Supplementary Figures for

An Expanded Genomic Representation of the Phylum Cyanobacteria

Rochelle M. Soo, Connor T. Skennerton, Yuji Sekiguchi, Michael Imelfort, Samuel J. Paech, Paul G. Dennis, Jason A. Steen, Donovan H. Parks, Gene W. Tyson, and Philip Hugenholtz[†]

[†] Correspondence to: Philip Hugenholtz, p.hugenholtz@uq.edu.au

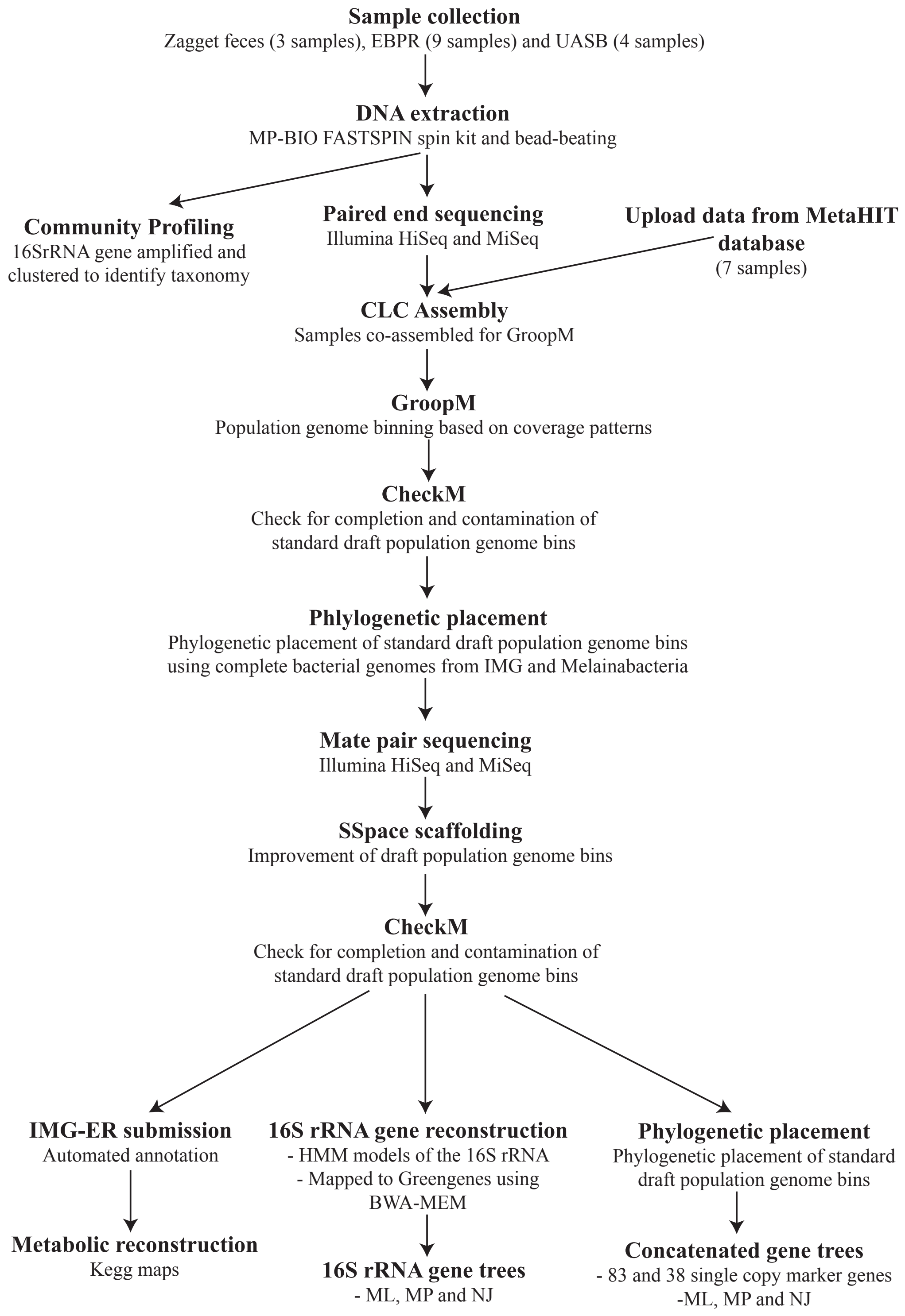
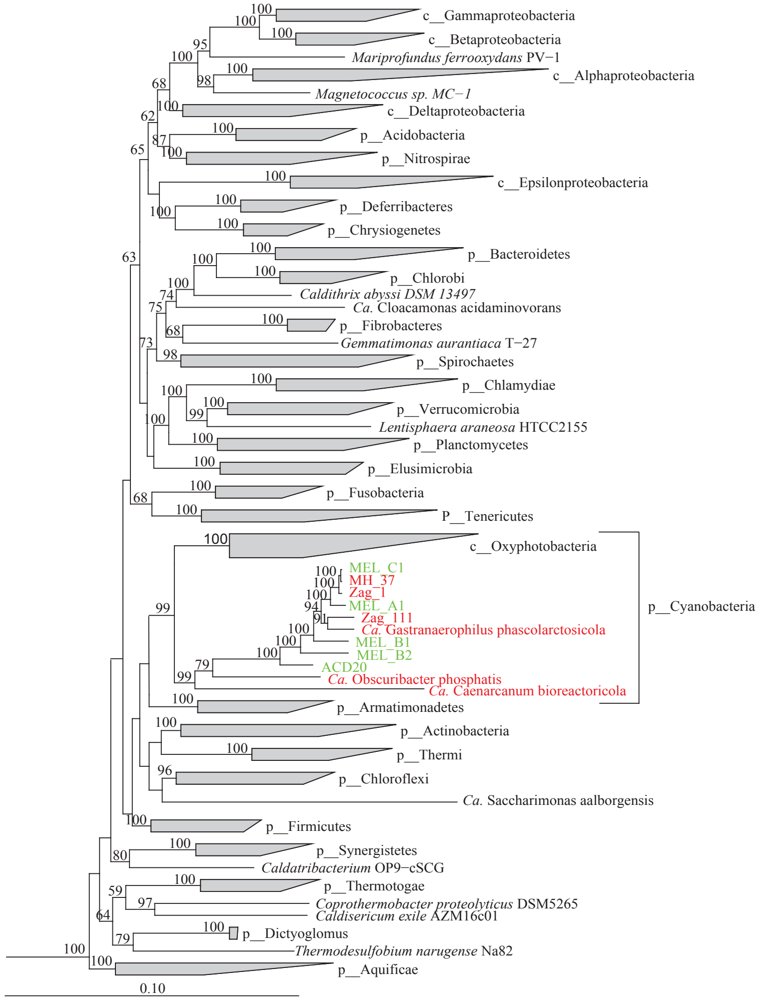
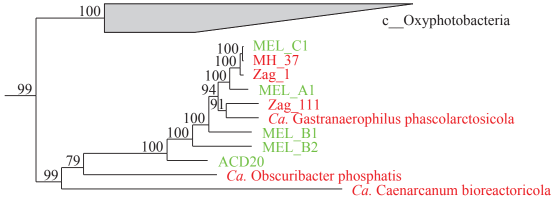


