# **Supporting Information**

### Welander et al. 10.1073/pnas.0912949107



**Fig. S1.** Chromatograms of total lipid extract separated by GC-MS from the following strains: *a*, *Rhodopseudomonas palustris* TIE-1. *b*, *R. palustris* ΔhpnP. *c*, *R. palustris* ΔhpnP complemented with the ΔhpnP gene on a self-replicating plasmid. Numbered compounds: I, 2- methylhopenes; II, hopenes; III, 2-methyldiplopterol; IV, diplopterol; IV, 2-methyltetrahymanol; VI, tetrahymanol; VI, 2-methylbacteriohopanetetrol; and VIII, bacteriohopanetetrol.

Thermosyn Nostoc_pun	MULRRSSDEIKLVPEMTSGRVPYTPRNHRRILCIFPRY-S MTVNLKSLENPTYIAADSKTASNTKRTSVVPSNHRRILCIFPKY-	RSFGTFHHAYPLMGT RSFGTFHHAYPLMGN	/ <mark>RAFMPPOGILLVAAYL</mark> PE / <mark>RAFMPPOGILIVAAYL</mark> PO	SWEVRLIDENVRPATDADYA KWEVRFIDENVKSATRADYO
Cand_Korib		EKSFGTFDNAYPLLG-V	VKAFMPPOGILVVAAYLPA VKAFMPPOGMLVIAAYL <mark>PK</mark>	RWQVRFVDENVKPARSEDIR RWQVRFVDENVKPASDADYR
Nitrobact Rhodopseud	MRGVG0YGIRRVLCVFPRY-I MKAESG0TSRRILCVFPRY-I	SSFGTFEHAYPLTGG SFGTFOHSYPLMDD	/QAFMPPOGLLLIAAYL <mark>PA</mark> /AAFMPPOGLLVIAAYLPD	DWQVRFIDENIRFATNEDFE EWSVRFVDENIRAATADDFA
Oligotroph	MDKOAELSPRRILCVFPKY-S	PSFGTFEYAYPLTDG	KAFMPPOGLLLIAAALPP	SWSVRFIDENIRSATDADFA
Methylobac	MRSTVRPLADFAPAAAGAERKILCVFPHY-V	PSFGTFSHAYPLMGG	/KAFMPPOGLLVIAAALPO /KAFMPPOGLLLIAAYLPE	TWOVEFIDENIGRATPADLA
Stappia_ag	MSECNVLLVYPRFNA	AGSFWNYGATCEMVG-F	KYPAPPLGLITVAAMLPD	TWNVRLIDRNTGELTNDDLD
Acaryochlo	MRVLLIYPLF-I	KSFWSFEKTLSLVD-F	KALLPPLGLVTVAAILPQ	EWEFRLIDRNVEAVPESDWD
RELAXED STRINGENT	* *************************************	* * * * * * * * * * * * * * * * * * *	****************	* * * * * * * * * * * * * * * * * * *
10	01			
	WADIVITSGMHIORPOILEINEIAHRHGKLTALGGPSVSSCPE	YPDVDILHLGELGDA	DEMIAYFDEYGSORPPOO	LIFATT-ARLPLSEFPVPAY
	WADAVIVSGMHIORPOIRKINELAHRWGKITALGGPSVSGCPEY	YYPDFDLIHIGELGDA	DRMIEYID-MHTERPASO	MRFETA-ERLPLORFPIPAY
	WADVVMTSGMHIORNOINAINRRAHSFGKVTVIGGPSVSGCPE	YPEFDYLHVGELGDA' Y <mark>PS</mark> FDYLHVGELGDA'	DOLIERLA-RDPSRPDRO	VRFETK-ERLPLADFPTPAY VVLKTT-DRLAMDOFPVPAY
	WADAVFVSGMHIOROOMNDICRRAHDFDLPVALGGPSVSACPD	YPNFDYLHVGELGDA	DOLIAKLT-HOVTRPKRO	VVFTTE-DRLDMTLFPIPAY
	WADAVFVSGMHIORROMEDICARAHEOGKPVALGGPSVSAAPEC	YPAFDYLHVGELGDA	EALIELIA-RDVSRPPRO	IILKTK-DRRPLAEFPAPAY
	WADAVFVSGMHVQAAQIRDIAERAHAAGRVAVLGGPSVSGSPEC WADMVMTGGMLPQQADCLAVIRRSQAKNVPVVVGGPDATSSPEV	)YPGFDYLHIGEIGDA /YEAADFLVLGEAEGII	DELIRRLD-ADAGRPPAO TTFIDAWKEGVRN	VRLETK-ERLGLADFPAPAY GRFSAEKFKADVASSPVPRF
