

Molecular Phylogeny and Intricate Evolutionary History of the Three Isofunctional Enzymes Involved in the Oxidation of Protoporphyrinogen IX

Koichi Kobayashi^{1,*}, Tatsuru Masuda², Naoyuki Tajima¹, Hajime Wada^{1,3}, and Naoki Sato^{1,3,*}

¹Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Japan

²Department of General Systems Studies, Graduate School of Arts and Sciences, The University of Tokyo, Japan

³CREST, JST, Saitama, Japan

*Corresponding author: E-mail: kkobayashi@bio.c.u-tokyo.ac.jp; naokisat@bio.c.u-tokyo.ac.jp.

Accepted: August 1, 2014

Abstract

Tetrapyrroles such as heme and chlorophyll are essential for biological processes, including oxygenation, respiration, and photosynthesis. In the tetrapyrrole biosynthesis pathway, protoporphyrinogen IX oxidase (Protox) catalyzes the formation of protoporphyrin IX, the last common intermediate for the biosynthesis of heme and chlorophyll. Three nonhomologous isofunctional enzymes, HemG, HemJ, and HemY, for Protox have been identified. To reveal the distribution and evolution of the three Protox enzymes, we identified homologs of each along with other heme biosynthetic enzymes by whole-genome clustering across three domains of life. Most organisms possess only one of the three Protox types, with some exceptions. Detailed phylogenetic analysis revealed that HemG is mostly limited to γ -Proteobacteria whereas HemJ may have originated within α -Proteobacteria and transferred to other Proteobacteria and Cyanobacteria. In contrast, HemY is ubiquitous in prokaryotes and is the only Protox in eukaryotes, so this type may be the ancestral Protox. Land plants have a unique HemY homolog that is also shared by *Chloroflexus* species, in addition to the main HemY homolog originating from Cyanobacteria. Meanwhile, organisms missing any Protox can be classified into two groups; those lacking most heme synthetic genes, which necessarily depend on external heme supply, and those lacking only genes involved in the conversion of uroporphyrinogen III into heme, which would use a precorrin2-dependent alternative pathway. However, *hemN* encoding coproporphyrinogen IX oxidase was frequently found in organisms lacking Protox enzyme, which suggests a unique role of this gene other than in heme biosynthesis.

Key words: heme, HemG, HemJ, HemY, protoporphyrin IX, tetrapyrrole.

Introduction

Tetrapyrroles such as heme and chlorophyll (Chl) are the most abundant and probably the most ancient pigments found in living organisms. As cofactors, they have a wide range of chemical properties, including light absorption, electron transfer, and oxygen binding, and so are essential components of critical biological processes such as respiration and photosynthesis across all kingdoms of life.

Heme and Chl share the same biosynthetic pathway from the first universal tetrapyrrole precursor, 5-aminolevulinic acid (ALA), to the last common intermediate, protoporphyrin IX (Proto) (Tanaka et al. 2011). The pathway of ALA synthesis differs in various organisms (Panek and Brian 2002;

Heinemann et al. 2008; Tanaka et al. 2011), whereas the common pathway from ALA to Proto is highly conserved in virtually all living organisms except some archaea (Storbeck et al. 2010). Some other exceptions include parasitic organisms, which depend on heme produced by host organisms (Panek and Brian 2002; Heinemann et al. 2008).

The last common step toward heme and Chl synthesis is oxidation of protoporphyrinogen IX (Protox) to Proto by Protox (EC 1.3.3.4). In photosynthetic organisms, the pathway is then branched into heme and Chl synthesis by the insertion of a ferrous ion into Proto by ferrochelatase and the insertion of a magnesium ion by Mg-chelatase, respectively (Tanaka et al. 2011).

Genes involved in Protox activity were identified first from *Escherichia coli* (Sasarman et al. 1993) and *Bacillus subtilis* (Hansson and Hederstedt 1992) and designated *hemG* and *hemY*, respectively. The HemY-type Protox (≈ 55 kDa) is an oxygen-dependent oxidase with a flavin-adenine dinucleotide (FAD) cofactor (Koch et al. 2004; Corradi et al. 2006; Qin et al. 2010). Most eukaryotes and many aerobic or facultative bacteria possess the *hemY* gene for Protogen oxidation (Oborník and Green 2005). Meanwhile, γ -Proteobacteria such as *E. coli* (Sasarman et al. 1993) and *Salmonella typhimurium* (Xu et al. 1992) use HemG for Protogen oxidation. The HemG protein of *E. coli* is a 21-kDa protein with a flavin mononucleotide (FMN) as a cofactor and possesses the Protox activity in aerobic or anaerobic conditions (Boynton et al. 2009; Möbius et al. 2010). Recently, another gene responsible for Protogen oxidation, *hemJ*, was identified in a cyanobacterium *Synechocystis* sp. PCC 6803 as being distantly related to a subunit of NADH dehydrogenase complex I (Kato et al. 2010). This gene encodes a 22-kDa enzyme featuring Protox activity under aerobic conditions. *HemJ* homologs are found in most Cyanobacteria (Kato et al. 2010) and many other bacteria such as *Acinetobacter baylyi* (Boynton et al. 2011). Therefore, three phylogenetically unrelated enzymes catalyze the same oxidation step of Protogen in different organisms. These are examples of nonhomologous isofunctional enzymes (Galperin et al. 1998; Omelchenko et al. 2010) or analogous enzymes, which hereafter we call isofunctional enzymes.

Although Protogen can be converted to Proto without the Protox activity by a spontaneous reaction with molecular oxygen or by nonspecific peroxidases (Jacobs and Jacobs 1993; Lee et al. 1993), the requirement of Protox in tetrapyrrole biosynthesis *in vivo* has been demonstrated in a number of mutants for *hemG* (Xu et al. 1992; Sasarman et al. 1993), *hemY* (Hansson and Hederstedt 1992; Camadro and Labbe 1996; Meissner et al. 1996; Warnich 1996; Molina et al. 1999), and *hemJ* (Kato et al. 2010; Boynton et al. 2011). Moreover, Protox is a target of diphenyl ether herbicides, which inhibit the activity of HemY-type Protox and result in the accumulation of Proto that causes photobleaching in plants (Matringe et al. 1989; Witkowski and Halling 1989). Because Proto is a strong photosensitizer, which causes rapid oxidation of various molecules within the cell, the accumulation of this intermediate should be critically kept at a low level. HemY-type Protox may form a complex with ferrochelatase to enable efficient channeling of Proto between these two enzymes (Ferreiras et al. 1988; Koch et al. 2004; Masoumi et al. 2008). Moreover, in photosynthetic organisms synthesizing Chls or bacteriochlorophylls, Protox also transfers Proto to Mg-chelatase (Tanaka et al. 2011), which suggests a more complex role of this oxidase in regulating the substrate channeling between the heme and Chl branches. In this respect, it is curious to find three different enzymes involved in the Protogen oxidation across eukaryotes and bacteria, considering the need for strict and complex regulation of this

potentially harmful step in coordination with up- and downstream pathways. In addition, some organisms such as plants possess two isoforms of HemY (Oborník and Green 2005). The phylogenetic relationship as well as the differentiation of cellular functions of these two enzymes remains to be answered.

In this study, we identified homologs of each type of Protox in a wide range of organisms, including photosynthetic and nonphotosynthetic eukaryotes and prokaryotes, by using the recently developed automatic clustering method (Sato 2009). Then, we performed detailed phylogenetic analysis with sequence-based maximum-likelihood (ML) methods and detection of insertions/deletions within the amino acid sequences. As background information, we searched for all possible homologs in all bacterial genomes reported at the time of our analysis and extensively surveyed all homologs involved in heme biosynthesis together with Protox. We analyzed the coevolution of Protox enzymes with other heme biosynthesis enzymes and assessed replacement of different types of Protox in different lineages of bacteria.

Materials and Methods

Sequences

All homologs of HemY, HemG, and HemJ were retrieved from the Gclust database (Gclust2010e29b data set at <http://gclust.c.u-tokyo.ac.jp/> [last accessed August 12, 2014, now available under "Old versions"] including selected genomes covering plants, algae, nonphotosynthetic eukaryotes, and prokaryotes.) according to the published list of homologs. Gclust is a comparative genomic database of homologous protein clusters suitable for phylogenetic profiling (Sato 2002, 2009; Sato et al. 2005). The sources of the original databases are described on the website.

We identified all homologs of enzymes involved in heme biosynthesis in all prokaryotic genomes in RefSeq of the National Center for Biotechnology Information (NCBI) database as of November 2010 (data set AllBact2010) and used them for the statistics shown in table 2. In most analyses, the two data sets were sufficient for obtaining an idea of the overall distribution of enzymes in various phyla. However, some new sequences were retrieved from RefSeq as of February 2013 to complement the data in detailed analyses for [supplementary figures S4 and S5, Supplementary Material online](#). All 3,916,828 proteins in the 1,196 prokaryotic genomes were used in all-against-all BLASTP (mostly v2.2.22) analysis (options: $-m8 -FF -CO$). The results in a single table (specified by option $-m8$) were used for clustering by use of gclust v3.56 (parallel version) (Sato 2009). Each of the enzymes in the heme biosynthetic pathway was first identified by name, then all homologs were retrieved as several clusters. Additionally, some singletons were retrieved by following the "related groups" link. A list of enzymes in all analyzed

organisms was compiled from the lists of homologs. The huge file of original data of clusters (7.12 GB) is not currently available on the web site but may be obtained from the corresponding author upon request.