Fig. S1. Flow diagram of the Methods used in this paper

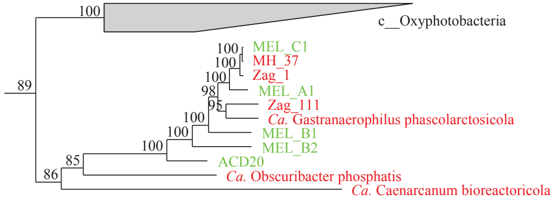
ML (RAxML, JTT+Gamma, 100x bootstraping) rooted with Archaea



ML (RAxML, JTT+Gamma, 100x bootstraping)



ML (FastTree, JTT+CAT, 100x bootstraping) on RAxML tree



MP (PAUP*, 100x bootstraping, >50% consensus tree)

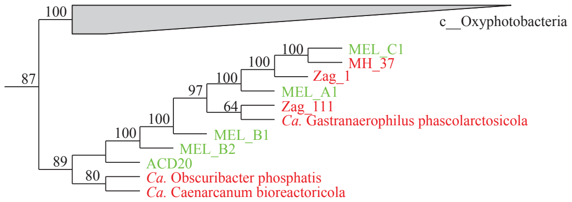
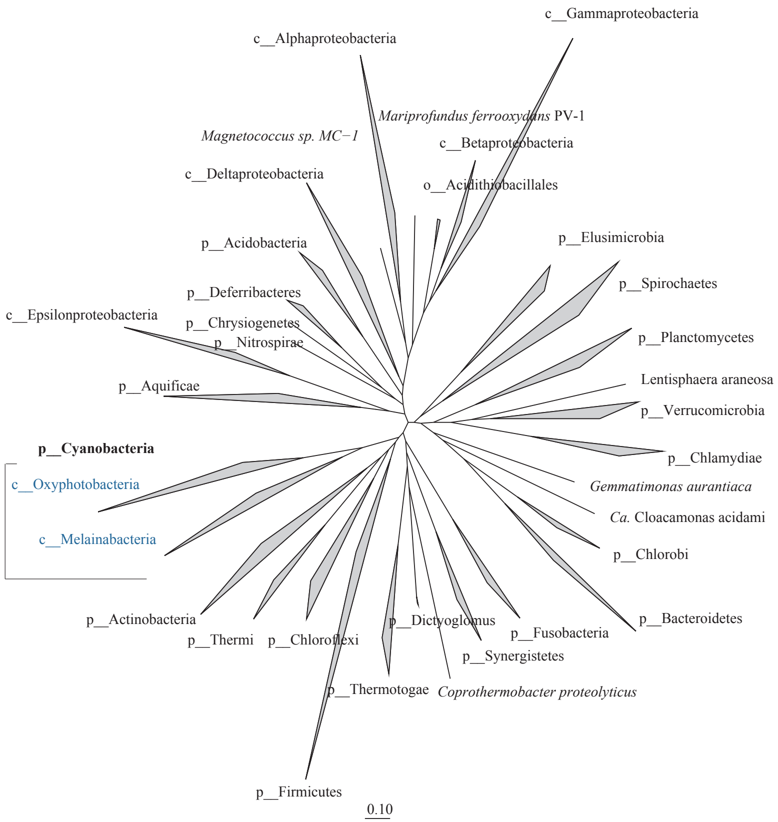


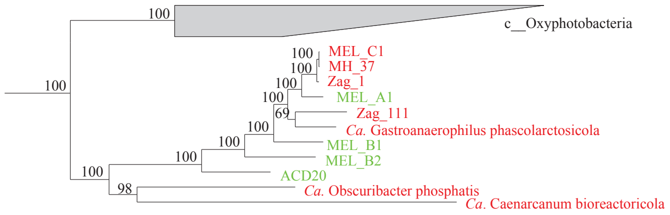
Fig. S2. Phylogeny of Oxyphotobacteria and Melainabacteria genomes among the bacterial phyla based on up to 38 marker genes

Phylogenetic maximum likelihood (RAxML, JTT, G) and maximum parsimony (PAUP*) trees based on a concatenated alignment of up to 38 marker genes, showing the phylogenetic robustness of the phylum Cyanobacteria and the classes Oxyphotobacteria and Melainabacteria. 422 OTUs (operational taxonomic units) from Bacteria and Archaea were used to produce the phylogenetic tree (**Table S4**). Bootstrap analyses (100 times) were performed for the data set with maximum likelihood (RAxML, JTT, G; FastTree, JTT, CAT) and maximum parsimony (PAUP*) methods, and the values obtained are shown in respective trees, except for the values from FastTree, which are shown on the tree generated with RAxML bootstrapping. Genomes in green are representatives from Di Rienzi et al., 2013 and genomes in red are Melainabacteria from this study. Bootstrap values >60% are shown. p_ represents phylum and c_ represents class. *Candidatus* has been abbreviated to *Ca.* for the most complete genomes.

