SYMBIOSIS (2005) 39, 83–90 ©2005 Balaban, Philadelphia/Rehovot ISSN 0334-5114

Characterization of a gene locus containing squalene-hopene cyclase (*shc*) in *Frankia alni* ACN14a, and an *shc* homolog in *Acidothermus cellulolyticus*

Nicole Alloisio[1], Joelle Maréchal[1], Brian Vanden Heuvel[2], Philippe Normand[1], and Alison M. Berry[2*]

[1]UMR CNRS 5557 Ecologie Microbienne, Université Claude Bernard, Lyon 1, Villeurbanne 69622 Cedex, France;
[2]Department of Plant Sciences, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA,
Tel. +1-530-752-7683, Fax. +1-530-752-4361, Email. amberry@ucdavis.edu

(Received October 24, 2004; Accepted March 31, 2005)

Abstract
Hopanoid lipids are pentacyclic triterpenes that are synthesized by eubacterial prokaryotes, including the two phylogenetically-related actinobacteria *Frankia* and *Acidothermus*, and that function in protection against environmental stress. In the nitrogen-fixing *Frankia*, hopanoids in the vesicle envelope reduce oxygen diffusion to nitrogenase, while in the thermophile *Acidothermus*, hopanoids may play a role in adaptation to high temperature and low pH. A 26,335 nucleotide region of the *Frankia alni* genome encoding the squalene-hopene cyclase gene (*shc*) was cloned and sequenced. Contrary to several characterized *shc* genes in other organisms, the *Frankia shc* gene was not part of a hopanoid biosynthetic cluster. Rather, the neighboring genes included several regulators, oxidoreductases, transmembrane proteins as well as multidrug transporters. A second copy of the *shc* gene, detected by Southern hybridization, is likely to be present in *F. alni*. A homologous gene was amplified and sequenced in *Acidothermus cellulolyticus* and found to have a sequence distance of 36.9% from that of *Frankia*. Within the actinobacteria, phylogenetic analysis shows a close relationship between the shc gene sequences of *F. alni* and *A. cellulolyticus*.

Keywords: Actinorhizal, hopanoid, oxygen, nitrogenase, temperature, actinobacteria, *Frankia*, *Acidothermus*

1. Introduction

Hopanoid lipids are pentacyclic triterpenes synthesized by eubacterial prokaryotes which play a role in stabilizing cell membranes, particularly in extreme environments, that is directly analogous to the role of sterols such as cholesterol in eukaryotic cells (Ourisson and Rohmer, 1992). Squalene-hopene cyclase is the defining enzyme in the biosynthesis of hopanoid lipids, catalyzing the cyclization of squalene into the characteristic pentacyclic triterpenoid structure that forms the basis of this class of molecules. The gene coding for squalene-hopene cyclase, *shc*, has been identified in a number of eubacteria, particularly those that inhabit extreme environments like ethanol solutions in the case of the alpha-Proteobacterium *Zymomonas mobilis* (Reipen et al., 1995), or hot springs for the Firmicute *Alicyclobacillus acidocaldarius* (Full and Poralla, 2000) or the recently characterized actinobacterium *Rubrobacter xylanophilus* (Ferreira et al., 1999).

In several organisms, *shc* is located within a cluster of genes that code for several other steps in the hopanoid biosynthetic pathway (Poralla et al., 2000), including squalene synthases, squalene dehydrogenase, and 1-deoxy-D-xylulose 5-phosphate synthase, the first step in the non-mevalonate pathway leading to terpenoid formation (Eisenreich et al., 2001).

Frankia is an aerobic, filamentous actinobacterium that forms nitrogen-fixing root nodule symbioses in association with actinorhizal plant hosts, members of eight angiosperm families. *Frankia* is capable of fixing nitrogen at atmospheric oxygen concentrations, even though the nitrogenase enzyme is irreversibly inactivated by oxygen. Nitrogen fixation in *Frankia* is localized in specialized cellular structures called vesicles, in which a thick lipid envelope provides a barrier to oxygen diffusion (Benson and

*The author to whom correspondence should be sent.

Presented at the 13th International Symposium on *Frankia* and Actinorhizal Plants, August 1–4, 2004, Durham, NH, USA

Silvester, 1993). The vesicle envelope is made of multiple layers of two particular hopanoid lipids, bacteriohopanetetrol (Berry et al., 1991), and a phenylacetic acid ester derivative (Berry et al., 1993) not described in any other organism to date. Phenylacetic acid (PA) was shown to be released from *in vitro* grown *F. alni* cells (Hammad et al., 2003). PA functions as a plant auxin, and thus may play a role in growth regulation of the nodule symbiosis.

Acidothermus is closely related to *Frankia* by phylogenetic analysis with both 16S and *recA*, and like *Frankia*, produces high levels of hopanoids (Maréchal et al., 2000; Normand et al., 1996). As an extremophile that was isolated from a hot spring in Yellowstone National Park (Mohagheghi et al., 1986), *Acidothermus cellulolyticus* has an optimal growth temperature of 55°C. *A. cellulolyticus* is thus an unexpected phylogenetic neighbor of *Frankia*. The contrasting ecological adaptations of these two organisms (high temperature tolerance, oxygen protection) may be related to their capacity to produce hopanoids. It is hypothesized that the biosynthesis of high levels of hopanoids, a feature shared by several bacteria living in various extreme environments, may be evidence of the evolutionary link between *Frankia* and *Acidothermus*.