Phylogenetic Analyses

Amino acid alignment was performed for each isofunctional enzymes by using Muscle v3.6 (Edgar 2004). Sequences not aligned in the entire length were removed, and alignment was repeated. The sequence alignment was used for subsequent phylogenetic analysis and homology modeling. The sites with gaps in more than 20% of sequences were removed by use of SISEQ v1.59 (Sato 2000), which was also used for conversion of various sequence formats. ClustalX v1.83 was used for profile alignment and to manage aligned sequences (Thompson et al. 1994).

Each alignment was used in the phylogenetic analysis as follows. A neighbor-joining (NJ) tree was estimated with MEGA v4 (Tamura et al. 2007) with the Jones–Taylor–Thorton (JTT) model and an equal evolutionary rate. Calculation with the ML method involved use of TreeFinder March 2008 version (Jobb et al. 2004) with the Whelan–Goldman (WAG) model, and with RAxML v7.0.4 (Stamatakis 2006) with $-f d -i 10 -m PROTCATWAG$ options. The exact parameters were determined by initial trials with $-c 10, 40, 55$ with or without $-i 10$. Bootstrap was based on 1,000 replicates. Bayesian inference (BI) involved use of MrBayes v3.2 (Ronquist and Huelsenbeck 2003), with the following options: $aamodelpr=fixed(wag)$, $ratepr=variable$, $ngen=2,000,000$, $samplefreq=200$, $burnin=3,000$ (for HemG, $ngen=1,000,000$, $samplefreq=100$), and with PhyloBayes v3.2e (Lartillot et al. 2009) with the CAT+gtr model. A 16S–23S-based phylogenetic tree in Cyanobacteria (fig. 2) was constructed with the BI method as described (Sasaki and Sato 2010).

The approximately unbiased (AU) test, intended to test the relative likelihood of various forms of trees based on support levels of individual sites, involved use of the CONSEL program (Shimodaira and Hasegawa 2001), with the output of Protml in MOLPHY v2.3beta (Adachi and Hasegawa 1996), based on the 19 most probable trees for 20 representative taxa. For this purpose, the trees were selected by Protml with constrained trees according to the results of ML and BI analyses. The WAG and JTT models were used. Phylogenetic trees were drawn with use of NJplot (Perrière and Gouy 1996) and Mesquite utility v2.5 (<http://mesquiteproject.org>, last accessed August 12, 2014). The alignment files used for phylogenetic analysis are available in [supplementary material, Supplementary Material](#) online.

Homology Modeling

Homology modeling of HemY homologues involved use of Modeller v8v1 or 8v2 (Sali and Blundell 1993) with the

automodel script, modified as necessary. Both of the structures in the Protein Data Bank (PDB) entry 1SEZ for tobacco Protox (PPO2) were used as templates. This file describes two monomers arranged in a symmetric position, but some loops and the N-terminus are invisible because of high flexibility. The corresponding parts in the inferred models are an invention of the software and are only meaningful as an indication of the presence of a structural part. Cartoon models of structure were prepared with use of Molscript (Kraulis 1991). The coordinate files of homology modeling are available from the corresponding author upon request.

Results and Discussion

Distribution of HemY, HemG, and HemJ across Eukaryotes and Prokaryotes

To examine the distribution pattern of HemY, HemG, and HemJ across eukaryotes and prokaryotes, homologous proteins clustered with each known Protox were identified in the Gclust2010 database including 169 organisms (Sato 2002, 2009; Sato et al. 2005). Because conventions for gene names differ in different organisms, all gene names are described according to the prokaryotic convention in this study. Table 1 shows that all 33 eukaryotes analyzed (plants, algae, protists, fungi, and animals) contain at least a copy of *hemY*, but no copies of *hemG* and *hemJ*. In addition, HemY was found in various bacterial phyla, which is consistent with a previous report (Obornik and Green 2005). HemG was found in γ -Proteobacteria, Cyanobacteria, and green bacteria. HemJ was found in most species of α - and β -Proteobacteria and Cyanobacteria. Some Chlorobi such as *Chlorobium chlorochromatii*, *Chlorobium tepidum*, *Chlorobium (Pelodictyon) luteolum*, and *Prosthecochloris vibrioformis* were reported to lack any types of Protox (Kato et al. 2010), but we found HemY homologs in all Chlorobi species investigated, which suggests that Chlorobi use HemY for Protogen oxidation. However, no Protox types were found in *Buchnera aphidicola*, *Mycoplasma genitalium*, *Archaeoglobus fulgidus*, *Aeropyrum pernix*, *Nanoarchaeum equitans*, or *Dehalococcoides CBDB1*. As discussed later, these prokaryotes lack de novo heme biosynthesis or use alternative pathways for heme production.

To exhaustively detect all isofunctional enzymes of Protox in all prokaryotes, we performed complete clustering of all proteins in all known prokaryotic genomes ([supplementary table S1, Supplementary Material](#) online). As reported previously (Panek and Brian 2002), ALA synthase (ALAS), which catalyzes the single-step condensation of succinyl-CoA and glycine to form ALA, was almost exclusively found in α -Proteobacteria (table 2), which shows the reliability of our clustering analysis. Of 1,196 prokaryotes, 834 species (70% of total prokaryotes) possess at least one Protox. Among them, 732 species (87%) possess only a single Protox. Because Proto is a strong photosensitizer and therefore should be controlled carefully during

Table 1

Distribution of Protox Isofunctional Enzymes in Eukaryotes and Prokaryotes in Gclust2010 Database

ID	Species	HemG	HemJ	HemY	Total
Eukaryotes					
ATH	<i>Arabidopsis thaliana</i>	0	0	3	3
CME	<i>Cyanidioschyzon merolae</i>	0	0	1	1
OSA	<i>Oryza sativa</i>	0	0	3	3
PPT	<i>Physcomitrella patens</i>	0	0	2	2
PoTR	<i>Populus trichocarpa</i>	0	0	3	3
SMOL	<i>Selaginella moellendorffii</i>	0	0	4	4
VVI	<i>Vitis vinifera</i>	0	0	2	2
SoBI	<i>Sorghum bicolor</i> BTx623	0	0	2	2
BrDI	<i>Brachypodium distachyon</i>	0	0	2	2
LoJA	<i>Lotus japonicus</i> MG-20	0	0	2	2
MeTR	<i>Medicago truncatula</i>	0	0	2	2
MZE	<i>Zea mays</i> B73	0	0	4	4
CRE	<i>Chlamydomonas reinhardtii</i>	0	0	3	3
TPS	<i>Thalassiosira pseudonana</i>	0	0	1	1
PTR	<i>Phaeodactylum tricorutum</i>	0	0	1	1
OTAU	<i>Ostreococcus tauri</i> OTTH095	0	0	1	1
OLUC	<i>Ostreococcus lucimarinus</i> CCE9901	0	0	1	1
CEL	<i>Caenorhabditis elegans</i>	0	0	1	1
DME	<i>Drosophila melanogaster</i>	0	0	1	1
HSA	<i>Homo sapiens</i>	0	0	1	1
HYMA	<i>Hydra magnipapillata</i>	0	0	1	1
PLAT	<i>Ornithorhynchus anatinus</i>	0	0	1	1
SCE	<i>Saccharomyces cerevisiae</i>	0	0	1	1
SPO	<i>Schyzosaccharomyces pombe</i>	0	0	1	1
PFA	<i>Plasmodium falciparum</i>	0	0	1	1
NGR	<i>Naegleria gruberi</i>	0	0	1	1
DCGR	<i>Candida glabrata</i> CBS138	0	0	1	1
DKLA	<i>Kluyveromyces lactis</i> NRRL Y-1140	0	0	1	1
NCR	<i>Neurospora crassa</i> 74-OR23-1A	0	0	2	2
TET	<i>Tetrahymena thermophila</i> SB210	0	0	1	1
PHRA	<i>Phytophthora ramorum</i>	0	0	1	1
PHSO	<i>Phytophthora sojae</i>	0	0	1	1
DPTM	<i>Paramecium tetraurelia</i>	0	0	2	2
Bacterial phyla					
	α -Proteobacteria	1	32	3	34
	γ -Proteobacteria	3	2	2	7
	β -Proteobacteria	0	6	1	6
	Firmicutes	0	2	0	3
	Actinobacteria	1	0	4	4
	Small genome bacteria	1	0	5	6
	Archaea	1	0	3	7
	Cyanobacteria	2	34	9	42
	Green bacteria	5	2	21	26

substrate channeling, the presence of a single copy of Protox gene might be preferred. Of note, most organisms with a Protox enzyme also have a ferrochelatase HemH. Tight coupling of Protox and HemH might be advantageous for keeping free Proto at a low level to prevent photoactive and harmful

action of this compound to the cell. Although Protox and ferrochelatase are presumed to have evolved in parallel, the mode of interaction of the three types of Protox isofunctional enzymes with ferrochelatase remains unknown.

Table 2 displays the distribution of the three isofunctional enzymes of Protox in prokaryotes. The distribution generally depended on phylum. HemY is found in most phyla, whereas HemG is mostly restricted to γ -Proteobacteria and a part of Firmicutes. HemJ is found in various Proteobacteria, *Bacteroides*, and Cyanobacteria. However, some species in various phyla contain an isofunctional Protox different from the major one in the phylum, which could indicate frequent horizontal gene transfer events. A number of species do not contain any isofunctional Protox (see the difference between the overall number of species and the number of species containing Protox). In particular, Protox is not found in Tenericutes and Crenarchaea. Some parasitic or symbiotic organisms such as *Haemophilus influenzae* (Panek and Brian 2002), *Caenorhabditis elegans*, and *Leishmania major* (Rao et al. 2005) lack not only Protox but also most of the heme biosynthetic enzymes, and their life depends on external heme supply. In fact, our analysis of all homologs of enzymes involved in heme biosynthesis shows that parasitic or symbiotic bacteria such as *Mycoplasma*, *Borrelia*, *Streptococcus*, *Lactobacillus*, *Bifidobacterium*, and *Buchnera* lack most enzymes for heme biosynthesis including Protox (supplementary table S2, Supplementary Material online), so these bacteria may require heme supply from the hosts/symbiotic partners. Furthermore, *Bacteroides thetaiotaomicron*, *Bacteroides vulgatus*, and *Candidatus Sulcia*, for which the possible presence of yet-unidentified genes for Protox was discussed (Kato et al. 2010), also lack the known heme biosynthetic pathway, so these organisms may not require Protox activity. However, some bacteria that do not seem parasitic or symbiotic (e.g., *Dehalococcoides*, *Thermotoga*, *Dictyoglomus*) also lack most of the heme biosynthetic enzymes. Thus far, heme sources in these organisms are an enigma.