	WADLVILSAMIVOKEDLLAOIOEAKRRGKLVAVGGPYPTALPNEVI WARMULLSCMIVOKEDMINIVGESHPRCKKVAVGGPEATSVEEDIG	DVGADYLILDEGEITI	PLFIDAIGRGESS	GIFRSGGEKPDVTNTPIPRF
	*****	****	******	*** ********
	***************	**********	* * * * * * * *	*********
20	01 OHVRMENYFLGSIOFSSGCPFRCEFCDIPELYGRNPRLKTPOOIL	ELDTML-ASGNPGAVY	VFVDDNFIGNRRAVMELLP	HLIDWOKANGYPLOFACEAT
	HLLDIDDYFLANVQFSSGCPYRCEFCDIPELYGNSPRMKTPEQVV/	AELDAMR-OSGNMGAV	FVDDNFVGDRRAAMKLLP	HLIDWOKRNGYPIOFACEAT
	DMAAMDKYFLANVQFSSGCPYRCEFCDIPELYGRNPRLKTPQQIL	ELDAIN-AGGAIGAV	FVDDNF IGNRRAVADLLP FVDD <mark>NFV</mark> GNRRAALELLP	HLVEWOKKNNYPVEFACEAT
	ELADIKRYFLGSIQYSSGCPYQCEFCDIPGLYGRNPRLKTPEQITF ELAECSKYLLGSIQYSSGCPYQCEFCDIPGLYGRNPRLKTPEQIT	RELDKLL-ECGISGSVY TELDRMI-ECGIRGSVY	FVDDNFIGNRKAALELLP FVDDNFIGNRKAALDLLP	HLIEWOKKTGYVLRFSCEAT HLVEWOKRTGFOLOLACEAT
	ELIPLRKYFLGSIQFSSGCPYQCEFCDIPGLYGRNPRLKSPQQVLA	AELDKMV-ANGLSGAVY	VFVDDNFIGNRRAVLELLP	HLVEWOKRNGYAVOFACEAT
	DOVPLKRYLIGSLOFSSCOPYTCEFCDIPGLYGRVPRLKTPAOVVC DOVPLKRYLIGSLOFSSCOPYRCEFCDIPALYGROPRLKTPEOLLS	ELDAIVAOPGHPAVVY	FVDDNLVGNRRALLDLLP FVDDNFIGNRRATREMLP	HLVEWOKRNGYPLOFACEAT
	DLLEFDAYAEMSVOFSRGCPFTCEFCDIIELYGRVPRTKTNEOMLA DLLEFDAYAEMSVOFSRGCPFOCEFCDIIVLYGRKPRTKTPAOLLA	AELDRLY-ELGYRGHVI	OFVDDNLIGNKKAVKGFLP MVDDNFIGNKRNVKLFLK	ELAKWOKERNYPFEMSTEAS
	DLLKFPAYDNMSVQFSRGCPFQCEFCDIIVLYGRKPRTKTPQQLI7	AELDRLC-ELGWRG <mark>SV</mark> F	MVDDNFIGNKRNVKLLLA	EL <mark>REWIAEKNYPFTFATEAS</mark>
	***************************************	******	******************	*****************
3 (	01			
	LNIAOSPKLLEMMREAYFCTIFCGIETPEPEALHAIHKDONLSMP LNLAOSPKLLEMMREAYFCTIFCGIETPEPDALHAISKDONLSMP	LLOAVOTLNSYGMEVVS LLKAIOVLNSYGMEVVS	GIIIGFDTDTPQTGDRIL GIIIGLDTDTPETADRII	EFIRASHIPTLTINLLHALP EFIRASOIPMLTINLLHALP
	LNIAOSPKILEMMREAYFCTVFCGIETPEPDALKSISKTHNLSMP	LEAIGTLNSYGMEVVS	GIIIGLDTDTPETGERIL	EFIRLSOLPMLTINLLHALP
	LNIVKSPDLLALMREADFWIIFVGIETPDPAALALISKDHNNIVP LNIAKRPEILSOMRDAFFATIFCGIETPDPVALKAMHKDHNMVP	IMDAI <mark>RTLNDYGMEVV</mark> S	GIILGLDSDIVEIGERIL GIILGLDTD <mark>KPET</mark> GQHLL	EFIDHSQIPLLTINLLEALP
	LNIAKRPEILELMREAYFCTIFVGIETPDPTALKAMHKDHNMMVP LNIAKRPEILOLMHDAFFLTVFCGIETPDPKALOAMSKOHNMMVP	LEGVRTISSYGIEVVS	GIILGLDTDTPETGEFLM	OFIEOSOIPLLTINLLOALP EFVDOSOIPLLTINLLOALP
	LNIARSPOILKLMREAAFDTIFCGIETPEPEALOAIOKSHNMVRPI	LESVKAINDHGMEVV	GIILGLDTDTAESGDRIL	EFLEOSNIPLATINLLOALP
