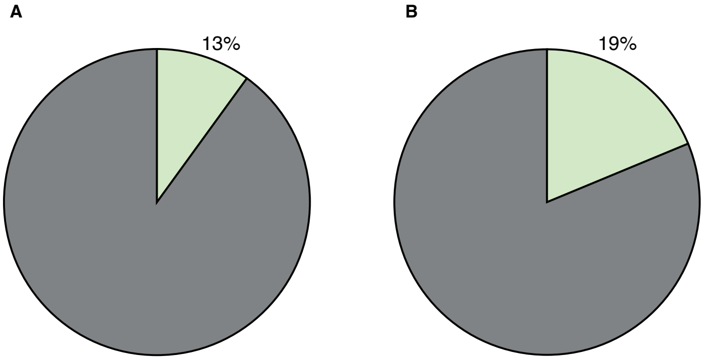

Figure S1. Distribution of *hpnP* in cyanobacteria. Genomes with *hpnP* are represented in green and without *hpnP* in gray. (A) Distribution of *hpnP* among all finished cyanobacterial genomes on IMG. (B) Distribution of *hpnP* among cyanobacterial genera to account for biased sequencing among genomes. If one member of a genus had *hpnP* that genus was counted as having the gene.

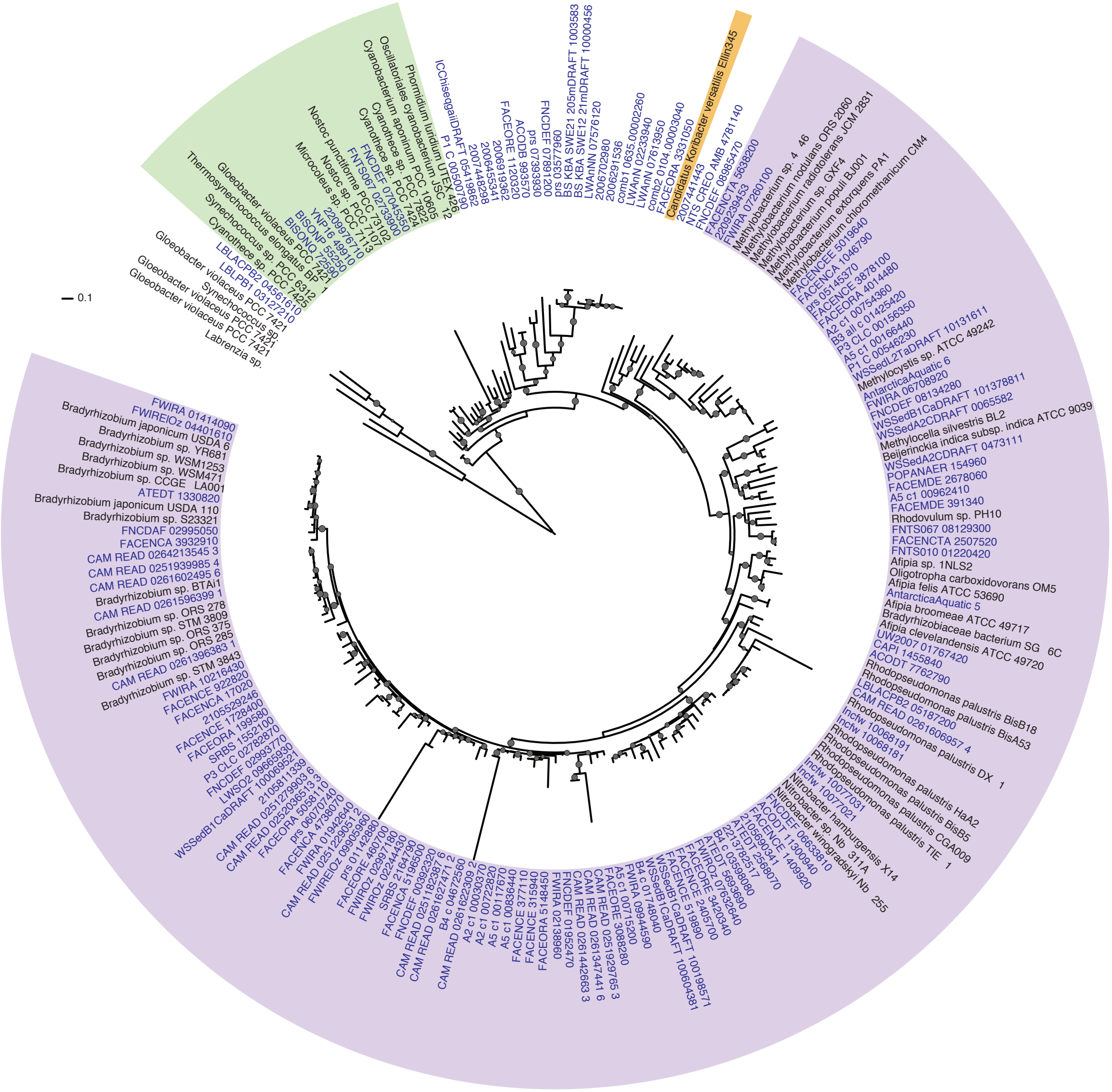

Figure S2. Phylogeny of metagenomic HpnP sequences. Metagenomic HpnP sequences, blue, appear in this maximum likelihood phylogeny with known HpnP sequences from genomes. Phylogeny is the same as Fig. 1. Additional information on metagenomic *hpnP*s can be found in Dataset S1. Clades are colored in purple for alphaproteobacteria, green for cyanobacteria, orange for acidobacteria, and