Some bacteria (e.g., *Clostridium* and *Desulfovibrio* species) and most archaea lack the enzymes involved in the conversion of uroporphyrinogen III (Urogen) into heme but possessed the enzymes required for Urogen formation (supplementary table S3, Supplementary Material online) (Panek and Brian 2002; Cavallaro et al. 2008). Indeed, *Desulfovibrio vulgaris* and *Methanosarcina barkeri* may synthesize protoheme through precorrin 2, which is formed by the methylation of Urogen and thus bypasses the pathway from Coprogen to Proto. We confirmed that *Desulfovibrio* and *Methanosarcina* species indeed lack any types of Protox in their genome but possessed enzymes required for Urogen formation (supplementary table S3, Supplementary Material online). As proposed by Cavallaro et al. (2008), many archaeal species lack most enzymes required for the conversion of Urogen into Proto. Specifically, almost all enzymes required for Proto formation from Urogen are missing in Crenarchaeota, so these species may exclusively

Table 2

Distribution of Heme Biosynthetic Proteins in 1,176 Known Prokaryotes

Phylum	All	Protox				ALA Synthesis			HemB-E	Coprox			HemH
		Y	J	G	Total	HemAL	ALAS	Total		F	N/Z	Total	
γ -Proteobacteria	264	46	82	153	247	246	1	246	255	234	259	260	250
β -Proteobacteria	83	29	82	2	82	82	2	82	82	75	82	82	82
ϵ -Proteobacteria	32	1	30	0	31	32	0	32	32	0	32	32	31
δ -Proteobacteria	35	18	1	3	19	35	0	35	35	6	35	35	19
α -Proteobacteria	128	10	122	3	122	1	121	121	122	119	109	127	122
Other Proteobacteria	1	1	0	0	1	1	0	1	1	0	1	1	1
Firmicutes	231	82	0	20	99	128	1	128	145	0	208	208	120
Tenericutes	29	0	0	0	0	0	0	0	0	0	6	0	0
Actinobacteria	109	88	0	4	88	95	2	95	98	3	104	104	94
Chlamydiae	16	16	0	0	16	16	0	16	16	2	16	16	16
Spirochaetes	22	6	0	0	6	7	0	7	8	0	14	14	9
Acidobacteria	3	3	0	0	3	3	0	3	3	0	3	3	3
Bacteroidetes	34	12	12	1	23	17	0	17	29	15	28	28	19
Other bacteria	17	5	1	1	6	13	0	13	13	3	17	17	6
Cyanobacteria	40	15	34	2	40	40	0	40	40	40	40	40	40
Green sulfur bacteria	11	11	0	1	11	11	0	11	11	0	11	11	11
Green nonsulfur bacteria	14	8	0	4	8	9	0	9	9	3	14	14	5
<i>Deinococcus-Thermus</i>	8	7	0	2	7	8	0	8	8	1	8	8	7
Hyperthermophilic bacteria	25	8	3	3	14	14	0	14	16	0	24	24	12
Additional bacteria	5	2	1	1	3	3	1	4	4	1	5	5	2
Archaea													
Euryarchaeota	58	11	0	13	21	51	0	51	51	0	0	0	6
Crenarchaeota	28	0	0	0	0	23	0	23	22	0	3	3	0
Other archaea	3	0	0	0	0	2	0	2	1	0	0	0	0

NOTE.—Data are number of species.

use the alternative heme biosynthetic pathway through precorrin-2.

Several bacteria from various phyla such as *Mycobacterium avium* and *Thermus thermophilus* lacked known Protox types but possessed all other enzymes for heme biosynthesis (supplementary table S4, Supplementary Material online). Some of these bacteria (*Geobacter sulfurreducens*, *Arthrobacter arilaitensis*, *Opitutus terrae*, *Thermomicrobium roseum*, and *T. thermophilus*) seem to be free-living and thus might have yet unidentified genes for Protox activity.

A number of bacteria lacking most of the heme biosynthetic pathway (e.g., *Buchnera*, *Bartonella*, *Streptococcus*, *Bifidobacterium*, *Dehalococcoides*, *Thermotoga*) possess homologs for *hemN*, which encodes an oxygen-independent type of coproporphyrinogen III oxidase (supplementary table S2, Supplementary Material online). In addition, *Desulfovibrio*, *Clostridium*, and some Archaea (*Ignicoccus hospitalis*, *Hyperthermus butylicus*, and *Acidilobus saccharovorans*), which are supposed to use the precorrin-2 pathway for heme biosynthesis, also have *hemN* homologs (supplementary table S3, Supplementary Material online). In *Lactococcus lactis*, which lacks the heme biosynthetic pathway, the HemN homolog named HemW is not a functional Coprox but a

heme-binding protein with a putative role in heme trafficking (Abicht et al. 2012). Thus, the prokaryotes lacking the common heme biosynthetic pathway may use the *hemN* homologs for heme trafficking or other processes than heme biosynthesis as proposed for *L. lactis* (Abicht et al. 2012).

Phylogenetic Analysis of HemG

hemG is mostly limited to γ -Proteobacteria, but is also found in various bacteria (table 2). Phylogenetic analysis revealed that HemG sequences from γ -Proteobacteria are clustered in a major clade that is highly supported by various tests (fig. 1). The diversification of *hemG* within γ -Proteobacteria is more obvious in the phylogenetic analysis of all *hemG* genes (supplementary fig. S1, Supplementary Material online), so HemG may be conserved predominantly within this class. The γ -proteobacterial *hemG* clade contains genes from two strains of *Prochlorococcus marinus* (Cyanobacteria), which suggests a horizontal transfer. Because these *Pr. marinus* strains have no HemJ or HemY, the major Protox types in Cyanobacteria, these strains might have lost these Protox after acquisition of HemG. However, a detailed comparison of the genomic sequences near the *hemG* or *hemJ* genes in closely related *Prochlorococcus* strains (fig. 2) revealed that the

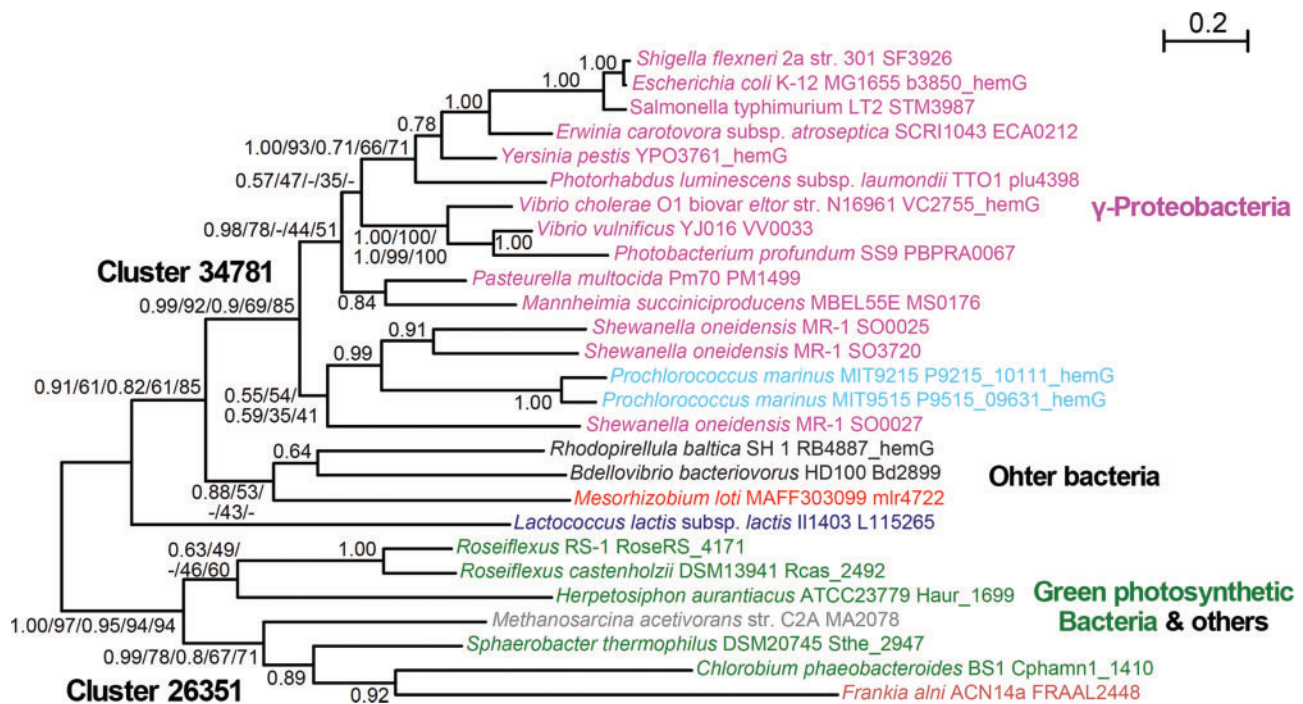


FIG. 1.—Phylogenetic tree of HemG. BI tree of HemG proteins with confidence support values obtained with BI/ML/PB/RA/NJ methods shown at each branch. Only BI score is shown in the interior of branches for visibility. Distance scale is indicated by the bar on top. The cluster numbers refer to those in the Clustal 2010 data set. Cluster 26351 is an outgroup consisting of flavodoxins.

hemG gene fits into the identical place of *hemJ* in related strains. There are no significant changes in the surrounding genes, such as *uvrC* and *php*. We confirmed the connecting sequences by direct sequencing. The probable recombination indeed took place perfectly.

