. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.* **6**:279–291.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **11**:4673–4680.
- White, R. S., and W. B. Dempsey. 1970. Purification and properties of VB6 kinase from *Escherichia coli* B. *Biochemistry* **9**:4057–4064.
- Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**:4576–4579.
- Yang, Y., G. Zhao, and M. E. Winkler. 1996. Identification of the *pdxK* gene that encodes pyridoxine (VB6) kinase in *Escherichia coli* K-12. *FEMS Microbiol. Lett.* **141**:89–95.
- Yang, Y., G. Zhao, T. K. Man, and M. E. Winkler. 1998a. Involvement of the *gapA*- and *epd* (*gapB*)-encoded dehydrogenases in pyridoxal 5'-phosphate coenzyme biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.* **180**:4294–4299.
- Yang, Y., H. C. Tsui, T. K. Man, and M. E. Winkler. 1998b. Identification and function of the *pdxY* gene, which encodes a novel pyridoxal kinase involved in the salvage pathway of pyridoxal 5'-phosphate biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.* **180**:1814–1821.
- Zhao, G., A. J. Pease, N. Bharani, and M. E. Winkler. 1995. Biochemical characterization of *gapB*-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and its possible role in pyridoxal 5'-phosphate biosynthesis. *J. Bacteriol.* **177**:2804–2812.

Manolo Gouy, Associate Editor

Accepted October 4, 2004