colorless for unknown. Branch support was calculated with aLRT. Gray dots represent aLRT values over 0.8 with larger dot equaling larger aLRT values. The scale bar is a measure of evolutionary distance equaling 0.1 substitutions per site.

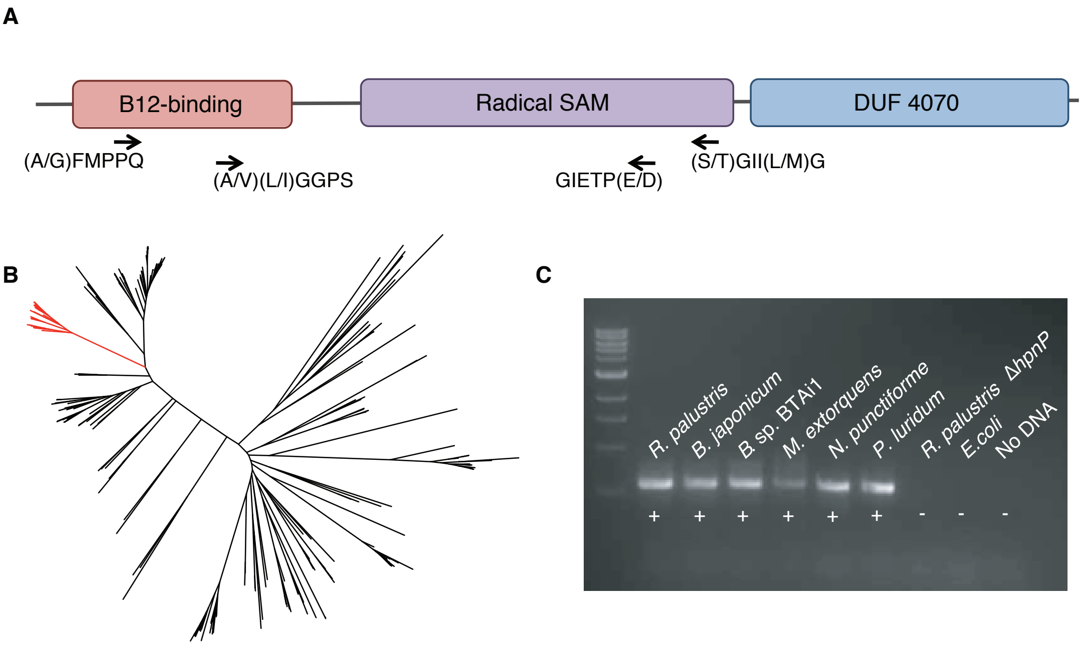

Figure S3. Design of *hpnP* degenerate PCR primers. (A) Protein domain structure of HpnP determined by InterProScan and location of degenerate PCR primers with corresponding protein consensus sequences. HpnP is 527 amino acids in length. The domains B$_{12}$ binding, radical SAM, and DUF4070 are 105, 162, and 140 amino acids in length, respectively, in *R. palustris* TIE-1. (B) Phylogeny of *hpnP*, red, and related homologous proteins. (C) Control PCR reactions for degenerate *hpnP* primers with genomic DNA from diverse cultured 2-MeBHP producers and negative controls.

