Evolutionary Implications of Anoxygenic Phototrophy in the Bacterial Phylum *Candidatus* Palusbacterota (WPS-2)

Lewis M. Ward[1], Tanai Cardona[2], Hannah Holland-Moritz[3,4]

[1]Department of Earth and Planetary Sciences, Harvard University
[2]Department of Life Sciences, Imperial College London
[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder
[4]Department of Ecology and Evolutionary Biology, University of Colorado at Boulder

Corresponding authors: lewis_ward@fas.harvard.edu, t.cardona@imperial.ac.uk, hannah.hollandmoritz@colorado.edu

Abstract

Genome-resolved environmental metagenomic sequencing has uncovered substantial previously unrecognized microbial diversity relevant for understanding the ecology and evolution of the biosphere, providing a more nuanced view of the distribution and ecological significance of traits including phototrophy across diverse niches. Recently, the capacity for bacteriochlorophyll-based anoxygenic photosynthesis has been found in the uncultured bacterial WPS-2 clade that are in close association with boreal moss. Here, we use phylogenomic analysis to investigate the diversity and evolution of phototrophic WPS-2. We demonstrate that phototrophic WPS-2 show significant genetic and metabolic divergence from other phototrophic and non-phototrophic lineages. The genomes of these organisms encode a completely new family of anoxygenic Type II photochemical reaction centers and other phototrophy-related proteins that are both phylogenetically and structurally distinct from those found in previously described phototrophs. We propose the name *Candidatus* Palusbacterota for the phylum-level aerobic WPS-2 clade which contains phototrophic lineages, from the Latin for "bog bacteria", distinguishing it from the anaerobic, non-phototrophic sister phylum *Candidatus* Eremiobacterota for "desert bacteria", typically found in dry environments.

Introduction

The vast majority of primary productivity on Earth is fueled by photosynthesis, both today[1] and through most of the history of life[2-4], making the organisms and proteins responsible for driving photosynthesis into critical bases for the carbon cycle. While the quantitatively most significant form of photosynthesis today is oxygenic photosynthesis, which uses water as an electron donor to power carbon fixation, there is also a great diversity of bacteria capable of anoxygenic photosynthesis using compounds such as sulfide, molecular hydrogen, ferrous iron, or arsenic as electron donors[5,6]. Anoxygenic photosynthesis is restricted to members of the bacterial domain, where it has a scattered phylogenetic distribution (Figure 1). While the first observations of anoxygenic photosynthesis were made in the late 1800s[7], the known diversity of bacteria capable of anoxygenic phototrophy has exploded in recent years thanks largely to genome-resolved metagenomics[8,9] (Figure 1). As our understanding of the physiology and evolution of phototrophy is contingent on adequate sampling of the

diversity of phototrophic organisms, continuing discoveries of phylogenetically, ecologically, and biochemically novel phototrophs spurs new insights into phototrophy today and in Earth's history.

Unlike oxygenic photosynthesis which requires the work of Type I and Type II photochemical reaction centers linked in series for water oxidation, anoxygenic phototrophs use exclusively either Type I or Type II reaction centers. Type I and II reaction centers can be differentiated by the nature of their electron acceptor[5]. Type I reduce ferredoxin, while Type II reduce quinones. There are only four phyla of bacteria known to have phototrophic representatives that use anoxygenic Type II reaction centers: Proteobacteria, Chloroflexi, Gemmatimonadetes[10] and the newly discovered WPS-2[11] (Figure 1). The photochemical pigments of anoxygenic Type II reaction centers are bound by two homologous subunits known as L (*PufL*) and M (*PufM*). These have traditionally been subdivided into two types, the L and M found in the reaction centers of phototrophic Proteobacteria (PbRC) and the L and M found in those of the phototrophic Chloroflexi (CfRC), with each set making separate monophyletic clades[12,13]. Therefore, substantial differences exist between the PbRC and the CfRC not only at the level of photochemistry, but also at the level of sequence identity of the core subunits, pigment and subunit composition[14,15]. Phototrophic Gemmatimonadetes, on the other hand, encode Proteobacteia-like reaction centers, as Zeng et al.[10] demonstrated that this group obtained phototrophy via horizontal gene transfer of an entire photosynthesis gene cluster from a gammaproteobacterium.