We are interested in characterizing and comparing hopanoid biosynthesis in *Frankia* and *Acidothermus*. Earlier, we cloned *shc* partial sequences from several *Frankia* strains (Dobritsa et al., 2001). In the current work, we report the cloning, sequencing, and characterization of a gene region containing the full-length shc gene and several flanking genes, from *Frankia alni* ACN14a. *F. alni* strain ACN14a nodulates *Alnus* spp., and belongs to genomic cluster I in the Frankiaceae (Normand et al., 1996). In addition, we report the amplification of a DNA sequence from *A. cellulolyticus* with high homology to *F. alni shc*.

2. Materials and Methods

Cloning of the shc *gene region*

Total genomic DNA from *Frankia alni* strain ACN14a (Normand and Lalonde, 1982) was isolated and purified with the Dneasy Plant Mini kit according to manufacturer's instructions (Qiagen, Courtaboeuf, France). To amplify a *shc* gene fragment of this strain, we used primers SHS (5'-CSWCTACRWGATCCCGTTCG-3') and SHC6 (5'-CCTCGCCCCAGCCGCCATCCTCGTTCTG-3') previously designed to amplify an *shc* gene fragment from *Frankia* strain CpI1 (Dobritsa et al., 2001). PCR amplification was carried out according to Dobritsa et al. (2001) with minor modifications, i.e. 50 ng of ACN14a DNA were used and 40 PCR cycles were performed at 95°C for 60 sec, 54°C for 45 sec and 72°C for 45 sec.

The resulting amplified fragment of about 320 nucleotides (nt) was sequenced and used to design a second generation of primers specific to *Frankia shc* gene (1694F: 5'-CCGTCGTCCGCCGACGT-3' and 1584R: 5'-AGCAGCCASTSGATGCC-3'). The *Frankia* strain ACN14a gene library (Cournoyer and Lavire, 1999) built with the low-copy vector pRK404 (Ditta et al., 1985) was screened by PCR amplification with these primers (1694F and 1584R) according to the instructions of the *Taq* DNA polymerase manufacturer (Invitrogen™ Life Technologies, Cergy Pontoise, France).

After initial denaturation at 95°C for 5 min, 40 cycles were performed at 95°C for 60 sec, 54°C for 45 sec, 72°C for 45 sec, with final extension at 72°C for 5 min. A clone designated pFAQshc was isolated. pFAQshc plasmid DNA was purified with the QIAprep Spin miniprep kit (Qiagen) and digested with *Bam*HI, *Bgl*II and *Xho*I restriction enzymes, respectively. *Bam*HI and *Bgl*II plasmid fragments were ligated with dephosphorylated pUC18 Ready-to-use vector cut by *Bam*HI (Qbiogene, Illkirch, France), and *Xho*I fragments with dephosphorylated pUC19 Ready-to-use vector cut by *Sal*I (Qbiogene) and subcloned in Efficiency™ DH5α™ chemically-competent *Escherichia coli* cells (Invitrogen™ Life Technologies, Cergy Pontoise, France). pUC plasmid inserts were completely sequenced by GENOME Express (Meylan, France) by primer walking. Sequence inserts were assembled and open reading frames (ORFs) were identified with the FramePlot program (Ishikawa and Hotta, 1999) and confirmed with the Genostar software (Genostar consortium: Institut Pasteur-INRA Hybrigenics SA-GENOME Express, France). Database searches were performed with the BLAST programs (Altschul et al., 1997).

In order to sequence a longer DNA region upstream of *shc* gene, we screened the above library by PCR for a clone containing the transcriptional regulator immediately upstream of *shc* gene but not the *shc* gene. Transcriptional regulator amplified fragment was obtained with primers 1919F (5'-TGTCACACGGACGCAGTAG-3') and 1809R (5'-AGGCCGTCAGGGTGAGCT-3') and *shc* amplified fragment with primers 1694F and 2072R (5'-CAGCAGCCAGTCGACGCC-3'), instead of 1584R. PCR amplifications were performed with *Taq* DNA polymerase (Invitrogen™) as described above, with addition of 5% v/v dimethylsulfoxide (Sigma-Aldrich, Saint Quentin Fallavier, France) in the PCR reactions and a 55°C temperature for hybridization. A clone designated pFAQ18 was isolated. pFAQ18 plasmid insert was directly sequenced by primer walking and by sequencing *Xho*I restriction enzyme fragments subcloned as described above. Gene identification was performed as described above. The Genbank accession number of the *Frankia alni* nucleotide sequence and identified ORFs is AJ854181 and that of the *A. cellulolyticus shc* sequence is AY954503.

Global comparisons of individual ORFs were carried out using the Genome utilities at NCBI http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome

The translated nucleotide sequence of the N-terminal domain of ORF 11 was compared with conserved signaling

Table 1. List of ORFs present on plasmids pFAQ18 and pFAQshc and of corresponding homologous proteins. pFAQ18: nucleotides 1 to 15618; pFAQshc: nucleotides 10718 to 26335.