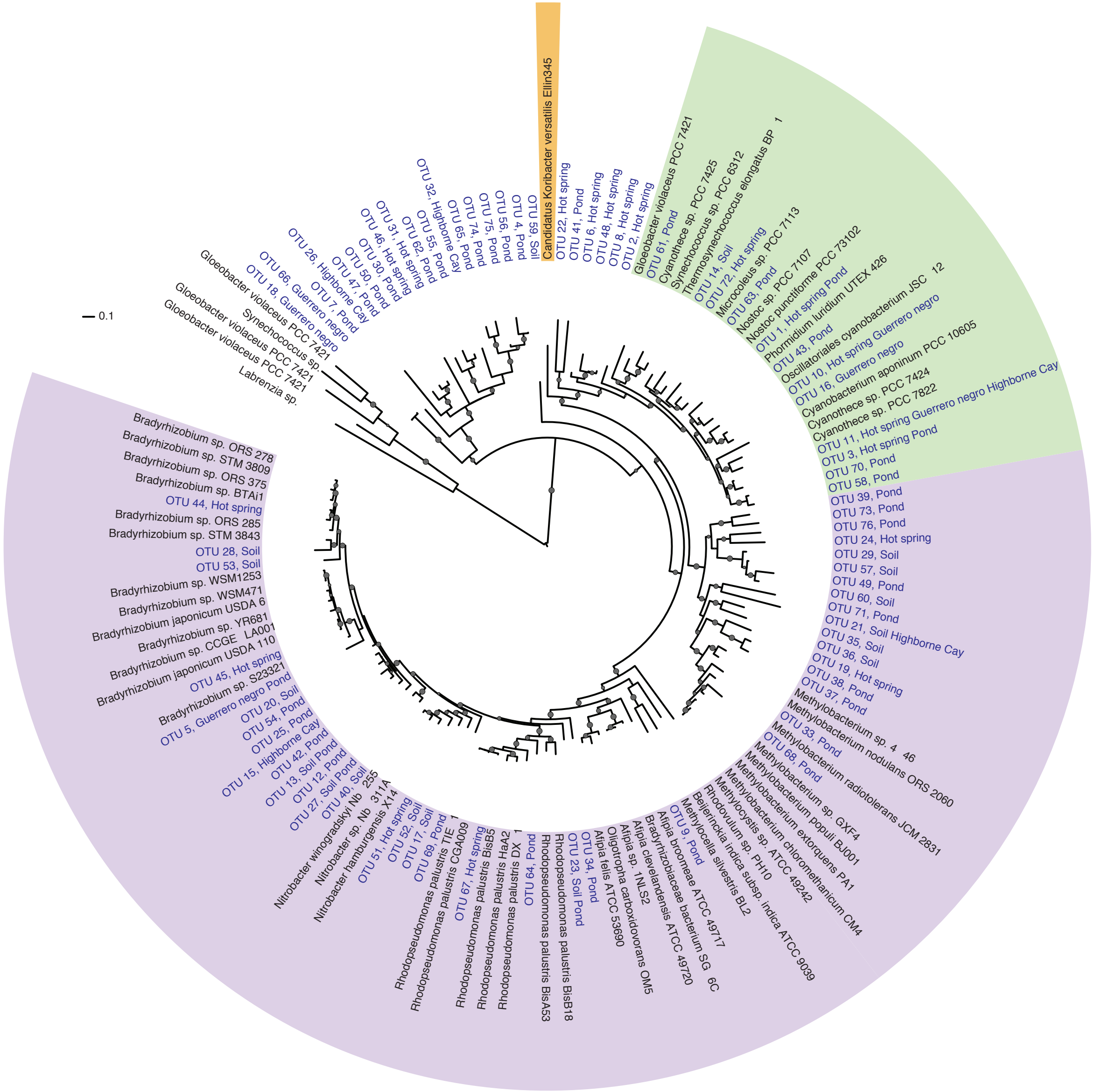

Figure S4. Phylogenetic tree comparing PCR clone library generated HpnP sequences to reference HpnP sequences from genomes. Environmental *hpnP* sequences were clustered at 95% similarity in CD-HIT. The representative sequences recovered were aligned in MAFFT. The phylogeny was generated in PhyML and edited in iTOL. Leaf names reflect all sequences that clustered with the representative sequence in blue. Clades are colored in purple for alphaproteobacteria, green for cyanobacteria, and orange for acidobacteria. Unknown group 1 is the early diverging uncolored region, and unknown group 2 is the uncolored region between the acidobacterium and cyanobacteria. Branch supports were calculated with aLRT. Gray dots represent aLRT values over 0.8 with larger dot equaling larger aLRT values. The scale bar is a measure of evolutionary distance equaling 0.1 substitutions per site.