Here we show that the genomes of the newly discovered phototrophs of the uncultivated candidate phylum WPS-2 encode a third distinct lineage of anoxygenic Type II reaction centers with novel and unusual characteristics—only the third of this type after the discovery of the Chloroflexi nearly half a century ago[16]. We also show that this phylum is distinct from its relatives in terms of traits including the capacity for aerobic respiration, membrane architecture, and environmental distribution, reflecting a divergent evolutionary history and likely novel ecological roles.

Methods

Metagenome Analyses

Metagenome-assembled genomes (MAGs) of WPS-2 bacteria were downloaded from NCBI WGS and JGI IMG databases. Completeness and contamination of genomes was estimated based on presence and copy number of conserved single-copy proteins by CheckM[17]. Sequences of ribosomal and metabolic proteins used in analyses (see below) were identified locally with the *tblastn* function of BLAST+[18], aligned with MUSCLE[19], and manually curated in Jalview[20]. Positive BLAST hits were considered to be full length (e.g. >90% the shortest reference sequence from an isolate genome) with *e*-values greater than $1e^{-20}$. Genes of interest were screened against outlier (e.g. likely contaminant) contigs as determined by CheckM[17] and RefineM[21] using tetranucleotide, GC, and coding density content. Presence of metabolic pathways of interest was predicted with MetaPOAP[22] to check for False Positives (contamination) or False Negatives (genes present in source genome but not recovered in metagenome-assembled genomes). Phylogenetic trees were calculated using RAxML[23] on the Cipres science gateway[24]. Transfer bootstrap support values were calculated by BOOSTER[25], and trees were visualized with the Interactive Tree of Life viewer[26]. Taxonomic assignment was confirmed with GTDB-Tk[27] and by placement in a concatenated ribosomal protein phylogeny following methods from Hug et al.[28] Histories of vertical versus horizontal inheritance of metabolic genes was inferred by comparison of organismal and metabolic protein phylogenies to determine topological congruence[9,29].

Reaction center evolution analyses

A total of 14 L and 12 M amino acid sequences were collected from the compiled metagenome data. The sequences were added to a dataset of Type II reaction center subunits compiled before[30], which included sequences from Cyanobacteria, Proteobacteria, and Chloroflexi. Sequence alignments were done in Clustal Omega[31] using 10 combined guide trees and Hidden Markov Model iterations. Maximum Likelihood phylogenetic analysis was performed with the PhyML online service (http://www.atgc-montpellier.fr/phyml/)[32] using the Smart Model Selection and Bayesian information criterion for the computation of parameters from the dataset[33]. Tree search operations were performed with the Nearest Neighbor Interchange approach and the approximate likelihood-ratio test method was chosen for the computation of branch support values. A phylogenetic tree was also constructed using only L and M subunits and excluding cyanobacterial homologs.

Transmembrane helix prediction on WPS-2 L and M subunits was computed using the TMHMM Server, v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)[34] and the ΔG transmembrane helix prediction tool v. 1.0 (http://dgpred.cbr.su.se/)[35]. Structural homology models were carried out with SWISS-MODEL server (https://swissmodel.expasy.org/)[36] using the 1.9 Å crystal structure of the gammaproteobacterium *Thermochromatium tepidum* as template[14] (PDB ID: 5y5s). The WPS-2 L and M subunits annotated as Ga0175859_11240458 and Ga0175859_11402733 respectively were used for the structural reconstruction.

Results and Discussion

Ecology and distribution of WPS-2

Clonal libraries containing 16S rRNA sequences from the candidate WPS-2 (Wittenberg Polluted Soil) phylum were first identified in a 2001 study of soil polluted with polychlorinated biphenyl (PCB) in Germany[37]. Since then, WPS-2 has been found in 16S rRNA amplicon studies from diverse environments including acidic, polluted environments[37-39], alpine, high latitude and antarctic soils[40-43], cryoconite holes in Greenland[44], human/canine oral microbiomes[45], bogs and peatlands[46,47]. More recently, several metagenomic studies have successfully assembled MAGs from WPS-2. These studies include samples from an acidic mine, rich in arsenic[48] (MAGs from this study were later identified as WPS-2 by Camanocha and Dewhirst[45]), Arctic[21] and Antarctic desert soils[42], boreal mosses[11], and a peatland in Sweden[49].