ORF	Coordinates (nt)	Length (nt)	Coding strand	Similar protein in databases	% identity	E value
1 (partial)	3–479	477	+	Butyrate-acetoacetate CoA-transferase, subunit B (EC 2.8.3.9) *Deinococcus radiodurans* (pir/E75600)	70 (106/151)	4e-55
2	907–1647	741	+	Rhamnolipids biosynthesis 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) *Pseudomonas aeruginosa* (swp/Q9RPT1)	36 (79/216)	9e-24
3	4259–2055	2205	–	Putative AfsR-like regulatory protein *Streptomyces coelicolor* (ref/NP_628294.1)	33 (173/515)	3e-51
4	4380–5183	804	+	Hypothetical protein, *Corynebacterium efficiens* (ref/NP_438088.1)	37 (53/143)	3e-19
5	5836–5426	411	–	Lactoylglutathione lyase and related lyases *Ralstonia eutropha* (ref/ZP_00167123)	31 (42/135)	3e-18
6	6978–9140	2163	+	Putative two-component system sensor kinase, *Streptomyces avermitilis* (ref/NP_824808.1)	40 (94/230)	2e-36
7	9005–9682	678	+	Transposase TnpA, *Corynebacterium glutamicum* (ref/NP_478076.1)	42 (90/210)	2e-41
8	10526–9996	531	–	Putative integral membrane protein *Steptomyces coelicolor* (ref/NP_625342.1)	36 (64/174)	6e-14
9	10907–11560	654	+	Putative TetR-family transcriptional regulator, *Streptomyces avermitilis* (ref/NP_826973.1)	37 (69/184)	5e-18
10	11553–13796	2244	+	Putative squalene-hopene cyclase *Streptomyces avermitilis* (ref/NP_826973.1)	53 (380/713)	0.0
				Putative squalene-hopene cyclase *Streptomyces coelicolor* (ref/NP_630836.1)	53 (383/712)	0.0
11	13930–15642	1713	+	Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain *Magnetospirillum magnetotacticum* (ref/ZP_00053091.2)	30 (147/476)	2e-42
12	16613–15576	1038	–	Probable transmembrane protein *Ralstonia solanacearum* (ref/NP_519271.1)	31 (109/346)	2e-37
13	17542–16850	693	–	Conserved hypothetical protein *Streptomyces coelicolor* (ref/NP_624786.1)	59 (131/220)	8e-70
14	18753–17668	1086	–	Oxidoreductase, *Streptomyces coelicolor* (ref/NP_631415.1)	41 (138/334)	1e-59
				Phenylacetaldehyde reductase *Corynebacterium* sp. (dbj/BAA35108.1)	38 (123/316)	4e-48
15	19654–18821	834	–	Hypothetical protein, *Streptomyces coelicolor* (ref/NP_628799.1)	58 (159/273)	6e-86
				S-adenosylmethionine-dependent methyltransferases *Kineococcus radiotolerans* (ref/ZP_00226732.1)	31 (77/245)	1e-19
16	19785–20408	624	+	Conserved hypothetical protein *Streptomyces coelicolor* (ref/NP_628800.1)	62 (123/197)	3e-59
				Putative TetR-family transcriptional regulator, *Streptomyces coelicolor* (ref/NP_628800.1)	49 (94/191)	8e-38
17	20928–20371	558	–	Predicted ATPase, *Burkholderia cepacia* (ref/ZP_00212365.1)	55 (98/176)	6e-52
18	22075–23274	1200	+	Related to alkanesulfonate monooxygenase (EC 1.14.14.5) *Neurospora crassa* (emb/CAD70336.1)	53 (199/370)	e-107
19	23449–25299	1851	+	ABC-type multidrug transport system, ATPase and permease components *Kineococcus radiotolerans* (ref/ZP_00227645.1)	42 (258/604)	e-118
20 (partial)	25391–26335	945	+	ABC-type multidrug transport system, ATPase and permease components *Thermobifida fusca* (ref/ZP_00291556.1)	35 (106/298)	5 e-40

domains by successive alignments in ClustalW (http://www.ebi.ac.uk/clustalw/), and the resulting subsequence was searched for conserved residues and spacing between key residues, by visual comparison with a catalog of PAS domains (Taylor and Zhulin, 1999).

Southern blot analysis

A *shc* 304-nt probe (corresponding to nucleotides 13174 to 13477 of the 26335-nt sequence and to nucleotides 1622 to 1925 of the *shc* gene, see Table 1) was made by PCR with primers 2070F (5'-TCGTCTACCAGATTCCGTTC-3') and 2071R (5'-TCGGAGTTCTGCCGCGACA-3') and labeling with [α^{32}P] dCTP (ICN, Orsay, France) using the Random primed DNA labeling kit (Roche, Meylan, France). *Frankia alni* genomic DNA and plasmid pFAQshc were digested by *Pst*I, *Sma*I and *Bam*HI restriction enzymes, separated by electrophoresis, transferred to a nylon membrane and hybridized with the *shc* DNA probe as described by Sambrook et al. (1989), using the ExpressHyb hybridization solution (BD Biosciences Clontech, Palo Alto, CA).

shc *sequence from* Acidothermus cellulolyticus

A. cellulolyticus type strain 8747 (Mohagheghi et al., 1986) was grown for 48–72 h at 55°C in liquid ATCC 1473 medium (www.atcc.org.mediapdfs/1473.pdf) except that glucose (1 g/l) was substituted for Sigmacell Type 50. One liter of cell culture was harvested by centrifugation at 10,000 x g (4°C) for 10 minutes. The resulting pellets were combined, washed with dH_20 and pelleted again by centrifugation. Total genomic DNA was extracted using a modified CTAB (cetyl-trimethylammonium bromide) method (Doyle and Doyle, 1987). Briefly, 6 ml of 2x CTAB (2% w/v CTAB, 1.4 M NaCl, 100 mM Tris-HCl, and 20 mM EDTA, 65°C) was added to the washed cells and incubated at 65°C for 1 hour. Chloroform/isoamyl alcohol (4 ml; 14:1 v/v) was added with shaking to create an emulsion. The aqueous layer was harvested after centrifugation at 14,000 x g for 3 min, retreated with chloroform/isoamyl alcohol, and recentrifuged. The resulting aqueous layer was harvested.