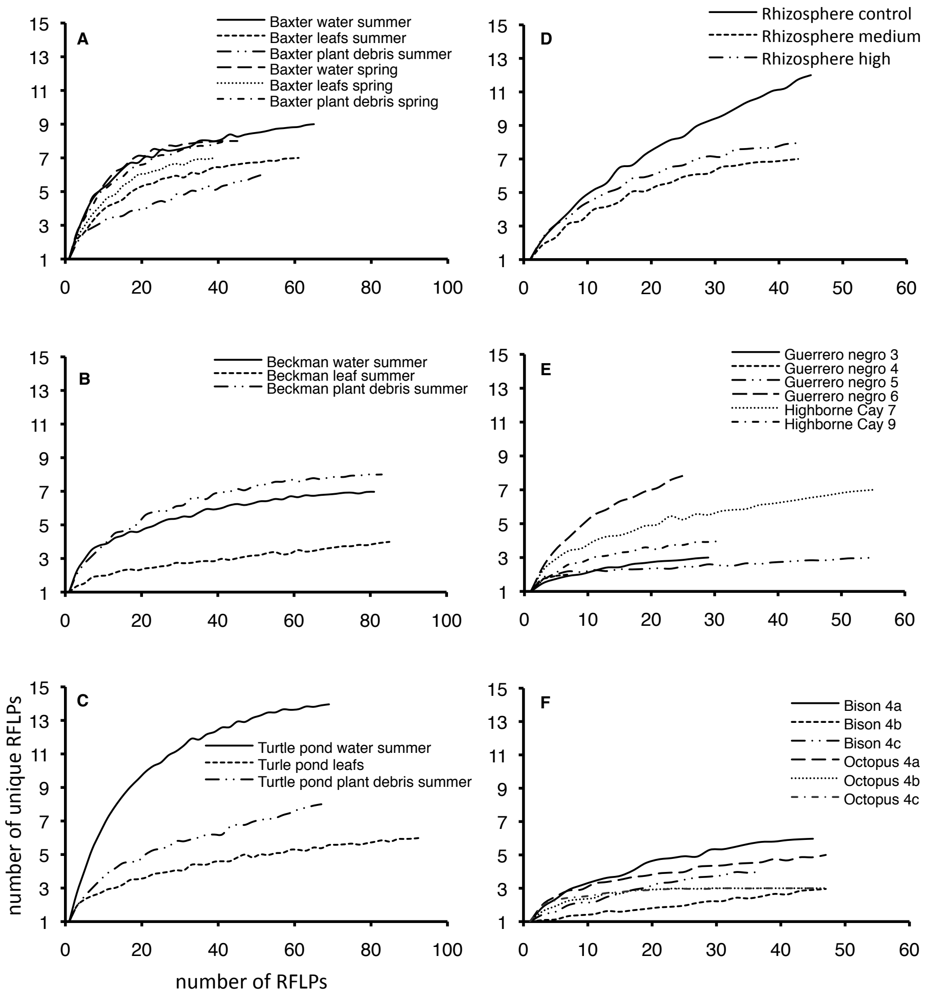

Figure S5. Rarefaction curves for select *hpnP* clone libraries. RFLP patterns confirmed to be *hpnP* by sequencing were used to generate rarefaction curves in QIIME using the observed species metric and rarefied 100 times in increments of 2. Rarefaction curves are grouped by environment and average richnesses (± standard deviation) were calculated when data was rarefied to 25 RFLP patterns **(A)** Baxter pond 6.4 ± 1.7 **(B)** Beckman pond 4.5 ± 1.2 **(C)** Turtle pond 6.6 ± 3.5 **(D)** Rhizosphere 7 ± 1.2 **(E)** Guerrero negro 4.4 ± 3 and Highborne Cay 4.5 ± 1 **(F)** selected hot spring microbial mats 3.4 ± 1.