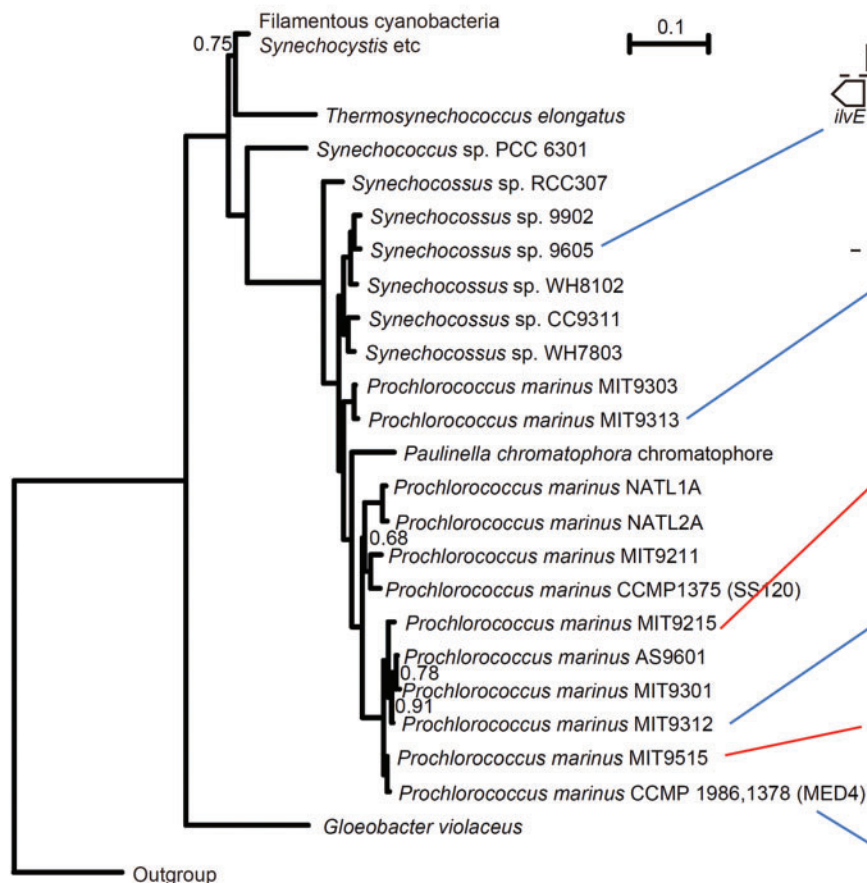
Homologous genes for HemG are also identified in some green photosynthetic bacteria and other prokaryotes. However, the sequences from green photosynthetic bacteria are phylogenetically distant from those of γ -Proteobacteria and form another cluster with those of *L. lactis* and Euryarchaea species, which probably lack the typical heme metabolic pathway (fig. 1 and [supplementary fig. S1, Supplementary Material](#) online). Considering that these green photosynthetic bacteria possess *hemY* genes, HemG homologs clustered in the outer clade likely do not function as Prottox in vivo. This possibility is supported by evidence that a long chain insert loop, a critical domain unique to HemG of γ -Proteobacteria (Boynton et al. 2009), is poorly conserved in HemG homologs of the outer clade group ([supplementary fig. S2, Supplementary Material](#) online). In fact, the HemG homolog from *Bdellovibrio bacteriovorus* (BD2899), also featuring a poorly conserved long chain insert loop, could not complement *hemG* deficiency in *E. coli* (Boynton et al. 2009). This bacterium also possesses a *hemY* gene presumably as the gene responsible for Protogen oxidation. HemG is a member of the long chain flavodoxins that are

FMN-containing electron transfer proteins involved in various biological processes. HemG homologs may have originated from long chain flavodoxins and evolved uniquely as the Prottox within γ -Proteobacteria.

Phylogenetic Analysis of HemJ

We next performed phylogenetic analysis of HemJ (fig. 3). HemJ sequences (Cluster 1819 or InterPro IPR005265) are found in many α -Proteobacteria and Cyanobacteria and in some other Proteobacteria. The tree was rooted with Cluster 28370 and Cluster 40856 as outgroups. These clusters (not included in InterPro IPR005265) differ from Cluster 1819 in the central part ([supplementary fig. S3, Supplementary Material](#) online) and are unlikely to be HemJ. Thus, HemJ may have originated within α -Proteobacteria. The sequences from other Proteobacteria and Cyanobacteria form sister groups to the α -proteobacterial HemJ, which shows that these sequences were transferred to these bacteria from α -Proteobacteria. Alternatively, HemJ emerged in a common ancestor of Proteobacteria and Cyanobacteria. However, the branching pattern in β - and γ -Proteobacteria might suggest multiple horizontal gene transfer events. If we admit that the acquisition of *hemJ* in Cyanobacteria is due to horizontal gene transfer, the transfer of *hemJ* gene into Cyanobacteria might have occurred near the base of cyanobacterial diversity before the separation of the *Anabaena* and

A. 16S-23S rRNA tree



B. Synteny

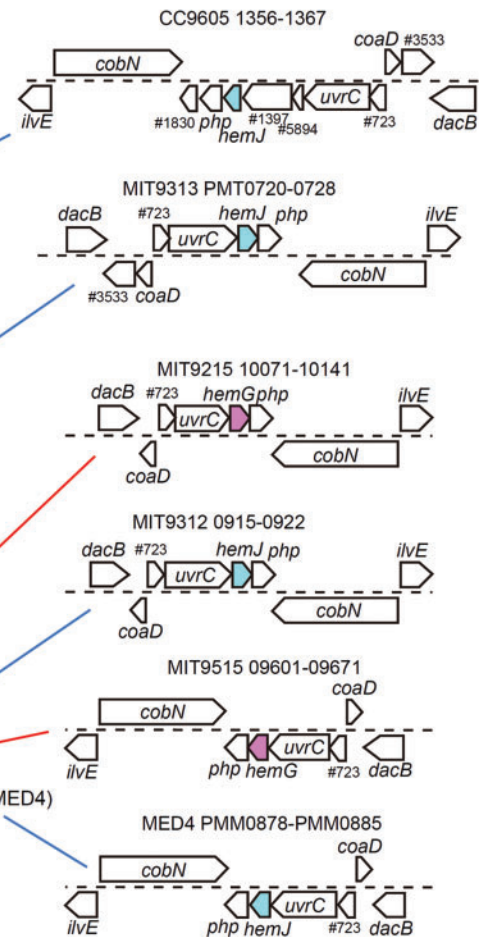


FIG. 2.—Synteny in the *hemG/hemJ* replacement region in *Prochlorococcus*. (A) A 16S–23S-based phylogenetic tree of the *Synechococcus*–*Prochlorococcus* lineage along with deeply branched *Gloeobacter violaceus*. (B) Arrangement of genes in the regions containing *hemG* or *hemJ* in related *Prochlorococcus* strains.

Synechococcus–*Prochlorococcus* lineages. Moreover, the *hemJ* from *Synechococcus* PCC 7502 (supplementary fig. S4, Supplementary Material online) is first branched from the root of the cyanobacterial clade. An extensive genome analysis showed that *Synechococcus* PCC 7502 constitutes a deep branch of the cyanobacterial phylogenetic tree with *Pseudanabaena* species, which is basal to the plastid branch (Shih et al. 2013). Thus, Cyanobacteria may have acquired *hemJ* before the endosymbiotic event in plant cells. *Gloeobacter violaceus*, the deepest branch within Cyanobacteria, possesses *hemJ* in addition to *hemY*. This species probably obtained *hemJ* through horizontal gene transfer because the *hemJ* sequence of *G. violaceus* is placed within the large branch including *Anabaena* and *Synechococcus*–*Prochlorococcus* lineages. However, an alternative explanation could be possible, whereby cyanobacterial HemJ is positioned at the root of HemJ diversification. If the enzymatic

mechanism including the active center of HemJ is elucidated, the rooting problem could be resolved.