Genomic DNA was precipitated with 6 ml of ice-cold isopropanol, and placed at –20°C for 12 hours. The genomic DNA was washed in 70% v/v ethanol and the DNA pellet air-dried, then redissolved in 100 µl of 10 mM Tris-HCl. For PCR primer design, an alignment based on the squalene-hopene cyclase (*shc*) gene sequence cloned from *Frankia* ACN14a as described above and in the Results, together with the *shc* full-length sequence from *Streptomyces coelicolor* (Genbank AL939129, positions 220066–222066), and *S. avermitilis* (GenBank AP005027 positions 54266–56308), revealed three conserved regions of 50 base pairs or longer with sequence similarity greater than 85% common to all three sequences.

Four primers were designed within these conserved regions. Positions within the primer that contained substitutions among shc sequences were designed as degenerate. The primer sequences used were:
976F (5'-CGCTGGATCATCGMSCGGCAGGAG-3');
1397F (5'-CTACCCCGACAYCGACGACACCGCC-3');
1397R (5'-TCGGCGGTGTCGTCGRTGTCGGGG-3') and
1645R (5'-TCGACGACGTGCGCGGTGACGTC-3').

Primer 976F corresponds to the *F. alni* ACN14a positions 12528–12552; primers 1397F and 1397R correspond to positions 12984–13011 and primer 1645R corresponds to positions 12984–13008.

The primers developed for the *shc* gene were tested using *A. cellulolyticus* genomic DNA. PCR reactions (50 µl) were performed using 30 mM Tricine (Tris), pH 8.4, 2 mM $MgCl_2$, 50 mM KCl, 5% w/v acetamide, 5% v/v DMSO, 100 µM of each dNTP, 1.0 unit of *Taq* polymerase, and 30 nM of each amplification primer. 50 ng of template DNA was added and the reactions run on a Perkin Elmer Thermal Cycler (Perkin-Elmer Corp, Norwalk, CT) programmed for a hot start (95°C, 2 min) and 30 cycles of 94°C for 30 sec, 50°C for 45 sec, and 72°C for 1 min with a final extension for 7 min. A positive control containing *Frankia* genomic DNA (strain CpI1) and a negative control, lacking template DNA, were also run. PCR amplifications were analyzed in a 1.5% w/v agarose gel run in 0.5X TBE buffer (45 mM Tris borate, 1 mM EDTA; Sambrook et al., 1989) and visualized with ethidium bromide under UV light.

Fragments were purified using a QIAquick Gel Extraction kit (Qiagen, Santa Clarita, CA). PCR products were cycle sequenced in both directions using Big Dye reagents (Applied Biosystems, Foster City, CA) and analyzed on an ABI Prism 3100 automated sequencer (Applied Biosystems, Foster City, CA) according to manufacturer protocols. Sequences were proofread using SeqMan (1998, DNAstar, Madison, WI) and then verified with BLAST (www.ncbi.nlm.nih.gov/BLAST/) to ensure that the *shc* region was being amplified.

Phylogenetic analysis

Putative *shc* gene fragments generated from *Acidothermus cellulolyticus* were aligned with the *shc* fragments from *Frankia alni* ACN14a, *Rubrobacter xylanophilus* (NZ_AAEB01000005), *Streptomyces avermitilis* (BA000030), *Streptomyces coelicolor* (NC003888), *Alicyclobacillus acidocaldarius* (AB007002) and *Alicyclobacillus acidoterrestris* (X89854) using ClustalW in MegAlign (DNAstar, 1998). A single dataset consisting of the conserved, alignable regions for all sequences including *Acidothermus* was imported into PAUP 4.0b810 (Swofford, 2003) where a maximum parsimony exhaustive search was conducted. Support for clades was estimated using 1000 bootstrap replicates (Felsenstein, 1985) as implemented in PAUP.

3. Results and Discussion

Southern blot analysis with the *shc* probe revealed (i) clear bands of similar sizes in both genomic DNA and plasmid pFAQshc, i.e. 1500 and 700 nt with *Bam*HI, 780 and 250 nt with *Pst*I and 750 nt with *Sma*I; and (ii) additional bands that hybridized only with genomic DNA, i.e. two *Bam*HI bands of 1400 and 900 nt and two *Sma*I bands of 1550 and 550 nt (Fig. 1). Analysis of the restriction sites present in the genomic DNA sequence showed that these additional bands could not be due to partial digestions. Moreover, except for the higher 1550-nt *Sma*I band, the bands common to genomic DNA and plasmid pFAQshc, and the additional bands, revealed similar intensities. These lines of evidence taken together suggest the presence of two shc genes in the *Frankia alni* ACN14a genome.