	LNMAROPEILELMROANFITVFVGIETPELDALASIDROHNAAVP LNLADDEALLKMMOOAGYFSVFIGIESPDPEVLVATRKKONTRRDI	IA <mark>KSVH</mark> KIYDAGIFVLA	GFIVGFDEESDRVADEIS	ALIEEAAIPVAMTGLLYALP
	VDLAQDQELMDAMVRCNFGAVFLGIETPDEESLTFTQKFQNTRDSI VDLAQDEELMDLMIECNFNSVFLGIETPDEDSLKVTKKFQNTRDSI	SESTDTMSEKGLETM	GFIIGFDGEKSGAGARIV	KFVEOTAIPTALFSMLOALP REVEKTNIPIALFSMLOALP
	***************************************	* * * * * * * * * * * * * * * * * *	*****	****
	^^^	******	*****	*****
40	TPLWRRLEAEGRINHDENRESNVEFIMPYEE	VEMWRRTITTAYEPE	FLY <mark>OR</mark> FAYOMEHTYPNRIA	VPNSPARLSRENILRGLRIM
	RTPLWRRLEAEGRLVFDESRESNVEFLMPYEON RTPLWDRLOOEGRLVSEVEGRESNVEFLLPHDON	VEMWRRCITTAYEPE	LYERFAYNMEHTYPNRIE LYERFAYNCEHTYPNRIA	VPNSPARTSGANIRKGLTYL VPNSPERTSWANIOKGLTIL
	RTPLYRRLOAAGRLNDAPGRESNVEFVLPYDO	VAMWRKTFONAFTPE	AIYERFSY00 <mark>KNTYPHR</mark> V0	PPPS-GKANLRNMLRGFTTL
	KTPLWDRLCREGRLVHDEGRDSNIDFRLPYEH KTPLWDRLOREGRLVHPHDON	VATWRNCMGRAYEPE	LLERFEYOIRETYPKRI-	HPSTPORASKANIKLGLIML
	KTPLWDRLARAGRLIE	VAMWRDCMGRAYTPE	MFARFOHOIEHTYPNRIH	PPPTKORLTAANI KRGLTMM RPWSRORLSPRNI RMAVIIL
	RTPLWDRLAREGRLVDDPRLESNVRFLRPHDEV	/VASWRRAIAHAYAPE	LFDRFRHOIDHTYSRRIE	TPAK-TRLTRANLRRGLVLG
	DTALWHRLEKENRLENKSANINOTTLMNFFPTRPLAE	LEDFKOVIDRVYGPE LASEYVEAFWELYEPSH	FLDR-AYRHYRILGOATY	PKKG-KGAKKPLNWKVLRAL
	NTALWHRLEKEGRLIQ GQEGTGNQTSLMNYVPTRPIDE	LANEYVEAFCNIYDPVE	YLDR- <mark>TFRHFMSMNIRPK</mark>	YGG <mark>S-GKSSVKVSWVDIRAL</mark> ****
	***********	*******	* * * * *	
50		THELOFCARCDEALC		
	TNILLRIGLFS-NYROTFWKMAKPALKAGDIENLIHVGLVGHHI	LINFAQECARGDEAAS	YSOKILTKVRS	
	ANLLIKVGVLS-DYRETFWKMAVPALRSGRIEELIHVGLVGHHI AKLMWTVGVKS-DYRRVFWKMARPALRELDTEDTHVSLVSHHI	LISFAQOCAVGKESASE	YSOKVRTASH	
	RNIFWKVGLRG-DYRRVFWKFALRRLARGEIEYLISSVLVAHHI	LIMFARSASRGKTNAS	YYSLKLRETAG	<mark>ASVP</mark> AE
	ARIIWHAGVRS-DYREFWKFAWPRLKAGDIENLLLVMVVAHHI ARIIWHAGVRS-DYRREFWKFAWPRLKAGDVERFISCSLVAHHI	LITFARGASSGROTASY	YY <mark>SIRLREAAV</mark> YY <mark>SDKPAE</mark> ALP	<mark>-</mark> PAE
	AKLFWHVGLRG-GYRRAFWTFALGRLIRGKIEPVISVGLVAHHI	JIMFARDACSGRSNASH	IYSAKNRPOGA	<mark>ATPLROEIAAE</mark> -
	GRLILAMTFKETTYRGPFWKTLAKTAFTNFRAIKPVLMMLALYVHI	GPFSRFVIA-QIDKQ	IAEIEKGNWEQ	PALVAAE
	STLAWRQGVVR-KTRWKFWINLFRIIQKNPQVLTRYLSASAFIEH	THEY-ROIVEDOIEVON	LAVHLANHPRVVVK	KLEEEKAQVLV PVEE <mark>SSQPEQIPAKTA</mark> -
	***************************************	**** ***** ****		