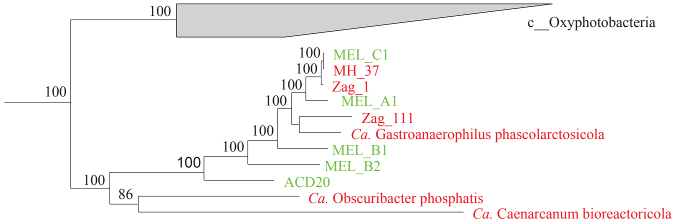
ML (RAxML, JTT+Gamma, 100x bootstraping) unrooted



ML (RAxML, JTT+Gamma, 100x bootstraping)



ML (FastTree, JTT+CAT, 100x bootstraping) on RAxML tree



MP (PAUP*, 100x bootstraping, >50% consensus tree)

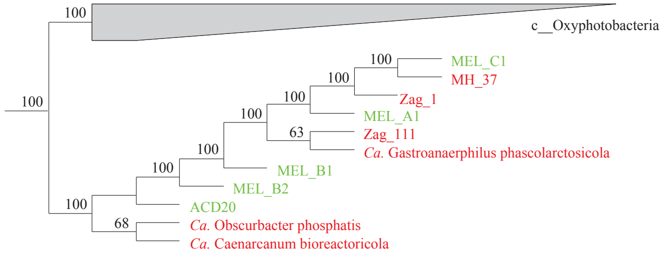
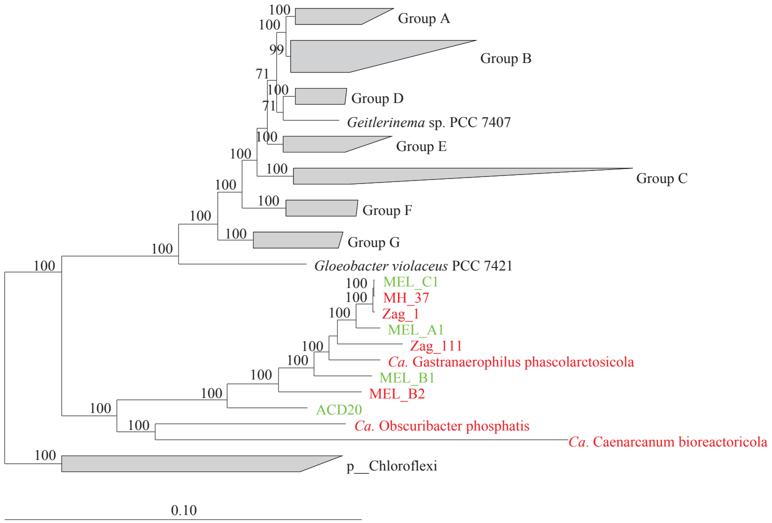
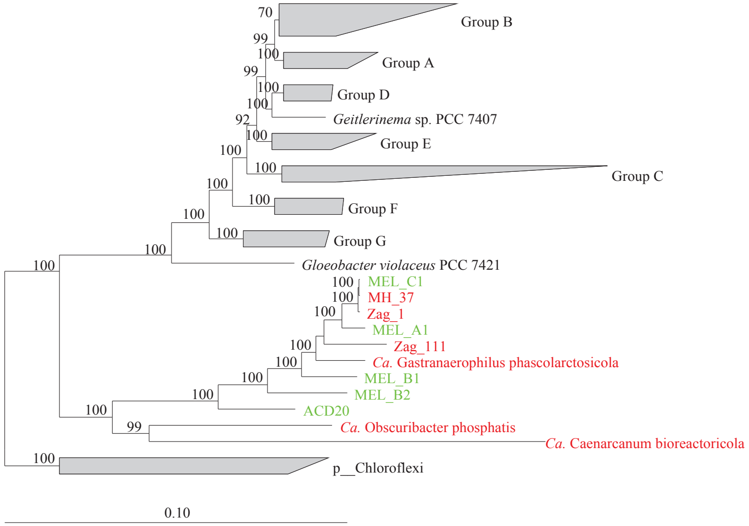


Fig. S3. Phylogeny of Oxyphotobacteria and Melainabacteria among the bacterial phyla based on up to 83 marker genes

Phylogenetic maximum likelihood (RAxML, JTT, G) and maximum parsimony (PAUP*) trees based on a concatenated alignment of up to 83 marker genes, showing the phylogenetic robustness of the phylum Cyanobacteria and the classes Oxyphotobacteria and Melainabacteria. 322 OTUs from Bacteria were used to produce the phylogenetic tree (**Table S4**). Bootstrap analyses (100 times) were performed for the data set with maximum likelihood (RAxML, JTT, G; FastTree, JTT, CAT) and maximum parsimony (PAUP*) methods, and the values obtained are shown in respective trees, except for the values from FastTree, which are shown on the tree generated with RAxML. Genomes in green are representatives from Di Rienzi et al., 2013 and genomes in red are Melainabacteria from this study. Bootstrap values >60% are shown. p_ represents phylum, c_ represents class and o_ represents order. *Candidatus* has been abbreviated to *Ca.* for the most complete genomes.


ML (RAxML, JTT+Gamma, 100x bootstraping)


ML (FastTree, JTT+CAT, 100x bootstraping) on RAxML tree

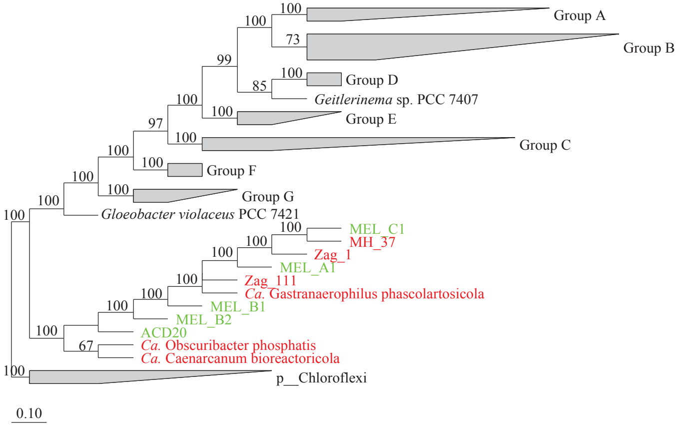

MP (PAUP*, 100x bootstraping, >50% consensus tree)

Fig. S4. Phylogeny of Oxyphotobacteria and Melainabacteria in the Cyanobacteria phylum based on up to 83 marker genes

Phylogenetic maximum likelihood (RAxML, JTT, G) and maximum parsimony (PAUP*) trees based on a concatenated alignment of up to 83 marker genes, showing the phylogenetic robustness of the phylum Cyanobacteria and the classes with Chloroflexi used as the outgroup. The Oxyphotobacteria were grouped as according to Shih et al., 2013 (**Table S3**). Bootstrap analyses (100 times) were performed for the data set with maximum likelihood (RAxML, JTT, G; FastTree, JTT, CAT) and maximum parsimony (PAUP*) methods, and the values obtained are shown in respective trees, except for the values from FastTree, which are shown on the tree generated with RAxML. Genomes in green are representatives from Di Rienzi et al., 2013 and genomes in red are Melainabacteria from this study. Bootstrap values >60% are shown. p_ represents phylum, c_ represents class and o_ represents order. *Candidatus* has been abbreviated to *Ca.* for the most complete genomes.