Although these environments are varied, several trends emerge: WPS-2 has a global distribution and is most often found in cool, acidic, and aerobic environments. While not exclusively found in low-temperature environments, most samples containing WPS-2 were collected from areas that are typically cold (i.e. Antarctica, Greenland), or undergo shorter growing seasons (high-latitude and alpine environments). Similarly, WPS-2 is more often found in sites with acidic to moderately acidic pH (between 3 and 6.5). In some cases, this acidity has come from pollution, in others, such as acidic springs, peatlands, bogs, and fir-spruce forest soil, it is a natural part of the habitat. Although WPS-2 has often been associated with polluted environments, it seems more likely that this is due the acidity accompanying these environments than to any special resistance to the varied pollutants from these studies. Oxygen-rich environments such as mosses, cryoconite holes, and the top few centimeters of soil are typical habitat for WPS-2 and several papers have speculated that members of this phylum are therefore aerobic or microaerobic[11,40,42].

Often WPS-2 makes up a minor part of the bacterial community, however in some notable cases, the phylum has either dominated the bacterial community (e.g. 23-25%)[41,43,44] or several phylotypes from WPS-2 have been among the dominant taxa[11,47]. Although most studies identifying WPS-2 as dominant are from specialized or unique environments (i.e. acidic springs, Antarctic soils, and cryoconite holes), an easily-accessible and commonplace environment in which phylotypes of WPS-2 are often abundant, is boreal mosses. Phylotypes from WPS-2 are the second-most dominant taxa across 6 common boreal moss species and absent in only one moss of the 7 species studied[11]. Interestingly, even when mosses are not the focus of a study, WPS-2 is often more common in environments that contain them[39,46,47,49]. Currently there are no isolates of WPS-2 to investigate its biochemistry, or to use in comparative genomics, yet given its phototrophic properties and evolutionary context, such an isolate would be useful. As an accessible and abundant host to phylotypes of WPS-2, boreal mosses provide a logical starting place from which to try to culture WPS-2—an essential step in upgrading this candidate phylum to taxonomically valid status as well as for characterizing its physiology.

Phylogenetic context of phototrophy in the WPS-2 phylum

The WPS-2 clade is located on a branch of the bacterial tree of life located near the phyla Chloroflexi, Armatimonadetes, and Dormibacteria (Figure 1). This placement has interesting implications for the evolutionary history of a variety of traits (as discussed below) but is particularly relevant for considerations of the evolutionary history of anoxygenic phototrophy.

Phylogenetic analysis of proteins for the Type II reaction center suggests that phototrophy in WPS-2 might be most closely related to that of the Chloroflexi (Figure 2, but see below). Given the relatedness of WPS-2 and Chloroflexi phototrophy proteins (Figure 2, Supplemental Figure S1) and their relative closeness in organismal phylogenies (Figure 1), an important outstanding question is whether the last common ancestor of these phyla was phototrophic, implying extensive loss of phototrophy in both phyla and their relatives, or whether WPS-2 and Chloroflexi have independently acquired phototrophy via HGT subsequent to the divergence of these lineages (likely from an earlier branching uncharacterized or extinct group of phototrophic bacteria). While this cannot be conclusively ascertained with available data, it is apparent that aside from bacteriochlorophyll synthesis and phototrophic reaction centers, necessary proteins involved in phototrophic electron transfer (e.g. a *bc* complex or alternative complex III) are not shared, suggesting that members of these phyla independently acquired respiratory electron transfer necessary for phototrophy after their divergence. This is consistent with a relatively recent origin of phototrophy in the Chloroflexi (i.e. within the last ~1 billion years) after the acquisition of aerobic respiration by the Chloroflexi and other phyla following the Great Oxidation Event (GOE) ~2.3 billion years ago[50,51]. Together with the relatively high protein sequence similarity between reaction center proteins among the phototrophic WPS-2 (discussed below) this may suggest that phototrophy in the WPS-2 has radiated on a similar or even shorter timescale.

Phototrophy in WPS-2 is not restricted to a single lineage, but appears in at least four discrete lineages scattered around the known diversity of this phylum (Figure 3). The topology of the organismal phylogeny of the WPS-2 is incongruent with that of phototrophy protein phylogenies (Figure 3, Supplemental Figure S1), and so this distribution may best be explained by a history of HGT between members of the WPS-2 and not by a deeper ancestry of phototrophy followed by extensive loss. Similar cases of multiple HGT events of complete photosynthesis gene clusters have been demonstrated within the family Rhodobacteraceae of

Alphaproteobacteria[52], between classes of the phylum Chloroflexi[9], and into the Gemmatimonadetes phylum from members of the Proteobacteria[10], which might suggest convergent evolutionary patterns driving the distribution of Type II reaction center-based anoxygenic phototrophy within distant lineages.