**Fig. S2.** Alignment of select HpnP and outgroup sequences showing filtered and unfiltered positions. The twelve representatives are as follows: *Thermosynechococcus elongatus, Nostoc punctiforme* PCC73102, *Gloeobacter violaceus* PCC7421, Candidatus *Koribacter versatilis* Ellin 345, *Nitrobacter* sp. Nb-311a, *R. palustris* TIE-1, *Oligotropha carboxidovorans* OM5, *Methylocella silvestris* BL2, *Methylobacterium nodulans* ORS2060, *Stappia aggregata* IAM12614 (outgroup), *Nostoc azollae* 0708 (outgroup), and *Acaryochloris marina* MBIC11017 (outgroup). This alignment is from trials 7–9 (Table S3) and was generated by Muscle with one outgroup family. The sites included in the filtered alignments under relaxed and stringent scenarios are marked with \*.

SANG SANG



Fig. S3. Major families of putative bacterial radical SAM proteins. Maximum likelihood tree was inferred from an alignment of 247 sequences and 679 positions. The clade labeled "outgroup family" was used in all trials, and the additional outgroup was used in trials 4–6, 10–12, and 16–18 (Table S3).





**DNAS** 

DNAS AND

Bacterial genome	2-Me C30*	2-Me C35 <sup>†</sup>	Ref.	Isolation habitat
α-Proteobacteria				
R. palustris TIE-1	+	+	(1)	Iron mat
R. palustris CGA009	+	+	(1)	Freshwater sediment
R. palustris BisB18	+	+	This study	Freshwater sediment
R. palustris BisA53	+	+	This study	Freshwater sediment
R. palustris BisB5	+	+	This study	Freshwater sediment
R. palustris HaA2	+	+	This study	Freshwater sediment
Bradyrhizobium sp. ORS278	nt‡	nt		Stem nodules
Bradyrhizobium sp. BTAi1	+	-	(2)	Stem nodules
Bradyrhizobium japonicum USDA110	+	-	(2)	Root nodules
Nitrobacter hamburgensis X14	nt	nt		Soil
Nitrobacter winogradskyi Nb-255	nt	nt		Soil
Nitrobacter sp. Nb-311A	nt	nt		Coastal marine waters
O. carboxidovorans OM5	nt	nt		Wastewater
Beijerinckia indica ATCC 9039	+	-	(3)	Acidic soil
M. silvestris BL2	nt	nt		Acidic forest soil
Methylobacterium extorquens AM1 <sup>§</sup>	nt	nt		Airborne contaminant
M. extorquens PA1 <sup>§</sup>	nt	nt		
M. extorquens DM4 <sup>§</sup>	nt	nt		Industrial wastewater
Methylobacterium chloromethanicum CM4 <sup>§</sup>	nt	nt		Contaminated soil
M. nodulans ORS 2060 <sup>§</sup>	nt	nt		Root nodules
Methylobacterium populi BJ001 <sup>§</sup>	nt	nt		Poplar tree
Methylobacterium sp. 4–46 <sup>§</sup>	nt	nt		Root nodules
Methylobacterium radiotolerans JCM 2831 <sup>§</sup>	nt	nt		
Cyanobacteria				
G. violaceus PCC 7421	+	+	(4)	Calcareous rocks
T. elongatus BP-1	nt	nt		Hot spring
Cyanothece sp. 7424 <sup>§</sup>	nt	nt		Rice field
Cyanothece sp. 7425 <sup>§</sup>	nt	nt		Rice field
Cyanothece sp. PCC7822 <sup>§</sup>	nt	nt		Rice field