A sequence similar to the *Prochlorococcus hemJ* is found in the chromatophore genome of the photosynthetic amoeba *Paulinella chromatophora*. This result is consistent with the hypothesis that chromatophores in *Paulinella* originated from a recent primary endosymbiosis of a member of the *Synechococcus*–*Prochlorococcus* group (Nowack et al. 2008). The amino acid sequence of *Pa. chromatophora* HemJ (PCHRC_PCC_0146) is highly conserved (supplementary fig. S3, Supplementary Material online), so this gene may not be a pseudogene. Because other genes for tetrapyrrole biosynthesis are also found in the chromatophore genome, except for *hemD* encoding Urogen synthase (Nowack et al. 2008), the cyanobacterial tetrapyrrole pathway including HemJ may be functional in the chromatophore. In plants, the mosaic origin of the heme biosynthetic pathway from

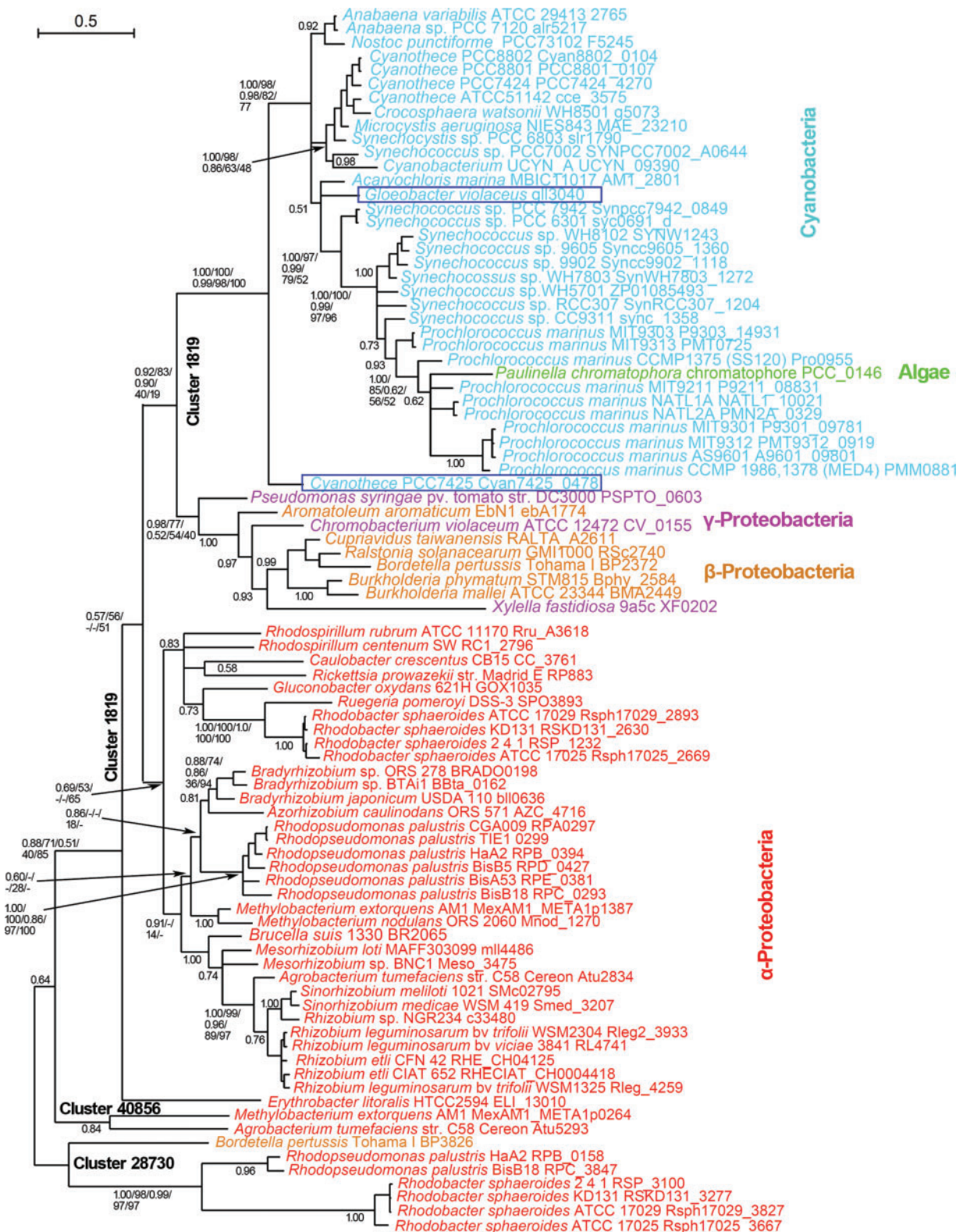


FIG. 3.—Phylogenetic tree of HemJ. BI tree of HemJ proteins with confidence support values obtained with BI/ML/PB/RA/NJ methods shown at each branch. Only a BI score is shown in the interior of branches for visibility. For multiple branches, only a BI score is assigned. Distance scale is indicated by the bar on top. The cluster numbers refer to those in the Gclust 2010 data set. Cluster 1819 contains HemJ proteins that were not yet characterized at the time of preparing the Gclust database. Clusters 40856 and 28730 are supposed to be outgroups having unknown function(s).

plastid (cyanobacterial), mitochondrial (α -proteobacterial), and nuclear genes is proposed (Oborník and Green 2005). Whether the *Pa. chromatophora* nuclear genome encodes a set of tetrapyrrole biosynthetic genes is unclear, and sequence analysis of the nuclear genome is needed to understand interactions of the pathway between the nucleus and the chromatophore in this unique organism.

Phylogenetic Analysis of HemY

HemY is widely distributed among almost all bacterial phyla and is the sole Protox in eukaryotes (tables 1 and 2). Its ubiquity indicates that HemY was used in the last common eukaryotic ancestor and was diversified according to species diversification. Because HemY is a member of an FAD protein superfamily showing sequence and structural homologies to amine oxidases and phytoene desaturases (Dailey and Dailey 1998; Koch et al. 2004), this enzyme probably originated from an FAD protein in the common ancestral lineage. Of note, green sulfur bacteria, which are obligate anaerobes, possess only the HemY-type Protox. Although HemY was characterized as the oxygen-dependent oxidase in several organisms (Heinemann et al. 2008), HemY from *Plasmodium falciparum* was found active only under anaerobic conditions in the presence of electron acceptors such as FAD, NAD, and NADP (Nagaraj et al. 2010). Therefore, HemY in green sulfur bacteria may not require molecular oxygen for Protox oxidation.

Although most of the Cyanobacteria we investigated feature the HemJ Protox, some Cyanobacteria possess HemY. The phylogenetic tree of HemY for Cyanobacteria is basically consistent with the tree for cyanobacterial species, with the deepest branch for *G. violaceus* HemY (fig. 4 and [supplementary fig. S5, Supplementary Material](#) online). Other deep-branching Cyanobacteria such as *Cyanobacterium* A-Prime (*Synechococcus* sp. JA-3-3Ab from Yellowstone) and *Pseudanabaena* PCC7367 also possess HemY for Protox, so HemY may be the ancestral Protox in Cyanobacteria but has been replaced by HemJ in most species.

Phylogenetic analysis showed that land plants have two distinct types of HemY, clustered into different clades (fig. 4). Both clusters for plant HemY are unambiguously separated from those for animals and fungi. As reported by Oborník and Green (2005), one of the plant HemY (PPO1) branches off from the cyanobacterial clade and thus could originate from a cyanobacterial HemY through endosymbiosis. In tobacco (Lermontova et al. 1997), PPO1 is targeted to plastids, which is consistent with its origin. Because deficiency of PPO1 causes severe growth defects along with necrotic leaf damage in both tobacco (Lermontova and Grimm 2006) and *Arabidopsis* (Molina et al. 1999), this plastid-localized enzyme could be the main PPO responsible for the tetrapyrrole biosynthesis in plastids.

Meanwhile, the other form of plant HemY (PPO2) is targeted to mitochondria in tobacco (Lermontova et al. 1997)

and dually targeted to plastids and mitochondria in spinach (Watanabe et al. 2001). PPO2-type HemY is unique to land plants and is not found in algae, which suggests that land plants acquired PPO2 genes after the divergence of an ancestral land plant from a green algal lineage. Oborník and Green (2005) reported that plant PPO2 was on the same branch as sequences from α -Proteobacteria and would be of proteobacterial origin. However, the branch formed by the α -proteobacterial sequences in their analysis had low bootstrap supports with substantial amino acid substitutions (Oborník and Green 2005). Because these α -proteobacterial sequences (ZP00004390 from *Rhodobacter sphaeroides*, NP767181 from *Bradyrhizobium japonicum*, and NP104729 from *Mesorhizobium loti*) are not clustered with HemY but rather with amine oxidases in the Gclust (ZP00010613 from *Rhodopseudomonas palustris* is DNA methylase and could be an annotation error), the relationship between plant PPO2 and α -proteobacterial sequences proposed in Oborník and Green (2005) could be an artifact due to long branch attraction. Indeed, these bacteria possess HemJ sequences as do most of the other α -Proteobacteria, and therefore HemJ probably functions as a Protox in these bacteria.

HemYs from *Chloroflexus aurantiacus* and *Chloroflexus aggregans* branch from the root of the plant PPO2 clade (fig. 4). The HemY sequences from other Chloroflexi such as *Roseiflexus* species belong to a large bacterial clade with those from Actinobacteria and Firmicutes. To ascertain this result, we performed the AU test for 19 phylogenetic trees constructed from 20 representative taxa (fig. 5 and table 3). All 15 trees in which the *Chloroflexus* hemY and plant PPO2 sequences formed sister branches (trees 1–15) gave AU probabilities >0.2 and therefore were not discarded. In contrast, trees in which the *Chloroflexus* HemY was placed on the bacterial branch (trees 16 and 17) or on the cyanobacterial branch (trees 18 and 19) gave AU values <0.05 , which are sufficiently low to reject these possibilities. These data confirm that *Chloroflexus* HemY and plant PPO2 sequences form a branch separate from the branches of plant PPO1 and bacterial HemY.

As indicated by Tang et al. (2011), except for *hemY* and adjacent *hemN*, which are approximately 2,000 genes away from the heme operon, genes involved in heme biosynthesis form an operon in *Chloroflexus* species ([supplementary fig. S6, Supplementary Material](#) online). Heme biosynthetic genes also form an operon in *Roseiflexus RS-1* and *Roseiflexus castenholzii*. Unlike in *Chloroflexus* species, the *hemY* gene is included in the operon in *Roseiflexus* species. Thus, *hemY* sequences found in *Roseiflexus* species may be the ancestral type of *hemY* in green nonsulfur bacteria, and in *Chloroflexus* species, it was replaced by another *hemY* gene related to plant PPO2 through horizontal gene transfer, although the ancestor of the PPO2 homologs remains unknown.