The pFAQshc genomic DNA segment contained 15,618 nt and the pFAQ18 genomic DNA segment contained 12,254 nt of which 10,717 nt were localized upstream of the pFAQshc segment and 1,537 nt overlapped it. Thus a total genomic region of 26,335 nt could be identified around the *shc* gene. The %G+C of this region is 71.6%, which lies between the overall genome value of 68% found for a number of strains (An et al., 1985) and the 73.4% found for a 8989-nt genome fragment comprising the *sodF* gene in *F. alni* (Hammad et al., 2001).

Eighteen complete and two partial putative ORFs were recognized in this fragment (Table 1). The *shc* gene was surrounded by nine upstream genes and ten downstream genes. The identified gene products were three transcriptional regulators (ORFs 3, 9, 16); two signal transduction regulators (ORFs 6, 11); two ABC-transporters (ORFs 19, 20); two transmembrane proteins (ORFs 8, 12); three enzymes implicated respectively in benzoic acid metabolism (ORF 1), rhamnolipid biosynthesis (ORF 2) and glyoxal pathway (ORF 5); one S-adenosylmethionine-dependent methyltransferase (ORF 15); one ATPase (ORF 17); an enzyme related to alkanesulfonate monooxygenase (ORF 18); one oxidoreductase (ORF 14); one transposase (ORF 7) and two hypothetical proteins (ORFs 4, 15).

The *F. alni shc* gene shows high sequence similarity to the *Streptomyces coelicolor shc* gene (Table 1). However the *shc* locus in *S. coelicolor*, like that reported for *Zymomonas mobilis* and *Bradyrhizobium japonicum* (Perzl et al., 1998), consists of a cluster of isoprenoid and hopanoid biosynthetic genes (Poralla et al., 2000), unlike the region around the *F. alni shc* gene. The presence of a gene coding for a transposase (ORF 7) near *shc* in this DNA segment suggests the possibility of a past transpositional event involving the *shc* gene.

Interestingly the TetR-family transcriptional regulator (ORF 9) overlaps the downstream *shc* gene, suggesting a control function in *shc* gene transcription. The genome of *Geobacter sulfurreducens* contains two copies of *shc*, one of which (*shc-1*) is located in a hopanoid biosynthetic cluster

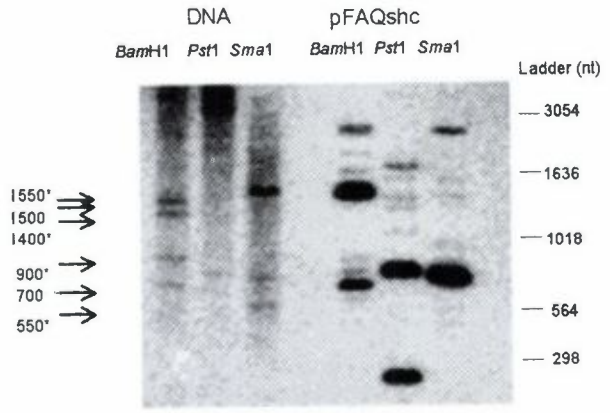

Figure 1. Genomic DNA and plasmid pFAQshc were digested by *Bam*H1, *Pst*I and *Sma*I, separated by electrophoresis, transferred to a nylon membrane and hybridized with the *shc* probe. *: Additional bands suggesting the presence of a second *shc* gene.

(Methe et al., 2003). The second copy (*shc-2*) has an adjacent ORF coding for a TetR-family promoter upstream of *shc*, but no genes related to hopanoid biosynthesis, similar to the *F. alni shc* gene locus described here. It will be of interest to determine whether a second copy of shc exists in *Frankia*, and if so, whether it is located in a hopanoid biosynthetic cluster. Screening of the *F. alni* strain ACN14a gene library by hybridization with the *shc* probe and Southern blot analysis of all the positive clones could answer this question; however genome sequencing of *F. alni* strain ACN14a in progress at Genoscope (www.genoscope.cns.fr/externe/English/Projets/Projet_HF/organisme_HF.html) will give the most precise answer.

Further, we point out that ORF 14, the fourth ORF downstream from shc, belongs to the family of zinc-containing, NAD^+-requiring oxidoreductases (pfam 00701.1), and possibly codes for a phenylacetaldehyde reductase that could convert the substrate to phenylalcohol. In *Corynebacterium*, this pathway is involved in xenobiotic degradation (Wang et al., 1999). In *Frankia*, phenylacetaldehyde is also a likely precursor for phenylacetic acid, which plays a unique role in relation to hopanoid lipid biosynthesis, thus contributing to oxygen protection of nitrogenase. During differentiation of the *Frankia* vesicle, where nitrogen fixation ultimately takes place, a multilayered envelope composed of hopanoid lipids forms around the vesicle. This vesicle envelope is comprised of bacteriohopanetetrol and its phenylacetate monoester, the PA-tetrol being specific to the vesicle while the tetrol is found in both hyphal cells and vesicles (Berry et al., 1993). The proximity of ORF 10 and ORF 14 suggests the possibility of an interaction between hopanoid biosynthesis and ORF 14 function, although the genes are transcribed from opposite DNA strands.