Nostoc puncitforme PCC73102	+	+	(4)	Gymnosperm symbiont
Acidobacteria				
Candidatus Koribacter versatilis Ellin345	nt	nt		Soil

#### Table S1. Presence of 2-methylhopanoids in bacteria with an HpnP homolog

\*2-MeC30 indicates 2-methylhopenes.

<sup>+</sup>2-MeC35 indicates 2-methylbacteriohopanetetrol.

<sup>\*</sup>Indicates not tested.

PNAS PNAS

<sup>§</sup>2-methylhopanoid production has been shown in other Methylobacterium and Cyanothece strains (4, 5).

1. Rashby SE, et al. (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. Proc Natl Acad Sci USA 104:15099–15104.

2. Bravo JM, et al. (2001) Novel methylated triterpenoids of the gammacerane series from the nitrogen-fixing bacterium Bradyrhizobium japonicum. Eur J Biochem 268:1323–1331.

3. Vilcheze C, et al. (1994) Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria Azotobacter vinelandii, Beijerinckia indica and Beijerinckia mobilis. Microbiology 140:2749–2753.

4. Talbot HM, et al. (2008) Cyanobacterial bacteriohopanepolyol signatures from cultures and natural environmental settings. Org Geochem 39:232–263.

5. Knani M, Corpe WA, Rohmer M (1994) Bacterial hopanoids from pink pigmented facultative methylotrophs (PPFMs) and from green plant surfaces. *Microbiology* 140:2755–2759.

Bacterial genome	2-methyls?	HpnP ortholog?	Ref.
α-Proteobacteria			
R. palustris TIE-1	Yes	Yes	(1)
R. palustris CGA009	Yes	Yes	(1)
R. palustris BisB18	Yes	Yes	This study
R. palustris BisA53	Yes	Yes	This study
R. palustris BisB5	Yes	Yes	This study
R. palustris HaA2	Yes	Yes	This study
Bradyrhizobium sp. BTAi1	Yes	Yes	(2)
Bradyrhizobium japonicum USDA110	Yes	Yes	(2)
Beijerinckia indica ATCC 9039	Yes	Yes	(3)
Zymomonas mobilis	No	No	(4)
Rhodospirillum rubrum	No	No	(5)
Gluconobacter oxydnas	No	No	(5)
Acetobacter pasteurianus	No	No	(6)
δ-Proteobacteria			
Geobacter metallireducens	No	No	(7)
Geobacter sulfurreducens	No	No	(7)
γ-Proteobacteria			
Methylococcus capsulatus	No	No	(5)
Azotobacter vinelandii	No	No	(5)
β-Proteobacteria			
Nitrosomonas europa	No	No	(5)
Burkholderia pseudomallei	No	No	(8)
Cyanobacteria			
G. violaceus PCC 7421	Yes	Yes	(9)
N. puncitforme PCC73102	Yes	Yes	(9)
Trichodesmium erythraeum	No	No	(9)
Anabaena variabilis	No	No	(9)
Microcystis aeruginosa	No	No	(9)
Crocosphaera watsonii	No	No	(9)
Synechocystis sp. PCC 6803	No	No	(9)
Actinobacteria			
Streptomyces coelicolor	No	No	(10)
Frankia alni	No	No	(11)
Firmicutes			(
Alicyclobacillus acidocaldarius	No	No	(12)

Table S2. Correlation of 2-methylhopanoid production with the presence of HpnP in bacteria tested for hopanoid production

 Rashby SE, et al. (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. Proc Natl Acad Sci USA 104:15099–15104.