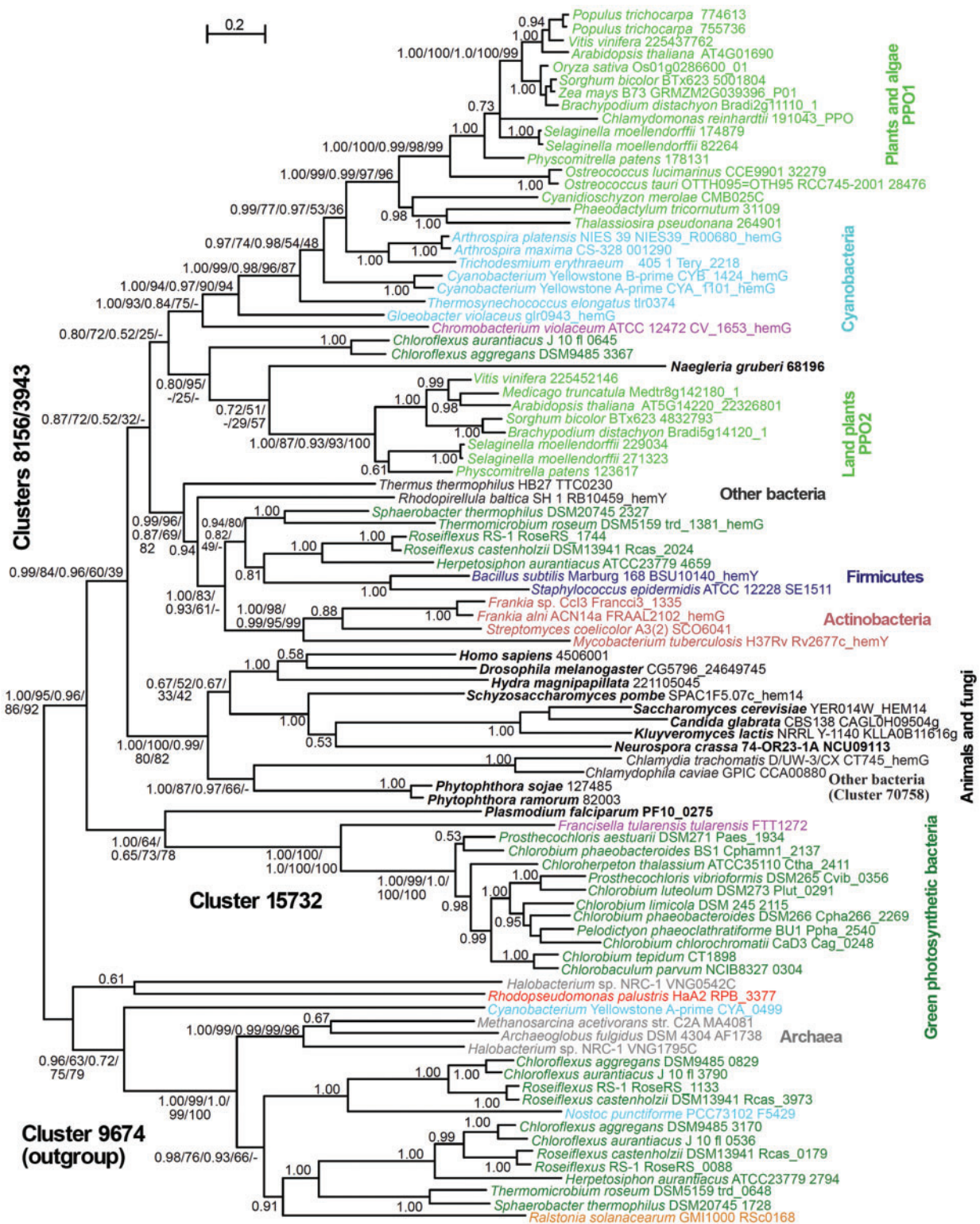


Fig. 4.—Phylogenetic tree of HemY. BI tree of HemY with confidence support values obtained with BI/ML/PB/RA/NJ methods shown at each branch. Only a BI score is shown in the interior of branches for visibility. Distance scale is indicated by the bar on top. The cluster numbers refer to those in the Gclust 2010 data set. Clusters 8156 and 3943 are mixed in the major HemY group. Land plants contain both PPO1 and PPO2. Cluster 15732 contains proteins that are probably not HemY. Cluster 9674 is an outgroup consisting of amine oxidases. Note that some proteins are erroneously annotated as HemG in the original databases.

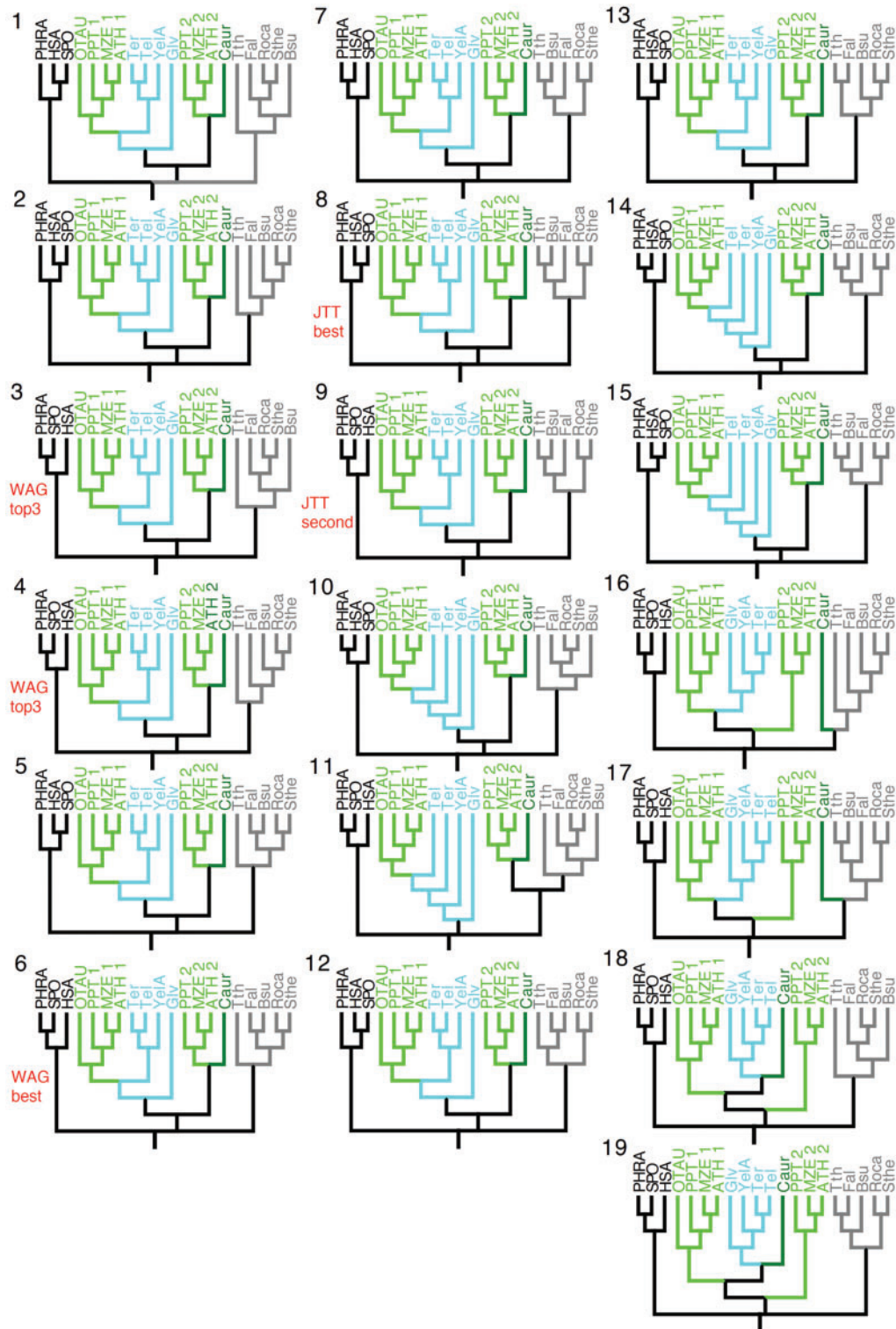


Fig. 5.—Phylogenetic trees used for AU testing. Twenty representative HemY proteins were selected, and the most likely trees were evaluated by AU test. Trees 1–15 were selected by the protml program. Trees 16–19 were tested as constrained trees. The results of the AU test are in table 3. Both JTT and WAG models were tested. Trees 1–15 should not be rejected as nonsignificant, whereas trees 16–19 were clearly abandoned. For names of eukaryotes, see table 1. Other names are the followings: Bsu, *Bacillus subtilis*; Caur, *Chloroflexus aurantiacus*; Fal, *Frankia alni*; Glv, *Gloeobacter violaceus*; Roca, *Roseiflexus castenholzii*; Sthe, *Sphaerobacter thermophilus*; Tel, *Thermosynechococcus elongatus*; Ter, *Trichodesmium erythraeum*; Tth, *Thermus thermophilus*; Yela, *Cyanobacterium Yellowstone A-prime*.