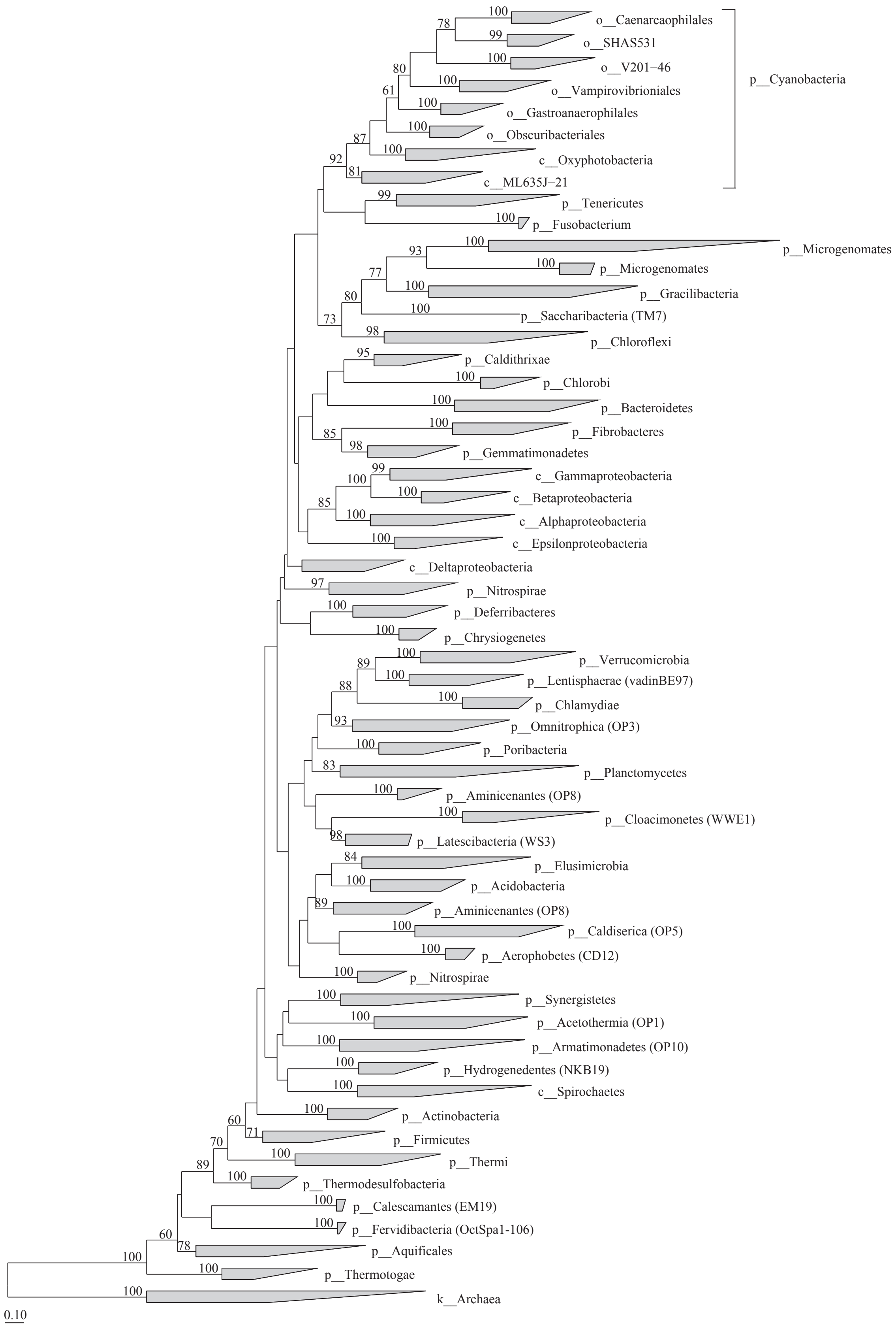

100
o__Caenarcaophilales
78
99
o__SHAS531
100
o__V201−46
80
100
o__Vampirovibrioniales
61
100
o__Gastroanaerophilales
100
o__Obscuribacteriales
87
100
92
c__Oxyphotobacteria
81
c__ML635J−21
p__Cyanobacteria
99
p__Tenericutes
100
p__Fusobacterium
100
93
p__Microgenomates
100
p__Microgenomates
77
100
p__Gracilibacteria
80
100
p__Saccharibacteria (TM7)
73
98
p__Chloroflexi
95
p__Caldithrixae
100
p__Chlorobi
100
p__Bacteroidetes
100
p__Fibrobacteres
85
98
p__Gemmatimonadetes
99
c__Gammaproteobacteria
100
100
c__Betaproteobacteria
85
100
c__Alphaproteobacteria
100
c__Epsilonproteobacteria
c__Deltaproteobacteria
97
p__Nitrospirae
100
p__Deferribacteres
100
p__Chrysiogenetes
100
89
p__Verrucomicrobia
100
p__Lentisphaerae (vadinBE97)
88
100
p__Chlamydiae
93
p__Omnitrophica (OP3)
100
p__Poribacteria
83
p__Planctomycetes
100
p__Aminicenantes (OP8)
100
p__Cloacimonetes (WWE1)
98
p__Latescibacteria (WS3)
84
p__Elusimicrobia
100
p__Acidobacteria
89
p__Aminicenantes (OP8)
100
p__Caldiserica (OP5)
100
p__Aerophobetes (CD12)
100
p__Nitrospirae
100
p__Synergistetes
100
p__Acetothermia (OP1)
100
p__Armatimonadetes (OP10)
100
p__Hydrogenedentes (NKB19)
100
c__Spirochaetes
100
p__Actinobacteria
60
71
p__Firmicutes
70
100
p__Thermi
89
100
p__Thermodesulfobacteria
100
p__Calescamantes (EM19)
100
p__Fervidibacteria (OctSpa1-106)
60
78
p__Aquificales
100
100
p__Thermotogae
100
k__Archaea
0.10

Fig. S5. 16S rRNA gene tree showing the phylogenetic robustness of the Cyanobacteria and its class level lineages in the bacterial domain

Maximum likelihood (RAxML, GTR, G, I; FastTree, JTT, CAT) trees based on the 16S rRNA gene, showing the phylogenetic robustness of the classes Oxyphotobacteria and Melainabacteria. 418 OTUs (operational taxonomic units) from Bacteria and Archaea were used to construct the trees. Bootstrap analyses (100 times) were performed for the data set with maximum likelihood (RAxML, GTR, G, I; FastTree, JTT, CAT). Bootstrap values >60% are shown.