 Bravo JM, et al. (2001) Novel methylated triterpenoids of the gammacerane series from the nitrogenfixing bacterium Bradyrhizobium japonicum. Eur J Biochem 268:1323–1331.

 Vilcheze C, et al. (1994) Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria Azotobacter vinelandii, Beijerinckia indica and Beijerinckia mobilis. Microbiology 140:2749–2753.

 Moreau RA, et al. (1995) Analysis of intact hopanoids and other lipids from the bacterium Zymomonas mobilis by high-performance liquid chromatography. Anal Biochem224:293–301.

 Rohmer M, et al. (1984) Distribution of hopanoid triterpenes in prokaryotes. J Gen Microbiol 130:1137–1150.

 Zundel M, et al. (1985) Prokaryotic triterpenoids. 1. 3 beta-Methylhopanoids from Acetobacter species and Methylococcus capsulatus. Eur J Biochem 150:23–27.

7. Hartner T, et al. (2005) Occurrence of hopanoid lipids in anaerobic Geobacter species. FEMS Microbiol Lett 243:59–64.

 Cvejic JH, et al. (2000) Bacterial triterpenoids of the hopane series as biomarkers for the chemotaxonomy of *Burkholderia, Pseudomonas* and *Ralstonia* spp. *FEMS Microbiol Lett* 183:295–299.

9. Talbot HM, et al, (2008) Cyanobacterial bacteriohopanepolyol signatures from cultures and natural environmental settings. Org Geochem 39:232–263.

10. Poralla K, et al. (2000) Hopanoids are formed during transition from substrate to aerial hyphae in Streptomyces coelicolor A3(2). FEMS Microbiol Lett 189:93–95.

 Nalin R, et al. (2000) High hopanoid/total lipids ratio in Frankia mycelia is not related to the nitrogen status. Microbiology 146:3013–3019.

 Poralla K, et al. (1980) A glycolipid containing hopane isolated from the acidophilic, thermophilic Bacillus acidocaldarius, has a cholesterol like function in membranes. FEBS Lett 113:107–110.

SANG SANG

Table S3. Sensitivity o	of rooted HpnP	topology to	alignment	parameters
-------------------------	----------------	-------------	-----------	------------

Trial	Total taxa incl. outgroup(s)	Alignment program	Filtering by Gblocks	No. of sites	-InL	Best topology*	
1	83	ClustalW	None	622	35,976.33	С	
2	83	ClustalW	Stringent	399	21,615.26	С	
3	83	ClustalW	Relaxed	469	27,187.91	А	
4	118	ClustalW	None	678	57,342.19	А	
5	118	ClustalW	Stringent	280	22,884.90	А	
6	118	ClustalW	Relaxed	422	38,833.93	А	
7	83	Muscle	None	626	36,221.90	Α	
8	83	Muscle	Stringent	392	20,958.59	Α	
9	83	Muscle	Relaxed	468	26,874.87	В	
10	118	Muscle	None	704	58,004.70	В	
11	118	Muscle	Stringent	306	25,363.81	А	
12	118	Muscle	Relaxed	411	36,969.22	А	
13	83	T-Coffee	None	689	35,924.79	А	
14	83	T-Coffee	Stringent	382	20,832.01	А	
15	83	T-Coffee	Relaxed	452	25,723.67	А	
16	118	T-Coffee	None	812	57,199.16	В	
17	118	T-Coffee	Stringent	312	25,942.82	А	
18	118	T-Coffee	Relaxed	394	34,708.95	А	

\*A indicates a topology similar to Fig. 4A, with early branching cyanobacteria. **B**, topology similar to Fig. 4B, with monophyletic cyanobacteria sister to the  $\alpha$ -proteobacteria. **C**, topology similar to Fig. 4C, with cyanobacteria branching within the  $\alpha$ -proteobacteria. We did not consider the position of the lone acidobacterium sequence as it was not well-supported in most trials.