Table 3

Results of AU Test Using CONSEL Analysis

Rank	Item	Obs	AU
JTT model			
1	8	0	0.665
2	9	0	0.659
3	4	2.7	0.543
4	2	2.7	0.521
5	5	2.1	0.52
6	6	2.3	0.498
7	7	1.5	0.439
8	15	5.4	0.432
9	12	1.5	0.423
10	3	3.8	0.379
11	14	6.4	0.354
12	1	3.6	0.349
13	13	4.6	0.334
14	10	10.2	0.274
15	11	10.8	0.252
16	17	27.9	0.033
17	16	30.6	0.022
18	19	63	0.001
19	18	65.8	2E-04
WAG model			
1	4	0.2	0.626
2	6	-0.1	0.605
3	3	0.1	0.594
4	9	2.5	0.538
5	1	0.8	0.524
6	5	0.7	0.509
7	2	1.1	0.502
8	8	3.4	0.478
9	7	3.6	0.423
10	10	6.5	0.361
11	12	4.4	0.327
12	14	8.7	0.315
13	11	7.5	0.3
14	15	8.8	0.286
15	13	5.1	0.265
16	16	31.7	0.017
17	17	34.5	0.011
18	19	71.3	5E-04
19	18	68.4	3E-05

NOTE.—Item, tree number; Obs, log-likelihood difference.

Structural Differences of HemY

Crystal structures of HemY from *Nicotiana tabacum* (ntPPO2) (Koch et al. 2004), *Myxococcus xanthus* (mxPPO) (Corradi et al. 2006), *B. subtilis* (bsPPO) (Qin et al. 2010), and *Homo sapiens* (hsPPO) (Qin et al. 2011) have revealed that the overall folding patterns are similar among HemY orthologs from these species (Qin et al. 2011). Amino acid residues involved in substrate- and FAD-binding were identified on the basis of their 3D structures and kinetic analyses of mutant proteins (Koch et al. 2004; Corradi et al. 2006; Heinemann et al.

2007; Qin et al. 2010, 2011; Hao et al. 2013). Multiple alignment of the HemY proteins showed that key amino acids of HemY are basically conserved across prokaryotes and eukaryotes (supplementary fig. S7, Supplementary Material online). However, some notable characteristics were found in each clade. 3D structures of HemY homologs of *Arabidopsis thaliana* (PPO1 and PPO2), *Cyanidioschyzon merolae*, *C. aurantiacus*, and *R. castenholzii* were modeled by using the known structures as templates. Figure 6 compares the structures of various HemY proteins. The loops that are characteristic to each protein are marked with a “#” plus an identification number, which are also shown in the alignment (supplementary fig. S7, Supplementary Material online).

From crystallographic analyses, residues that interact with the substrate were proposed. According to previous researches (Koch et al. 2004; Corradi et al. 2006; Qin et al. 2011), Arg98, Phe353, and Leu356 of ntPPO2 (corresponding to Arg97, Phe331 and Leu334 of hsPPO and Arg95, Phe329 and Leu332 of mxPPO, respectively) interact with the D-ring of Protogen. Furthermore, in the crystal structure of mxPPO, these residues make key interactions with the diphenyl ether herbicide acifluorfen, which suggests an involvement of these residues in acifluorfen sensitivity. These substrate-binding residues are not conserved in the HemY homologs from the bacterial clade including *B. subtilis*. In bsPPO, these residues are replaced by Ser95, Thr330, and Val333, respectively, and may not participate in acifluorfen binding or substrate recognition (Qin et al. 2010). Indeed, bsPPO has a tolerance for acifluorfen (Corrigall et al. 1998). Because HemY proteins from the bacterial clade commonly lack the residues involved in acifluorfen binding, these enzymes probably tolerate diphenyl ether herbicides, as does bsPPO.

Another characteristic of common amino acid sequences in the bacterial clade is a short insertion, in the case of bsPPO, between $\beta 7$ and $\alpha 3$ (supplementary fig. S7, Supplementary Material online). 3D structure of this region marked #1 was invisible in the crystallographic analysis of bsPPO, presumably because of conformational flexibility (Qin et al. 2010). This region belongs to the membrane-binding domain in the structure of ntPPO2 and is assumed to affect cellular localization patterns of the enzymes. Considering that bsPPO is a soluble monomer whereas ntPPO, mxPPO, and hsPPO are membrane-bound dimers, this insertion may contribute to the unique localization of bsPPO and presumably that of other orthologs in the bacterial clade.

When amino acid sequences were compared between plant PPO1 and PPO2 isoforms, a remarkable difference was observed, in the case of ntPPO2, between the last half of $\alpha 5$ and the linker to $\alpha 6$ as marked as #2 (fig. 6). The ntPPO2 $\alpha 5$, which forms a dimer interface at the base of membrane-binding domain in the crystal structure, is assumed to function in dimerization and/or interaction with membrane lipids (Koch et al. 2004). Therefore, this region may be associated with distinct localization patterns between PPO1 and PPO2 as

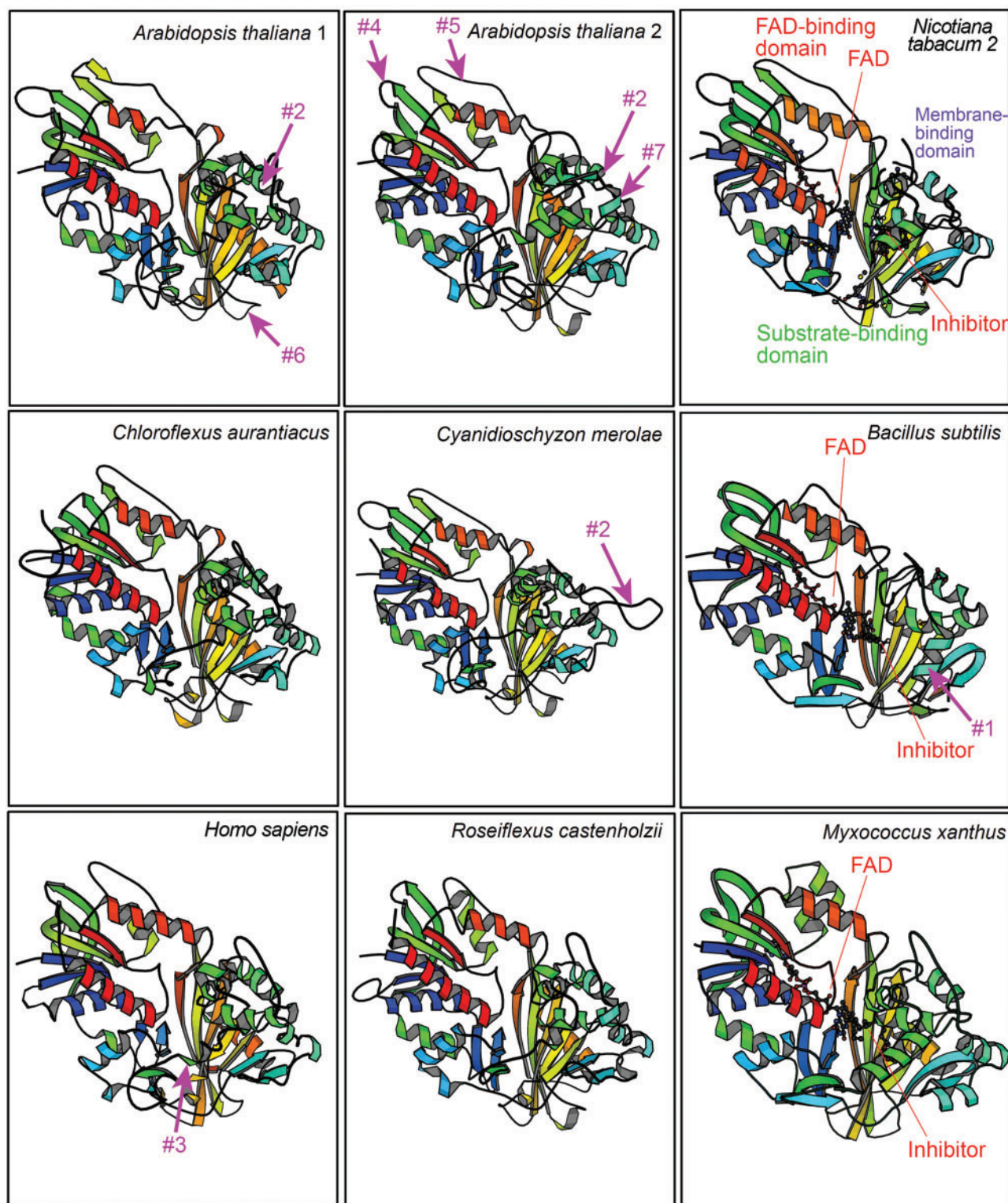


Fig. 6.—Comparison of the structures of various HemY proteins. Structures of various HemY proteins (shown on left and center) were estimated by homology modeling with *Nicotiana tabacum* PPO2, *Bacillus subtilis* HemY, and *Myxococcus xanthus* HemY. Characteristic structures are annotated with numbers, which correspond to the annotations in [supplementary figure S7, Supplementary Material](#) online. The rainbow color gradation from blue to red indicates direction from the N-terminus to C-terminus.

reported in tobacco (Lermontova et al. 1997) and spinach (Watanabe et al. 2001). Furthermore, the ntPPO2 α 5 is modeled to form an interface with ferrochelatase, which suggests an involvement of this region in protein–protein interaction with ferrochelatase and presumably also Mg-chelatase.

In addition, plant PPO2 isoforms have two unique amino acid insertions marked #5 and #7 (fig. 6), which correspond to β 11 and α 14 of ntPPO2, respectively. In the crystal structure of ntPPO2, β 11 is positioned within the FAD-binding domain but its specific function is unknown. α 14 of ntPPO2 forms dimer interface together with α 5, and may add unique characteristics to PPO2 isoforms in combination with α 5 for their localization. Meanwhile, plant PPO1 homologs have a unique insertion marked #6 in the loop between β 12 and β 13 of ntPPO2. In ntPPO2, hsPPO, and bsPPO, this region is positioned at the periphery of the substrate-binding domain near the dimer interface. Because the insertion in PPO1 homologs is rich in hydrophilic residues, this region may function in intermolecular interaction.