SI Table 1. Description of environmental samples that appear in Figure 2

Environment	Site/Sample type	Sample Name	Description	Previously Published Name
Hot springs	Carbonate	Narrow Gauge 1	Carb; floating pieces of mat from terrace pool	NG09-1
Hot springs	Carbonate	Narrow Gauge 2	Carb; thin orange mat from terrace edge	NG09-2
Hot springs	Carbonate	Narrow Gauge 3	Carb; thin orange mat from terrace overflow	NG09-3
Hot springs	Carbonate	Narrow Gauge 5	Carb; orange mat from runnoff channel	NG09-5
Hot springs	Carbonate	Narrow Gauge 6	Carb; white carbonate from spring source	NG09-6
Hot springs	Carbonate	Narrow Gauge 7	Carb; thin white streamers in source pool	NG09-7
Hot springs	Carbonate	Narrow Gauge 8	Carb; orange mat from runnoff channel	NG09-8
Hot springs	Pink streamer	Octopus Spring 1	PS; pink streamers	OS09-1
Hot springs	Pink streamer	Brain Pool 1	PS; pink streamers	BP09-1
Hot springs	Pink streamer	Bison Pool 1	PS; pink streamers	B09-1
Hot springs	Pink streamer	Ojo Caliente 1	PS; pink streamers	OC09-1
Hot springs	Pink streamer	Brain Pool 2	PS; pink streamers with green coatings forming in mixing zone between geyser and creek	BP09-2
Hot springs	Pink streamer	Narrow Gauge 4	PS; white opaque streamers from source old Narrow Gauge (carbonate system)	NG09-4
Hot springs	Pink streamer	Spent Kleenex 1	PS; white streamers	SK09-1
Hot springs	Pink streamer	Bison Pool 2	PS; yellow streamers	B09-2
Hot springs	Yellow biofilm	Bison Pool 3	YB; yellow biofilm	B09-3
Hot springs	Yellow biofilm	Octopus Spring 2	YB; yellow biofilm	OS09-2
Hot springs	Orange mat high temperature	Octopus Spring 3	OM-HT; thick green filaments growing off of stratified orange mat	OS09-3
Hot springs	Orange mat high temperature	Imperial Geyser 1	OM-HT; yellow topped stratified mat with orange and green layers below	IG09-1
Hot springs	Orange mat high temperature	Imperial Geyser 2a	OM-HT; <1mm think upper orange and green layer of stratified mat	IG09-2a
Hot springs	Orange mat high temperature	Imperial Geyser 2b	OM-HT; bright salmon color layers in stratified mat	IG09-2b
Hot springs	Orange mat high temperature	Imperial Geyser 2c	OM-HT; thick grey soft mat	IG09-2c
Hot springs	Orange mat high temperature	Imperial Geyser 3	OM-HT; thick green filaments growing off of stratified orange mat	IG09-3

Environment	Site/Sample type	Sample Name	Description	Previously Published Name
Hot springs	Orange mat low temperature	Bison Pool 4a	OM-LT; upper orange layer from conoform orange mat	B09-4a
Hot springs	Orange mat low temperature	Bison Pool 4b	OM-LT; green middle layer from conoform orange mat	B09-4b
Hot springs	Orange mat low temperature	Bison Pool 4c	OM-LT; salmon colored sinter pieces at base of conoform mat	B09-4c
Hot springs	Orange mat low temperature	Octopus Spring 4a	OM-LT; upper orange layer from conoform orange mat	OS09-4a
Hot springs	Orange mat low temperature	Octopus Spring 4b	OM-LT; green middle layer from conoform orange mat	OS09-4b
Hot springs	Orange mat low temperature	Octopus Spring 4c	OM-LT; salmon colored sinter pieces at base of conoform mat	OS09-4c
Hot springs	Black sediment	Norris zygogonium mat 1	BS; zygogonium mat	NR09-1
Hot springs	Black sediment	Norris zygogonium mat 2	BS; elemtal sulfur colored white mat from zygogoinum mat source spring	NR09-2
Hot springs	Black sediment	Norris zygogonium mat 3	BS; bright green biofilm from runoff channel	NR09-3
Hot springs	Black sediment	Washburn 1	BS; black gelatious sediment	WB09-1
Hot springs	Black sediment	Washburn 2	BS; black sediment	WB09-2
Hot springs	Black sediment	Boulder Spring 1	BS; black gelatious sediment	BS09-1
Hot springs	Black sediment	Boulder Spring 2	BS; black gelatious sediment	BS09-2
Terrestrial	Soil	Rhizosphere, control	Rancho Sierra Vista Newbury Park, CA ; Sage brush rhizosphere; 0 g N/m/yr	
Terrestrial	Soil	Rhizosphere, medium nitrogen	Rancho Sierra Vista Newbury Park, CA ; Sage brush rhizosphere; 1.5 g N/m/yr	
Terrestrial	Soil	Rhizosphere, high nitrogen	Rancho Sierra Vista Newbury Park, CA ; Sage brush rhizosphere; 3 g N/m/yr	
Freshwater	Man-made ponds	Baxter water, summer	Whole water	
Freshwater	Man-made ponds	Baxter leafs, summer	Lilypad leaf scrapping	
Freshwater	Man-made ponds	Baxter plant debris, summer	Decaying plant matter from the bottom of the pond	