## Table S4. Comparing alternative tree topologies using Shimodaira's Approximately Unbiased test

		-In	p-value for				
						topolog	3y⁺
Alignment	No. of sites	А	В	С	Α	В	С
ClustalWno filtering	622	36,001.91	35,997.57	35,976.33	0.20	0.26	0.81
ClustalW relaxed	469	27,187.91	27,199.24	27,199.77	0.69	0.32	0.38
ClustalW stringent	399	21,621.43	21,634.54	21,615.26	0.46	0.18	0.68
Muscleno filtering	626	36,221.90	36,244.39	36,259.22	0.84	0.24	0.10
Musclerelaxed	468	26,884.17	26,874.87	26,899.22	0.41	0.73	0.16
Musclestringent	392	20,958.59	20,982.64	20,964.40	0.72	0.14	0.42

\*For each alignment, the best tree was found by PhyML and its likelihood is shown in bold. The two alternative topologies were fixed for each alignment to topology A, B, or C from Fig. 4 and then branch lengths and substitution parameters were optimized on that topology, giving the likelihood values shown. These three trees were then compared using the AU test.

<sup>†</sup>p-values <0.05 allow us to reject that tree as the most likely. If no p-value is <0.05, then all three trees belong to the confidence set and the true tree could be any one of them.

#### Table S5. Strains and plasmids used in this study

Strains or plasmids	Genotype, markers, or characteristics			
Strains				
E. coli S17-1	thi pro hdsR hdsM <sup>+</sup> recA; chromosomal insertion of RP4-2 (Tc::Mu Km::Tn7)			
R. palustris TIE-1	Isolated from Woods Hole, MA			
R. palustris ∆hpnP	<i>R. palustris</i> TIE-1 ∆hpnP			
<i>R. palustris</i> $\Delta$ hpnP + pPVW24	<i>R. palustris</i> TIE-1 $\Delta$ hpnP transformed to Gm <sup>r</sup> with pPVW24			
Plasmids				
pJQ200SK	Mobilizable suicide vector; sacB Gm <sup>r</sup>			
pBBR1MCS-5	5.1-kb broad hoast range plasmid; Gm <sup>r</sup> lacZ			
pPVW9	1-kb hpnP upstream region PCR amplified with primers PW39 and PW40			
	and cloned into pJQ200SK at Notl and Xbal			
pPVW10	1-kb hpnP downstream region PCR amplified with primers PW41			
	and PW42 and cloned into pJQ200SK at Xbal and Spel			
pPVW17	Xbal and Spel cut hpnP downstream fragment from			
	pPVW10 subcloned into pPVW9			
pPVW24	3.5-kb hpnP complementation fragment PCR amplified with primers			
	PW85 and PW86 and cloned into pBBR1MCS-5 at HindIII and Spel			

Gm, gentamicin

PNAS PNAS

Analyte	Rt*	M+ (RI)	Base peak	Characteristic fragments (RI)
2-methylhopene	26.45	424(18)	189	205(73), 355(23), 381(25)
hopene	26.59	410(18)	189	191(93), 341(27), 367(23), 410(19)
2-methyldiplopterol acetate	27.68	484	149	189(92), 205(80), 381(5), 424(4)
diplopterol acetate	27.82	470	191	149 (93), 189(95), 367(5), 410(4)
2-methyltetrahymanol acetate	28.49	484	205	189(32), 263(5), 355(4), 424(4)
tetrahymanol acetate	28.64	470(6)	191	189(46), 249(6), 395(5), 410(4)
2-methylBHtetrol acetate	42.84	728(1)	205/95	253(10), 383(19), 433(5), 493(38), 713(1)
BHtetrol acetate	43.81	714(.5)	191	253(9), 369(18), 433(4), 493(34), 699(1)

\*Retention time (min) on 30 m DB-XLB column and using an Agilent 6890GC/5975MSD.

PNAS PNAS