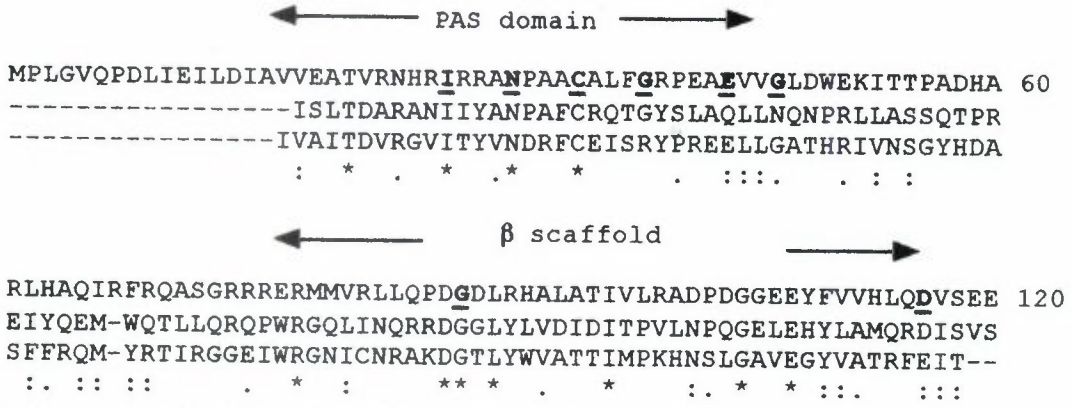

Figure 2. Alignment of N-terminal regions of three regulatory ORFs, corresponding to the conserved PAS redox sensor domain. *F. alni* ORF 11 aligns with the sensor domains in the region between residues 17–118 (residues 29–130 in AxDGC1). Asterisks (*) denotes identity; colon (:) indicates conserved substitutions within an amino acid class; period (.) indicates semi-conserved substitutions. Conserved residues of the PAS domain are underlined.

ORF 11 contains a GGDEF/EAL domain, which is involved in the regulation of prokaryotic cellular functions via c-di-GMP signaling (Tal et al., 1998). Further, a PAS sensor domain was identified in the N-terminal region of ORF 11, by a combination of sequence alignment and comparison with published data. The three sequences shown in Fig. 2 share several conserved amino acids within the domain. AxDGC1 (*Acetobacter xylinum* DGC1; AF062517) is a GGDEF/EAL enzyme that catalyzes the synthesis of c-di-GMP, acting as one component of an operon that regulates cellulose biosynthesis and cell differentiation post-transcriptionally (via c-di-GMP signaling), in response to environmental signals. The N-terminal sensor domain of NifL (KpNifL, *Klebsiella pneumoniae*; X13303.1) functions in oxygen perception, either through redox reactions or by direct oxidation (Bauer et al., 1999). Under oxidizing conditions, NifL regulates expression of nitrogen-fixing genes, acting as a suppressor of the positive regulatory gene *nifA*. The *F. alni* translated sequence exhibited 19–20% overall identity with the corresponding domains of AxDGC1 and KpNifL, respectively. The identity between AxDGC1 and KpNifL domains was higher, at 31%. By comparison, the *F. alni* sensor domain had relatively lower homology to a PAS redox sensor region of *E. coli* (13%; Aer), and a eukaryotic light sensor from *Arabidopsis* (16%; NPH1; sequence data not shown). There was no sequence homology between ORF 11 and KpNifL beyond this sensor domain, indicating that additional shared functional regions were not present.

The region of shared homology identified in Fig. 2 was compared with published databases for related sensor domains. *F. alni* ORF 11 showed extremely high conservation both in specific residues and the distance between them, in comparison with the consensus PAS/β-scaffold residues compiled by Taylor and Zhulin (1999). These key residues are underlined in bold (I, N, C, G, E and G in the PAS domain; G and D separated by 28 a.a. in the β-scaffold), and the parameters of the PAS/β-scaffold are indicated by horizontal arrows, in Fig. 2.

The evidence thus suggests that *F. alni* ORF 11 functions as a redox sensor, and also as a regulator of cell function through cyclic di-GMP. The homology with DGC1 may indicate that ORF 11 is a positive regulator under oxidizing conditions. Interestingly, during vesicle differentiation, the thickness of the hopanoid vesicle envelope is regulated by oxygen, such that an increase in oxygen tension results in an increase in the number of envelope layers deposited (Parsons et al., 1987). Given the location of this gene adjacent to and in the same frame as *shc*, it is conceivable that ORF 11 plays a regulatory role in the synthesis of vesicle hopanoid lipids in response to oxygen.