Environment	Site/Sample type	Sample Name	Description	Previously Published Name
Freshwater	Man-made ponds	Baxter water, spring	Whole water	
Freshwater	Man-made ponds	Baxter leafs, spring	Lilypad leaf scrapping	
Freshwater	Man-made ponds	Baxter plant debris, spring	Decaying plant matter from the bottom of the pond	
Freshwater	Man-made ponds	Beckman water, summer	Whole water	
Freshwater	Man-made ponds	Beckman leafs, summer	Lilypad leaf scrapping	
Freshwater	Man-made ponds	Beckman plant debris, summer	Decaying plant matter from the bottom of the pond	
Freshwater	Man-made ponds	Turtle pond water, summer	Whole water	
Freshwater	Man-made ponds	Turtle pond leafs, summer	Whole leaf	
Freshwater	Man-made ponds	Turtle pond sediment, summer	Pond sediment in contact with plant roots	
Marine	Open water	Whole seawater	SPOT 121; 33°33'N, 118°24'W; Sept. 21, 2012; 2L surface water filter on GF/F Whatman	
Marine	Open water	Plankton tow	SPOT 121; 33°33'N, 118°24'W; Sept. 21, 2012; horizontal surface tow filtered on GF/F Whatman; contained pennate diatoms, e.g. Rhizosolenia	
Marine	Hypersaline lagoon	Laguna Guerrero Negro 1	In situ Pond 4 near 1 (0-5mm); September 16, 2002	
Marine	Hypersaline lagoon	Guerrero negro 2	Flume experiment Pond 4 near 5 85 ppt salinity top 0 mm, orange	
Marine	Hypersaline lagoon	Guerrero negro 3	Flume experiment Pond 4 near 5 85 ppt salinity top 1-1.5 mm	
Marine	Hypersaline lagoon	Guerrero negro 4	Flume experiment Pond 4 near 5 diel 10:45 pm 0-1 mm	
Marine	Hypersaline lagoon	Guerrero negro 5	Flume experiment Pond 4 near 5 diel 2 pm 2-3 mm	
Marine	Hypersaline lagoon	Guerrero negro 6	Flume experiment Pond 4 near 5 diel 2 pm 0-1 mm	
Marine	Modern stromatolites	Highborne Cay 1	Subtidal mats, windward side of Highborne Cay Bahamas (76°49'W, 24°43'N), March 2010	
Marine	Modern stromatolites	Highborne Cay 2	As above	
Marine	Modern stromatolites	Highborne Cay 3	As above	

SI Table 2. Hypergeometric probability of *hpnP* or *shc* correlating with plant-associated organisms or environments.

	hpnP		*shc*	
	genomes	metagenomes	genomes	metagenomes
# genomes or metagenomes	1200	474	1200	474
# plant associated	107	93	107	93
# contain gene	26	59	183	221
# overlap	12	30	44	66
p value	4.6x10^{-7*}	5.7x10^{-9*}	6.6x10^{-12*}	1.2x10^{-7*}

*p values < 0.01 are significant.

SI Table 3. Hypergeometric probability of *hpnP* correlating with *nifD* or *moxF*.

	nifD	*moxF*
# genomes	1200	1200
# with *nifD* or *moxF*	199	124
# with *hpnP*	26	26
# overlap	16	13
p value	2.1x10^{-7*}	2.7x10^{-7*}

*p values < 0.01 are significant.