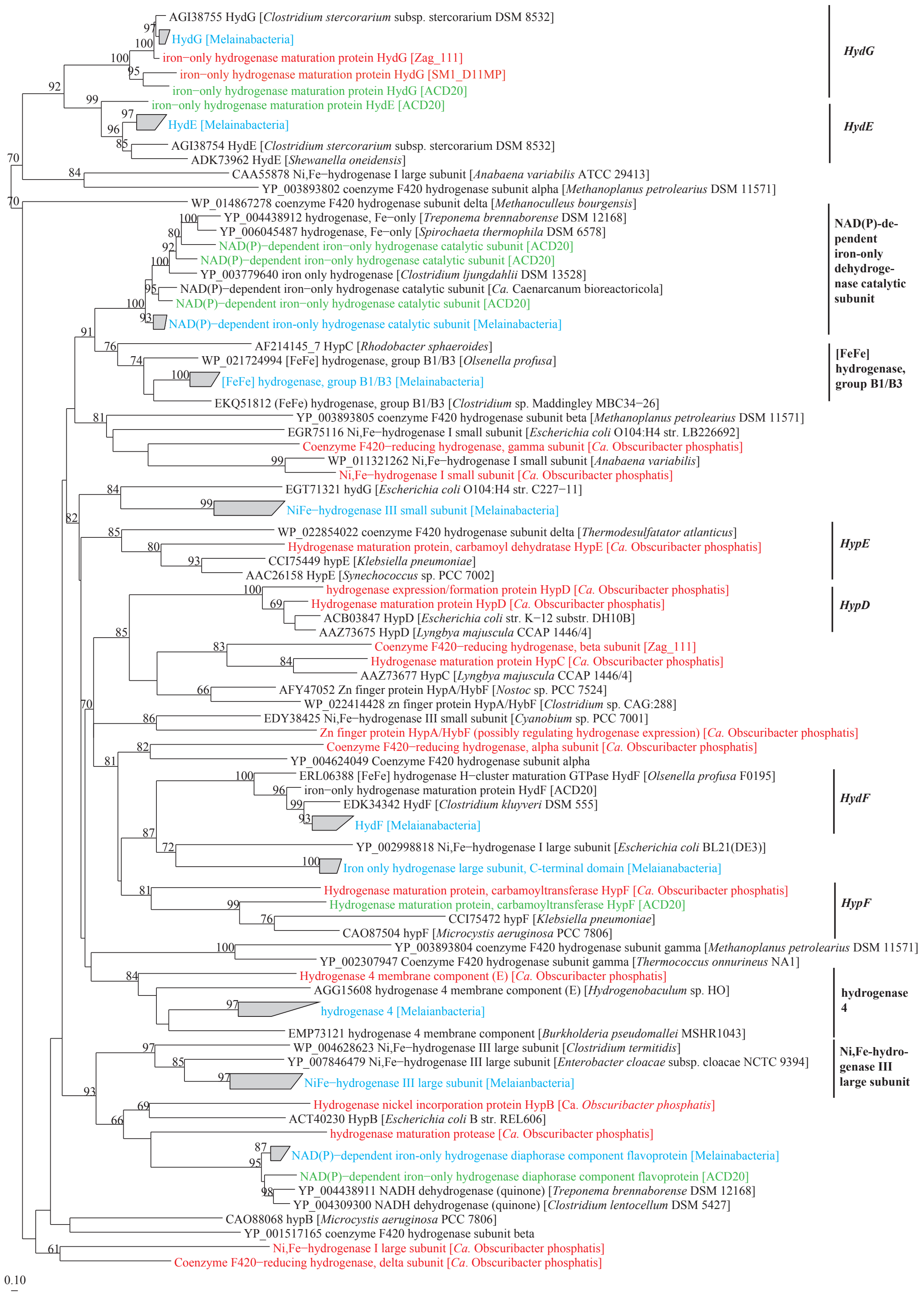

97
AGI38755 HydG [Clostridium stercorarium subsp. stercorarium DSM 8532]
100
HydG [Melainabacteria]
100
iron−only hydrogenase maturation protein HydG [Zag_111]
95
iron−only hydrogenase maturation protein HydG [SM1_D11MP]
92
iron−only hydrogenase maturation protein HydG [ACD20]
99
iron−only hydrogenase maturation protein HydE [ACD20]
97
HydE [Melainabacteria]
96
85
AGI38754 HydE [Clostridium stercorarium subsp. stercorarium DSM 8532]
ADK73962 HydE [Shewanella oneidensis]
70
84
CAA55878 Ni,Fe−hydrogenase I large subunit [Anabaena variabilis ATCC 29413]
YP_003893802 coenzyme F420 hydrogenase subunit alpha [Methanoplanus petrolearius DSM 11571]
70
WP_014867278 coenzyme F420 hydrogenase subunit delta [Methanoculleus bourgensis]
100
YP_004438912 hydrogenase, Fe−only [Treponema brennaborense DSM 12168]
80
YP_006045487 hydrogenase, Fe−only [Spirochaeta thermophila DSM 6578]
92
NAD(P)−dependent iron−only hydrogenase catalytic subunit [ACD20]
100
NAD(P)−dependent iron−only hydrogenase catalytic subunit [ACD20]
YP_003779640 iron only hydrogenase [Clostridium ljungdahlii DSM 13528]
95
NAD(P)−dependent iron−only hydrogenase catalytic subunit [Ca. Caenarcanum bioreactoricola]
100
NAD(P)−dependent iron−only hydrogenase catalytic subunit [ACD20]
93
NAD(P)−dependent iron-only hydrogenase catalytic subunit [Melainabacteria]
91
76
AF214145_7 HypC [Rhodobacter sphaeroides]
74
WP_021724994 [FeFe] hydrogenase, group B1/B3 [Olsenella profusa]
100
[FeFe] hydrogenase, group B1/B3 [Melainabacteria]
EKQ51812 (FeFe) hydrogenase, group B1/B3 [Clostridium sp. Maddingley MBC34−26]
81
YP_003893805 coenzyme F420 hydrogenase subunit beta [Methanoplanus petrolearius DSM 11571]
EGR75116 Ni,Fe−hydrogenase I small subunit [Escherichia coli O104:H4 str. LB226692]
Coenzyme F420−reducing hydrogenase, gamma subunit [Ca. Obscuribacter phosphatis]
99
WP_011321262 Ni,Fe−hydrogenase I small subunit [Anabaena variabilis]
Ni,Fe−hydrogenase I small subunit [Ca. Obscuribacter phosphatis]
84
EGT71321 hydG [Escherichia coli O104:H4 str. C227−11]
99
NiFe−hydrogenase III small subunit [Melainabacteria]
82
85
WP_022854022 coenzyme F420 hydrogenase subunit delta [Thermodesulfatator atlanticus]
80
Hydrogenase maturation protein, carbamoyl dehydratase HypE [Ca. Obscuribacter phosphatis]
93
CCI75449 hypE [Klebsiella pneumoniae]
AAC26158 HypE [Synechococcus sp. PCC 7002]
100
hydrogenase expression/formation protein HypD [Ca. Obscuribacter phosphatis]
69
Hydrogenase maturation protein HypD [Ca. Obscuribacter phosphatis]
ACB03847 HypD [Escherichia coli str. K−12 substr. DH10B]
85
AAZ73675 HypD [Lyngbya majuscula CCAP 1446/4]
83
Coenzyme F420−reducing hydrogenase, beta subunit [Zag_111]
84
Hydrogenase maturation protein HypC [Ca. Obscuribacter phosphatis]
AAZ73677 HypC [Lyngbya majuscula CCAP 1446/4]
66
AFY47052 Zn finger protein HypA/HybF [Nostoc sp. PCC 7524]
70
WP_022414428 zn finger protein HypA/HybF [Clostridium sp. CAG:288]
86
EDY38425 Ni,Fe−hydrogenase III small subunit [Cyanobium sp. PCC 7001]