The aligned dataset that included the putative shc fragment from *A. cellulolyticus* contained 295 base pairs, 19 of which were potentially phylogenetically informative. The shortest tree of length 96 can be seen in Fig. 3. This phenogram is topologically similar to a cladogram produced in a previous study using partial shc sequences (Dobritsa et al., 2001), except for the inclusion of *A. cellulolyticus* and *R. xylanophilus* in the current *shc* tree. Dobritsa et al. (2001) used relatively short, partial sequences of the *shc* gene for their phylogenetic reconstruction, yet the topology shared with our tree suggests the phylogenetic signal is spread throughout the *shc* gene, and all regions of the gene may be suitable for phylogenetic reconstruction if alignable between taxa. The *R. xylanophilus shc* sequence maps to a position intermediate between the *Frankia-Acidothermus-Streptomyces* cluster and the *Alicyclobacillus* sequences, although in the same larger sub-group as *Frankia-Acidothermus-Streptomyces*. The branch uniting *Frankia* and *Acidothermus* is strongly supported (Fig. 3). Using sequences from the 16S rRNA, Normand et al. (1996)

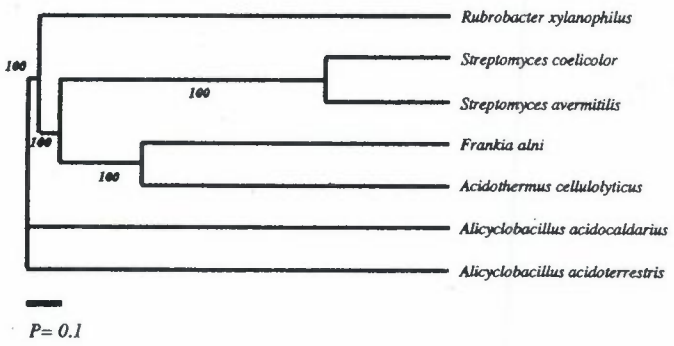

Figure 3. A Neighbor-joining phenogram based on 542 aligned characters representing five conserved regions located between ACN14a positions 12694–13609. Both *Streptomyces* sequences are complete in the alignment. The position of *Rubrobacter xylanophilus* is based on 75 aligned base pairs in the fifth conserved region. The position of *Alicyclobacillus* sp. is based on 191 aligned base pairs present in the third and fourth conserved regions located between ACN14a positions 13014 and 13449. The position of *A. cellulolyticus* is based on 415 aligned base pairs present in the first, second, and third conserved regions located between ACN14a positions 12694 and 13260. Bootstrap values are below the branches.

reported that the closest relative of the genus *Frankia* was not *Geodermatophilus* as previously thought (Hahn et al., 1989), but *A. cellulolyticus*. Maréchal et al. (2000) confirmed the sister relationship between *Frankia* and *Acidothermus* using sequences of the *recA* gene. The grouping of *Acidothermus* and *Frankia shc* sequences in our tree topology is concurrent with these previous studies, and being based on a gene present only in a small proportion of analyzed genomes, provides an independent support for the close relationship between the two bacteria.

Different modes of generating diversity in genomes have been noted including lateral transfers (Daubin et al., 2003) and gene duplication as the most powerful mechanisms (Cannon et al., 2004; Jordan et al., 2001; Kent et al., 2003). Transfers and gene duplication are typically followed by paralogous evolution (Carginale et al., 2004) to adapt to various niches. The analysis of the genomes of the two taxa *Frankia* and *Acidothermus* with particular emphasis on the hopanoid pathway, together with a physiological study of the hopanoid metabolites will one day permit us to reconstitute their evolution at the molecular level and understand how two such closely related bacteria came to adapt to such different ecological niches.

Acknowledgments

This research was supported by joint award NSF 0233366(AB)/CNRS (PN), by California Experiment Station Project 6258 (AB), and by a CNRS/Ministère Recherche Genome ACI program (PN).

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

An, C., Riggsby, W., and Mullin, B. 1985. Relationships of *Frankia* isolates based on deoxyribonucleic acid homology studies. *International Journal of Systematic Bacteriology* **35**: 140–146.

Bauer, C.E., Elsen, S., and Bird, T.H. 1999. Mechanisms for redox control of gene expression. *Annual Review of Microbiology* **53**: 495–523.

Benson, D.R. and Silvester, W.B. 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiological Reviews* **57**: 293–319.

Berry, A., Harriott, O., Moreau, R., Osman, S., Benson, D., and Jones, A. 1993. Hopanoid lipids compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proceedings of the National Academy of Sciences, USA* **90**: 6091–6094.

Berry, A., Moreau, R., and Jones, A. 1991. Bacteriohopanetetrol: abundant lipid in *Frankia* cells and in nitrogen-fixing nodule tissue. *Plant Physiology* **95**: 111–115.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D., and May, G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology* **4**: 10.

Carginale, V., Trinchella, F., Capasso, C., Scudiero, R., Riggio, M., and Parisi, E. 2004. Adaptive evolution and functional divergence of pepsin gene family. *Gene* **333**: 81–90.

Cournoyer, B. and Lavire, C. 1999. Analysis of *Frankia* evolutionary radiation using glnII sequences. *FEMS Microbiology Letters* **177**: 29–34.

Daubin, V., Lerat, E., and Perriere, G. 2003. The source of laterally transferred genes in bacterial genomes. *Genome Biology* **4**: R57.

Ditta, G., Schmidhauser, T., Yakobson, E., Lu, P., Liang, X.W., Finlay, D.R., Guiney, D., and Helinski, D.R. 1985. Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. *Plasmid* **13**: 149–153.

Dobritsa, S.V., Potter, D., Gookin, T.E., and Berry, A.M. 2001. Hopanoid lipids in *Frankia*: identification of squalene-hopene cyclase gene sequences. *Canadian Journal of Microbiology* **47**: 535–540.

Doyle, J.J. and Doyle, J.L. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochemical Bulletin* **19**: 11–15.