Concluding Remarks

We present an extensive analysis of isofunctional Protox enzymes in various organisms. Although the isofunctional enzymes seemed to move from one phylum to another often, we could provide a general view of the usage of isofunctional enzymes in prokaryotes and eukaryotes. Because these enzymatic activities are essential for producing Proto in almost all organisms for the synthesis of heme and Chl, studies are needed for the organisms that lack the genes for these isofunctional enzymes. The phylogenomic approach we used can shed more light on the still uncovered enzymes in the biosynthesis of Proto.

Supplementary Material

Supplementary figures S1–S7 and tables S1–S4 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

This work was supported in part by the Grants-in-Aid for Scientific Research on Priority Areas (No. 24770055), Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, and the Global Center of Excellence Program “From the Earth to Earths.” Most of the calculations for genome clustering involved use of the Supercomputers in the Human Genome Center, University of Tokyo.

Literature Cited

Abicht HK, Martinez J, Layer G, Jahn D, Solioz M. 2012. *Lactococcus lactis* HemW (HemN) is a haem-binding protein with a putative role in haem trafficking. *Biochem J.* 442:335–343.

- Adachi J, Hasegawa M. 1996. MOLPHY Version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monographs.* Vol. 28. Tokyo (Japan): Institute of Statistical Mathematics. p. 1–150.
- Boynton TO, et al. 2011. Discovery of a gene involved in a third bacterial protoporphyrinogen oxidase activity through comparative genomic analysis and functional complementation. *Appl Environ Microbiol.* 77:4795–4801.
- Boynton TO, Daugherty LE, Dailey TA, Dailey HA. 2009. Identification of *Escherichia coli* HemG as a novel, menadione-dependent flavodoxin with protoporphyrinogen oxidase activity. *Biochemistry* 48: 6705–6711.
- Camadro JM, Labbe P. 1996. Cloning and characterization of the yeast HEM14 gene coding for protoporphyrinogen oxidase, the molecular target of diphenyl ether-type herbicides. *J Biol Chem.* 271: 9120–9128.
- Cavallaro G, Decaria L, Rosato A. 2008. Genome-based analysis of heme biosynthesis and uptake in prokaryotic systems. *J Proteome Res.* 7: 4946–4954.
- Corradi HR, et al. 2006. Crystal structure of protoporphyrinogen oxidase from *Mycococcus xanthus* and its complex with the inhibitor acifluorfen. *J Biol Chem.* 281:38625–38633.
- Corrigall AV, et al. 1998. Purification of and kinetic studies on a cloned protoporphyrinogen oxidase from the aerobic bacterium *Bacillus subtilis*. *Arch Biochem Biophys.* 358:251–256.
- Dailey TA, Dailey HA. 1998. Identification of an FAD superfamily containing protoporphyrinogen oxidases, monoamine oxidases, and phytoene desaturase. Expression and characterization of phytoene desaturase of *Mycococcus xanthus*. *J Biol Chem.* 273:13658–13662.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Ferreiras GC, Andrew TL, Karr SW, Dailey HA. 1988. Organization of the terminal two enzymes of the heme biosynthetic pathway. Orientation of protoporphyrinogen oxidase and evidence for a membrane complex. *J Biol Chem.* 263:3835–3839.
- Galperin MY, Walker DR, Koonin EV. 1998. Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* 8:779–790.
- Hansson M, Hederstedt L. 1992. Cloning and characterization of the *Bacillus subtilis* hemEHY gene cluster, which encodes protoheme IX biosynthetic enzymes. *J Bacteriol.* 174:8081–8093.
- Hao G-F, et al. 2013. Computational and experimental insights into the mechanism of substrate recognition and feedback inhibition of protoporphyrinogen oxidase. *PLoS One* 8:e69198.
- Heinemann IU, et al. 2007. Functional definition of the tobacco protoporphyrinogen IX oxidase substrate-binding site. *Biochem J.* 402: 575–580.
- Heinemann IU, Jahn M, Jahn D. 2008. The biochemistry of heme biosynthesis. *Arch Biochem Biophys.* 474:238–251.
- Jacobs JM, Jacobs NJ. 1993. Porphyrin accumulation and export by isolated barley (*Hordeum vulgare*) plastids (effect of diphenyl ether herbicides). *Plant Physiol.* 101:1181–1187.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol.* 4:18.
- Kato K, Tanaka R, Sano S, Tanaka A, Hosaka H. 2010. Identification of a gene essential for protoporphyrinogen IX oxidase activity in the cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A.* 107:16649–16654.
- Koch M, et al. 2004. Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis. *EMBO J.* 23: 1720–1728.
- Kraulis PJ. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr.* 24: 946–950.

- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
- Lee HJ, Duke MV, Duke SO. 1993. Cellular localization of protoporphyrinogen-oxidizing activities of etiolated barley (*Hordeum vulgare* L.) leaves (relationship to mechanism of action of protoporphyrinogen oxidase-inhibiting herbicides). *Plant Physiol.* 102:881–889.
- Lermontova I, Grimm B. 2006. Reduced activity of plastid protoporphyrinogen oxidase causes attenuated photodynamic damage during high-light compared to low-light exposure. *Plant J.* 48:499–510.
- Lermontova I, Kruse E, Mock HP, Grimm B. 1997. Cloning and characterization of a plastidial and a mitochondrial isoform of tobacco protoporphyrinogen IX oxidase. *Proc Natl Acad Sci U S A.* 94:8895–8900.
- Masoumi A, et al. 2008. Complex formation between protoporphyrinogen IX oxidase and ferredoxin during haem biosynthesis in *Thermosynechococcus elongatus*. *Microbiology* 154:3707–3714.
- Matringe M, Camadro J, Labbet P, Scalla R. 1989. Protoporphyrinogen oxidase as a molecular target for diphenyl ether herbicides. *Biochem J.* 260:231–235.
- Meissner PN, et al. 1996. A R59W mutation in human protoporphyrinogen oxidase results in decreased enzyme activity and is prevalent in South Africans with variegate porphyria. *Nat Genet.* 13:95–97.
- Möbius K, et al. 2010. Heme biosynthesis is coupled to electron transport chains for energy generation. *Proc Natl Acad Sci U S A.* 107:10436–10441.
- Molina A, et al. 1999. Inhibition of protoporphyrinogen oxidase expression in *Arabidopsis* causes a lesion-mimic phenotype that induces systemic acquired resistance. *Plant J.* 17:667–678.
- Nagaraj VA, Arumugam R, Prasad D, Rangarajan PN, Padmanaban G. 2010. Protoporphyrinogen IX oxidase from *Plasmodium falciparum* is anaerobic and is localized to the mitochondrion. *Mol Biochem Parasitol.* 174:44–52.
- Nowack ECM, Melkonian M, Glöckner G. 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol.* 18:410–418.
- Obornik M, Green BR. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol.* 22:2343–2353.
- Omelchenko MV, Galperin MY, Wolf YI, Koonin EV. 2010. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol Direct.* 5:31.
- Panek H, Brian MRO. 2002. A whole genome view of prokaryotic haem biosynthesis. *Microbiology* 148:2273–2282.
- Perrière G, Gouy M. 1996. WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* 78:364–369.
- Qin X, et al. 2010. Structural insight into unique properties of protoporphyrinogen oxidase from *Bacillus subtilis*. *J Struct Biol.* 170:76–82.
- Qin X, et al. 2011. Structural insight into human variegate porphyria disease. *FASEB J.* 25:653–664.
- Rao AU, Carta LK, Lesuisse E, Hamza I. 2005. Lack of heme synthesis in a free-living eukaryote. *Proc Natl Acad Sci U S A.* 102:4270–4275.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 234:779–815.
- Sasaki NV, Sato N. 2010. Elucidating genome structure evolution by analysis of isoapostatic gene clusters using statistics of variance of gene distances. *Genome Biol Evol.* 2:1–12.
- Sasarman A, et al. 1993. Nucleotide sequence of the hemG gene involved in the protoporphyrinogen oxidase activity of *Escherichia coli* K12. *Can J Microbiol.* 39:1155–1161.
- Sato N. 2000. SISEQ: manipulation of multiple sequence and large database files for common platforms. *Bioinformatics* 16:180–181.
- Sato N. 2002. Comparative analysis of the genomes of Cyanobacteria and plants. *Genome Inform.* 13:173–182.
- Sato N. 2009. Gclust: trans-kingdom classification of proteins using automatic individual threshold setting. *Bioinformatics* 25:599–605.
- Sato N, Ishikawa M, Fujiwara M, Sonoike K. 2005. Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups. *Genome Inform.* 16:56–68.
- Shih PM, et al. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A.* 110:1053–1058.
- Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
- Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Storbeck S, et al. 2010. A novel pathway for the biosynthesis of heme in Archaea: genome-based bioinformatic predictions and experimental evidence. *Archaea* 2010: Article ID 175050.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
- Tanaka R, Kobayashi K, Masuda T. 2011. Tetrapyrrole metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* 9:e0145.
- Tang K-H, et al. 2011. Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC Genomics* 12:334.
- Thompson JD, Higgins DG, Gibson TJ. 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.* 22:4673–4680.
- Warnich L. 1996. Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South African families with variegate porphyria. *Hum Mol Genet.* 5:981–984.
- Watanabe N, et al. 2001. Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. *J Biol Chem.* 276:20474–20481.
- Witkowski DA, Halling BP. 1989. Inhibition of plant protoporphyrinogen oxidase by the herbicide acifluorfen-methyl. *Plant Physiol.* 90:1239–1242.
- Xu K, Dellinger J, Elliott T. 1992. The genes required for heme synthesis in *Salmonella typhimurium* include those encoding alternative functions for aerobic and anaerobic coproporphyrinogen oxidation. *J Bacteriol.* 174:3953–3963.

Associate editor: Geoff McFadden