The phylogenetic position of bacteriochlorophyll synthesis proteins including BchL of WPS-2 (Supplemental Figure S2) is somewhat in agreement with that of reaction center proteins (Figure 2, Supplemental Figure S1). However, BchL from the phylum Chlorobi is more closely related to that of the Chloroflexi than either is to the WPS-2. This is consistent with previous studies showing that HGT of protochlorophyllide and chlorophyllide reductase between stem group phototrophic Chlorobi and Chloroflexi had occurred[53,54], although the direction of transfer had remained ambiguous. The new topology of BchL including WPS-2 suggests exchange of bacteriochlorophyll synthesis proteins via HGT from stem group phototrophic Chloroflexi to stem group phototrophic Chlorobi. It is likely that exchange of photosynthetic components between Chloroflexi and Chlorobi had occurred more than once, in either direction, and at different time points and may have involved a swap of reaction center type as suggested before[53] and the acquisition of chlorosomes[55,56].

Additional metabolic traits in WPS-2 genomes

Aside from genes encoding phototrophy, the phototroph-containing WPS-2 clade is distinct from its relatives due to a variety of metabolic traits, including proteins used for aerobic respiration and carbon fixation, as well as evidence for an outer membrane.

Most WPS-2 genomes encode aerobic respiration via an A-family heme-copper oxidoreductase (HCO) and a *bc* complex III. This is in contrast to the Chloroflexi, which typically encode an alternative complex III instead of a *bc* complex. Protein sequences of respiratory proteins from WPS-2 genomes primarily form single closely related clades with topologies that broadly reflect organismal relationships (Figure 3, Supplemental Figures S3 and S4); these trends suggest vertical inheritance of these genes from the last common ancestor of this clade. The main exceptions to these trends are the divergent WPS-2 genomes assigned to the candidate phylum Eremiobacterota, which appear to be ancestrally anaerobic, and a handful of WPS-2 genomes that encode additional copies of respiratory proteins, which appear to have been acquired via horizontal gene transfer more recently and have since undergone additional HGT between members of the WPS-2.

Like the Chloroflexi, phototrophic WPS-2 typically encode a B-family HCO in addition to the ancestral A-family HCO used for respiration at relatively high O_2 concentrations[9]; the functional relationship between the B-family O_2 reductase and phototrophy is unclear, but may relate to oxygenic sensitivity of phototrophy-related proteins and the greater efficacy of B-family HCOs at low oxygen concentrations[57].

While phototrophic Chloroflexi typically perform carbon fixation via the 3-hydroxypropionate bicycle[50], carbon fixation in phototrophic WPS-2 is encoded via a Form I rubisco. All phototrophic genomes in this phylum encode closely related rubisco proteins related to that of *Kouleothrix aurantiaca*[9], with some also encoding a second Form I rubisco copy more closely related to sequences from *Bradyrhizobium* (Supplemental Figure S5). The PMHO01 and PMHP01 genomes encode rubisco despite not being phototrophic; this rubisco is not closely related to those encoded by phototrophic WPS-2, and so may reflect the independent acquisition of carbon fixation in some nonphototrophic WPS-2 lineages (perhaps to support a lithoautotrophic lifestyle). The

phylogeny of rubisco proteins is not congruent with organismal or phototrophy trees, suggesting that carbon fixation may have undergone an independent history of HGT, consistent with trends in the Chloroflexi[9,50].

The Chloroflexi are thought to possess a modified monoderm membrane[58,59]. WPS-2 genomes recovered genes for lipopolysaccharide synthesis and outer membrane proteins such as BamA, indicating that these organisms are diderm (i.e. possess an outer membrane), reinforcing interpretations of an ancestral diderm membrane architecture followed by secondary loss of the outer membrane in the phylum Chloroflexi[9,60-62].