Zn finger protein HypA/HybF (possibly regulating hydrogenase expression) [Ca. Obscuribacter phosphatis]
82
Coenzyme F420−reducing hydrogenase, alpha subunit [Ca. Obscuribacter phosphatis]
81
YP_004624049 Coenzyme F420 hydrogenase subunit alpha
100
ERL06388 [FeFe] hydrogenase H−cluster maturation GTPase HydF [Olsenella profusa F0195]
96
iron−only hydrogenase maturation protein HydF [ACD20]
99
EDK34342 HydF [Clostridium kluyveri DSM 555]
93
HydF [Melaianabacteria]
87
72
YP_002998818 Ni,Fe−hydrogenase I large subunit [Escherichia coli BL21(DE3)]
100
Iron only hydrogenase large subunit, C-terminal domain [Melaianabacteria]
81
Hydrogenase maturation protein, carbamoyltransferase HypF [Ca. Obscuribacter phosphatis]
99
Hydrogenase maturation protein, carbamoyltransferase HypF [ACD20]
76
CCI75472 hypF [Klebsiella pneumoniae]
CAO87504 hypF [Microcystis aeruginosa PCC 7806]
100
YP_003893804 coenzyme F420 hydrogenase subunit gamma [Methanoplanus petrolearius DSM 11571]
YP_002307947 Coenzyme F420 hydrogenase subunit gamma [Thermococcus onnurineus NA1]
84
Hydrogenase 4 membrane component (E) [Ca. Obscuribacter phosphatis]
AGG15608 hydrogenase 4 membrane component (E) [Hydrogenobaculum sp. HO]
97
hydrogenase 4 [Melaianbacteria]
EMP73121 hydrogenase 4 membrane component [Burkholderia pseudomallei MSHR1043]
97
WP_004628623 Ni,Fe−hydrogenase III large subunit [Clostridium termitidis]
85
YP_007846479 Ni,Fe−hydrogenase III large subunit [Enterobacter cloacae subsp. cloacae NCTC 9394]
97
NiFe−hydrogenase III large subunit [Melaianbacteria]
93
69
Hydrogenase nickel incorporation protein HypB [Ca. Obscuribacter phosphatis]
ACT40230 HypB [Escherichia coli B str. REL606]
66
hydrogenase maturation protease [Ca. Obscuribacter phosphatis]
87
NAD(P)−dependent iron-only hydrogenase diaphorase component flavoprotein [Melainabacteria]
95
NAD(P)−dependent iron−only hydrogenase diaphorase component flavoprotein [ACD20]
98
YP_004438911 NADH dehydrogenase (quinone) [Treponema brennaborense DSM 12168]
YP_004309300 NADH dehydrogenase (quinone) [Clostridium lentocellum DSM 5427]
CAO88068 hypB [Microcystis aeruginosa PCC 7806]
YP_001517165 coenzyme F420 hydrogenase subunit beta
61
Ni,Fe−hydrogenase I large subunit [Ca. Obscuribacter phosphatis]
Coenzyme F420−reducing hydrogenase, delta subunit [Ca. Obscuribacter phosphatis]
HydG
HydE
NAD(P)-dependent iron-only dehydrogenase catalytic subunit
[FeFe] hydrogenase, group B1/B3
HypE
HypD
HydF
HypF
hydrogenase 4
Ni,Fe-hydrogenase III large subunit
0.10

Fig. S6. Gene tree showing the hydrogenase genes found in the Melainabacteria representatives.

A representative collection of hydrogenase genes publicly available from NCBI was combined with genes identified from this study and from Di Rienzi et al., 2013. Hydrogenase genes common to genomes from both studies are colored blue and grouped together. Hydrogenases found only in this study are colored red and genes only identified in the genomes of Di Rienzi et al., 2013 are colored green.