Eisenreich, W., Rohdich, F., and Bacher, A. 2001. Deoxyxylulose phosphate pathway to terpenoids. *Trends in Plant Sciences* **6**: 78–84.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Ferreira, A.C., Nobre, M.F., Moore, E., Rainey, F.A., Battista, J.R., and da Costa, M.S. 1999. Characterization and radiation resistance of new isolates of *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus*. *Extremophiles* **3**: 235–8.

Full, C. and Poralla, K. 2000. Conserved tyr residues determine functions of *Alicyclobacillus acidocaldarius* squalene-hopene cyclase. *FEMS Microbiology Letters* **183**: 221–4.

Hahn, D., Lechevalier, M., Fischer, A., and Stackebrandt, E. 1989. Evidence for a close phylogenetic relationship between members of the genera *Frankia*, *Geodermatophilus*, and "*Blastococcus*" and emendation of the family Frankiaceae. *Systematic Applied Microbiology* **11**: 236–242.

Hammad, Y., Maréchal, J., Cournoyer, B., Normand, P., and Domenach, A.M. 2001. Modification of the protein expression pattern induced in the nitrogen-fixing actinomycete *Frankia* sp. strain ACN14a-tsr by root exudates of its symbiotic host *Alnus glutinosa* and cloning of the *sodF* gene. *Canadian Journal of Microbiology* **47**: 541–547.

Hammad, Y., Nalin, R., Maréchal, J., Fiasson, K., Pepin, R., Berry, A.M., Normand, P., and Domenach, A.-M. 2003. A possible role for phenylacetic acid (PAA) in *Alnus glutinosa* nodulation by *Frankia*. *Plant and Soil* **254**: 193–205.

Ishikawa, J. and Hotta, K. 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiology Letters* **174**: 251–253.

Jordan, I.K., Makarova, K.S., Spouge, J.L., Wolf, Y.I., and Koonin, E.V. 2001. Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Research* **11**: 555–565.

Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences, USA* **100**: 11484–11489.

Maréchal, J., Clement, B., Nalin, R., Gandon, C., Orso, S., Cvejic, J.H., Bruneteau, M., Berry, A., and Normand, P. 2000. A *recA* gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. *International Journal of Systematic and Evolutionary Microbiology* **50**: 781–785.

Methe, B.A., Nelson, K.E., Eisen, J.A., Paulsen, I.T., Nelson, W., Heidelberg, J.F., Wu, D., Wu, M., Ward, N., Beanan, M.J., Dodson, R.J., Madupu, R., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Durkin, A.S., Gwinn, M., Kolonay, J.F., Sullivan, S.A., Haft, D.H., Selengut, J., Davidsen, T.M., Zafar, N., White, O., Tran, B., Romero, C., Forberger, H.A., Weidman, J., Khouri, H., Feldblyum, T.V., Utterback, T.R., Van Aken, S.E., Lovley, D.R., and Fraser, C.M. 2003. Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. *Science* **302**: 1967–1969.

Mohagheghi, A., Grohmann, K., Himmel, M., Leighton, L., and Updegraff, D. 1986. Isolation and characterization of *Acidothermus cellulolyticus* gen. nov., sp. nov., a new genus of thermophilic, acidophilic, cellulolytic bacteria. *International Journal of Systematic Bacteriology* **36**: 435–443.

Normand, P. and Lalonde, M. 1982. Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. *Canadian Journal of Microbiology* **28**: 1133–1142.

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., Evtushenko, L., and Misra, A.K. 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. *International Journal of Systematic Bacteriology* **46**: 1–9.

Ourisson, G. and Rohmer, M. 1992. Hopanoids. 2. Biohopanoids: a novel class of bacterial lipids. *Accounts of Chemical Research* **25**: 403–408.

Parsons, R., Silvester, W.B., Harris, S., Gruijters, W.T., and Bullivant, S. 1987. *Frankia* vesicles provide inducible and absolute oxygen protection for nitrogenase. *Plant Physiology* **83**: 728–731.

Perzl, M., Reipen, I.G., Schmitz, S., Poralla, K., Sahm, H., Sprenger, G.A., and Kannenberg, E.L. 1998. Cloning of conserved genes from *Zymomonas mobilis* and *Bradyrhizobium japonicum* that function in the biosynthesis of hopanoid lipids. *Biochimica Biophysica Acta* **1393**: 108–118.

Poralla, K., Muth, G., and Hartner, T. 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiology Letters* **189**: 93–95.

Reipen, I.G., Poralla, K., Sahm, H., and Sprenger, G.A. 1995. *Zymomonas mobilis* squalene-hopene cyclase gene (*shc*): cloning, DNA sequence analysis, and expression in *Escherichia coli*. *Microbiology* **141**: 155–61.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring, N.Y., Cold Spring Harbor Laboratory.

Swofford, D.L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Associates., S., Sunderland, Massachusetts.

Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P., and Benziman, M. 1998. Three cdg operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. *Journal of Bacteriology* **180**: 4416–4425.

Taylor, B.L. and Zhulin, I.B. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology Molecular Biology Reviews* **63**: 479–506.

Wang, J.C., Sakakibara, M., Liu, J.Q., Dairi, T., and Itoh, N. 1999. Cloning, sequence analysis, and expression in *Escherichia coli* of the gene encoding phenylacetaldehyde reductase from styrene-assimilating *Corynebacterium* sp. strain ST-10. *Applied Microbiology Biotechnology* **52**: 386–392.