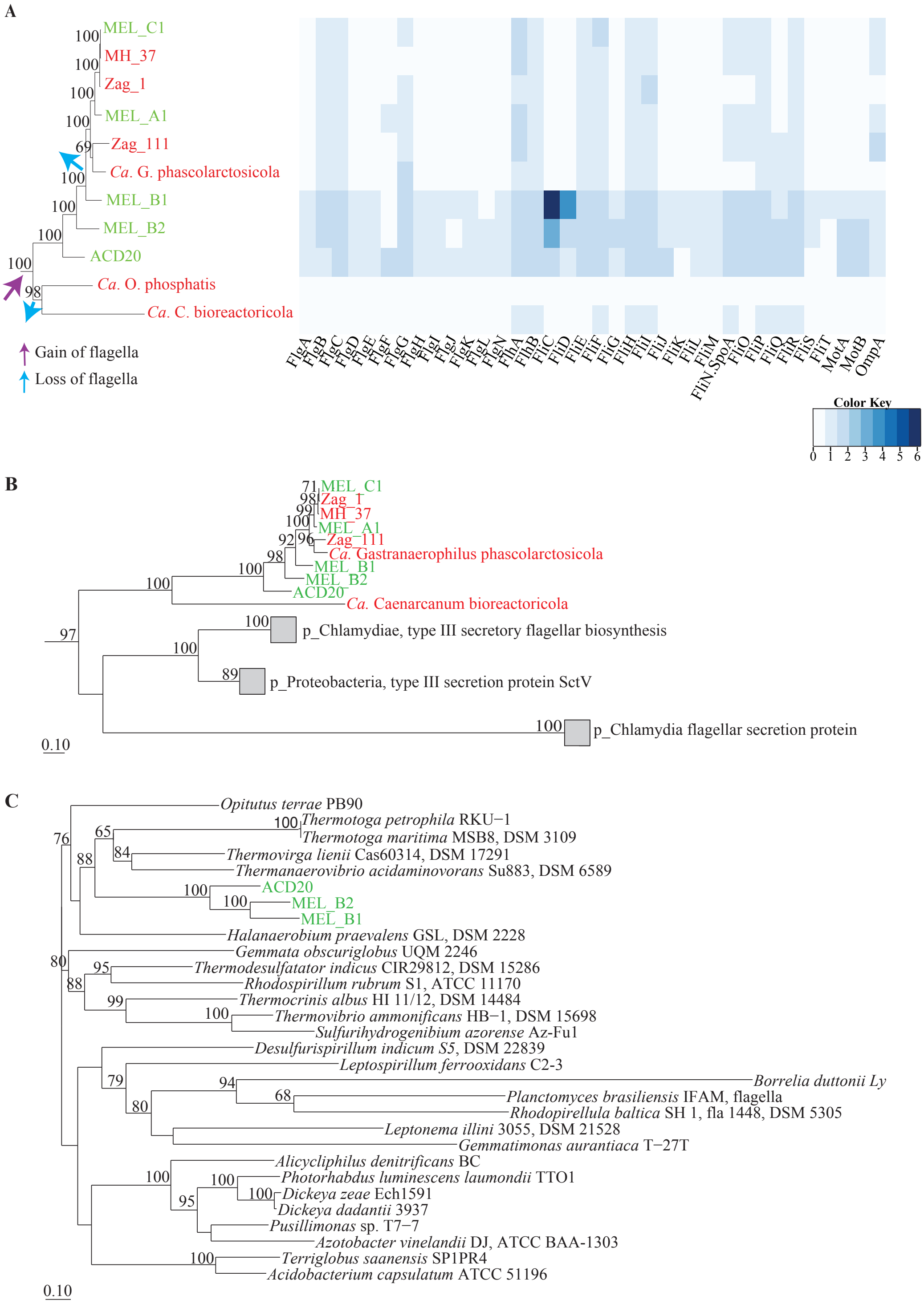

A
MEL_C1
MH_37
Zag_1
MEL_A1
Zag_111
Ca. G. phascolarctosicola
MEL_B1
MEL_B2
ACD20
Ca. O. phosphatis
Ca. C. bioreactoricola
100
100
100
100
69
100
100
100
100
98
Gain of flagella
Loss of flagella
FlgA
FlgB
FlgC
FlgD
FlgE
FlgF
FlgG
FlgH
FlgI
FlgJ
FlgK
FlgL
FlgN
FlhA
FlhB
FliC
FliD
FliE
FliF
FliG
FliH
FliI
FliJ
FliK
FliL
FliM
FliN.SpoA
FliO
FliP
FliQ
FliR
FliS
FliT
MotA
MotB
OmpA
Color Key
0 1 2 3 4 5 6
B
MEL_C1
Zag_1
MH_37
MEL_A1
Zag_111
Ca. Gastranaerophilus phascolarctosicola
MEL_B1
MEL_B2
ACD20
Ca. Caenarcanum bioreactoricola
71
98
99
100
92
96
98
100
100
97
100
p_Chlamydiae, type III secretory flagellar biosynthesis
100
89
p_Proteobacteria, type III secretion protein SctV
100
p_Chlamydia flagellar secretion protein
0.10
C
Opitutus terrae PB90
Thermotoga petrophila RKU−1
Thermotoga maritima MSB8, DSM 3109
Thermovirga lienii Cas60314, DSM 17291
Thermanaerovibrio acidaminovorans Su883, DSM 6589
ACD20
MEL_B2
MEL_B1
Halanaerobium praevalens GSL, DSM 2228
Gemmata obscuriglobus UQM 2246
Thermodesulfatator indicus CIR29812, DSM 15286
Rhodospirillum rubrum S1, ATCC 11170
Thermocrinis albus HI 11/12, DSM 14484
Thermovibrio ammonificans HB−1, DSM 15698
Sulfurihydrogenibium azorense Az-Fu1
Desulfurispirillum indicum S5, DSM 22839
Leptospirillum ferrooxidans C2-3
Borrelia duttonii Ly
Planctomyces brasiliensis IFAM, flagella
Rhodopirellula baltica SH 1, fla 1448, DSM 5305
Leptonema illini 3055, DSM 21528
Gemmatimonas aurantiaca T−27T
Alicycliphilus denitrificans BC
Photorhabdus luminescens laumondii TTO1
Dickeya zeae Ech1591
Dickeya dadantii 3937
Pusillimonas sp. T7−7
Azotobacter vinelandii DJ, ATCC BAA-1303
Terriglobus saanensis SP1PR4
Acidobacterium capsulatum ATCC 51196
76
65
100
88
84
100
100
80
95
88
99
100
79
94
68
80
100
100
95
100
100
0.10

Fig. S7. Heat map showing the presence and absence of flagella genes from the Melainabacteria representatives

(A) A maximum parsimony tree based on a concatenated alignment of up to 83 marker genes. Purple arrows indicate the gain of a functional flagella and blue arrows indicate the loss of a functional flagella. The heat map indicates the square root of the number of flagella genes found in each of the Melainabacteria representatives. (B) Flagella gene trees for FlhA and (C) flagella gene tree for FliM. Genomes used to make the trees are found in **Table S6**. Genomes in green are Melainabacteria representatives from Di Rienzi et al., 2013, genomes in red are Melaianabacteria representatives from this study. *Candidatus* has been abbreviated to *Ca.* for the most complete genomes.

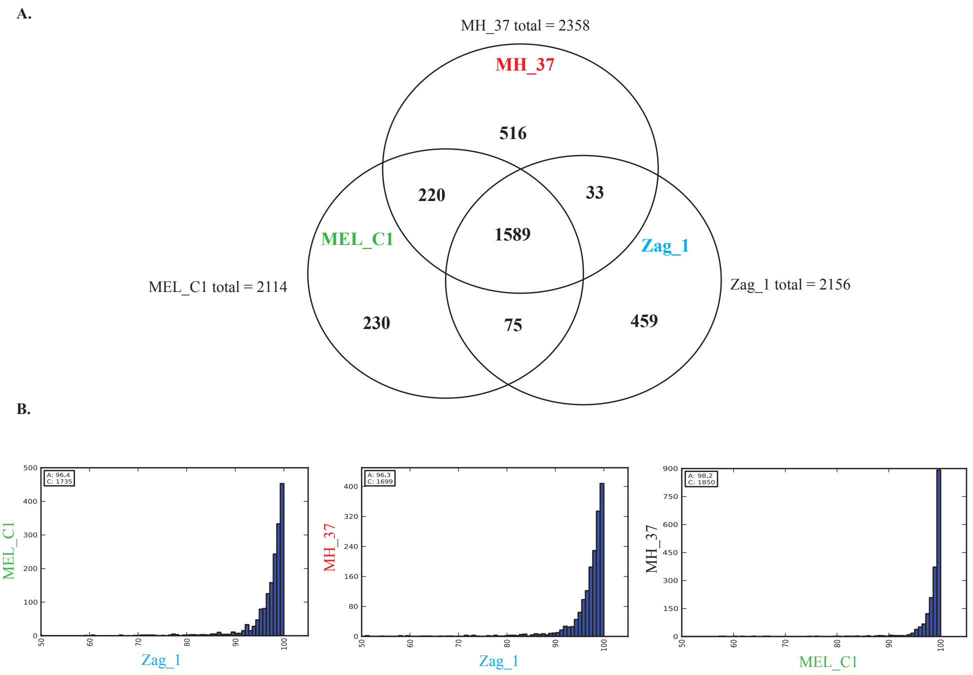


Fig. S8. Venn diagram of the three Gastranaerophilales genomes from the same species

(A) Shared orthologs between MH_37, MEL_C1 and Zag_1 showing the number of genes that are core to the three genomes, and those that are found between two genomes and individual genomes. The total number for each genome does not include paralogs. (B) Average nucleotide identity between MH_37, MEL_C1 and Zag_1, where ≥95% ANI is used to define a species.

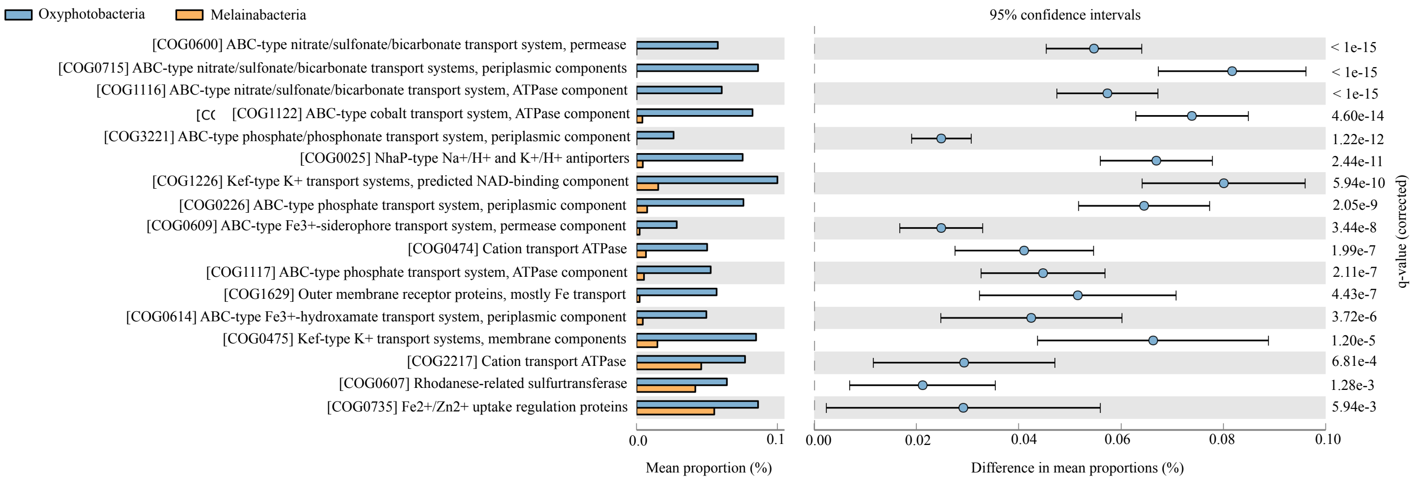


Fig. S9. COG category P (inorganic ion transport and metabolism) for Oxyphotobacteria and Melainabacteria

Shown are all COGs from category P with a difference in mean proportions >= 0.02% between Oxyphotobacteria and Melainabacteria and a Storey's q-value of <= 0.05.

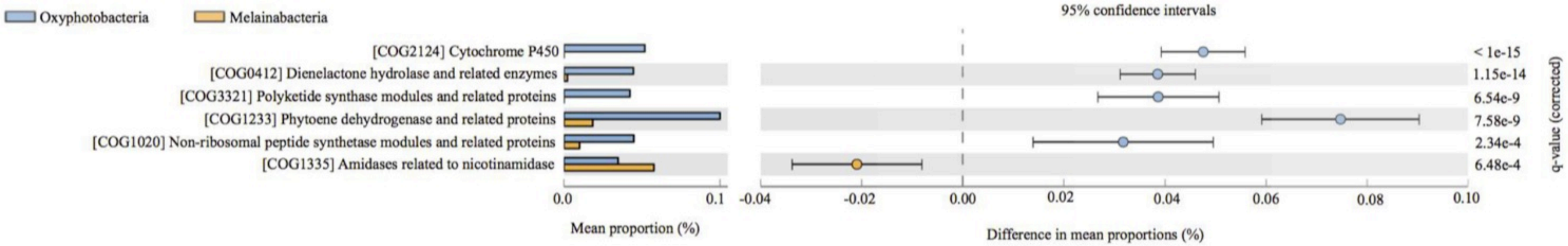


Fig. S10. COG category Q (secondary metabolites and biosynthesis, transport and catabolism) for Oxyphotobacteria and Melainabacteria

Shown are all COGs from category Q with a difference in mean proportions >= 0.02% between Oxyphotobacteria and Melainabacteria and a Storey's q-value of <= 0.05