Type II reaction centers in WPS-2

The phylogeny of Type II reaction center proteins showed that the L and M subunits of WPS-2 are distant, and clearly distinct, to those found in Proteobacteria and Chloroflexi (Figure 2), but the topology of the tree was unstable. It showed varying positions and levels of support for the WPS-2 sequences (Supplemental Figure S6). However, the sequence and structural comparison do reveal a slightly greater affinity and similarity to those of the Chloroflexi. The reaction center subunits show considerable diversity, with the level of sequence identity between the two most distant L and two most distant M subunits being about 59%. In comparison, the level of sequence identity between L or M subunits between two relatively distant strains of Chloroflexi, *Roseiflexus* spp. and *Chloroflexus* spp., is about 45% sequence identity. If the obtained sequences are representative of the diversity of phototrophic WPS-2, the greater sequence identity could potentially indicate a relatively more recent origin for their common ancestor in comparison to phototrophic Chloroflexi.

In addition to L and M subunits, the assembled genomes of WPS-2 also contain *pufC*, *puf2A* and *puf2C* encoding the tetraheme cytochrome direct electron donor to the photochemical pigments, and the light harvesting complex alpha and beta subunits, respectively[11]. No *puhA* was observed indicating that the reaction center of WPS-2 might lack an H subunit in a manner similar to the CfRC.

Sequence and structure prediction of the L and M subunits revealed a large number of unique characteristics (Figure 4), which is consistent with the level of distinctness seen in the phylogenetic analysis. In Proteobacteria and Cyanobacteria each Type II reaction center protein is made of 5 transmembrane helices each. The recent Cryo-EM structure of the reaction center from *Roseiflexus castenholzii* resolved a 6th helix in the L subunit[15]. This was located at the N-terminus and cannot be detected through secondary structure prediction methods. In the case of WPS-2 L subunit, transmembrane helix prediction only detected the standard 5 helices; instead a novel N-terminal 6th helix was detected in the M subunit (Figure 4 A).

Most of the ligands to the photochemical pigments and cofactor are conserved, but unique changes that might modulate the energetics of electron transfer can be noted. For example, in the PbRC both “monomeric” bacteriochlorophylls, Bch_L and Bch_M, are coordinated by histidine ligands (Figure 4 B and C). In the CfRC, a “third” bacteriopheophytin[15,63] occupies the position of Bch_M and accordingly there is no histidine ligand, and instead an isoleucine is found. In WPS-2 a histidine ligand to Bch_M is also absent like in the CfRC, however instead of isoleucine and strictly conserved aspartate (M-D169) is found, which may suggests the presence of bacteriochlorophyll rather than bacteriopheophytin at this position, but with modified energetics[64]. This M-D169 might play an important role in the inactivation of electron transfer via the M branch.

In the PbRC, L-H177 provides a hydrogen bond to the photochemical pigment P_L (Figure 4 D). Mutagenesis studies have shown that the strength or absence of this bond modulates the midpoint potential of the primary donor pigments P_L and P_M[65]. In the CfRC, the histidine is substituted by phenylalanine, which

should contribute to the lower oxidizing potential of P in comparison to the PbRC. In the CfRC the midpoint potential of the primary donor is +360-390 mV[66-68] while in the PbRC is +450-500 mV[69-72]. In the WPS-2 reaction center, a tyrosine is found at this position and it is potentially oriented towards Bch_M effectively breaking the hydrogen-bond to P_L like in the CfRC. The absence of the hydrogen-bond to P_L could lower the midpoint potential of the primary donor by about -95 mV[65], which may suggests that WPS-2 has a primary donor with a potential more similar to that of the CfRC.

In the PbRC, a carotenoid is found in contact with Bch_M and it is thought that this protects the system against the formation of reactive oxygen species by quenching bacteriochlorophyll triplets[73], see Figure 4 F. In the CfRC, no carotenoid has been reported at this position and the structure from *Roseiflexus* did not show a bound carotenoid[15]. In the CfRC and the WPS-2 reaction center three glycine residues, which in the PbRC give space to the carotenoid, are replaced by bulkier residues that should hinder the binding of a potential carotenoid.

In the PbRC and the CfRC a strictly conserved tyrosine (L-Y162) is located in between heme 3 (c-559) of PufC and the special pair (Fig. 5 E). Mutagenesis studies have shown that this tyrosine modulates the midpoint potential of both the heme and the special pair, but it is not required for fast electron transfer[74]. While a gene was identified for PufC in WPS-2[11], this tyrosine was found to be a phenylalanine in all WPS-2 L sequences. A deprotonation of this conserved tyrosine residue stabilizes the primary charge separation reactions[75], therefore this mechanism is not operational in the reaction center of WPS-2.

Conclusions

Bacterial genomes previously assigned to WPS-2 have been assigned to a candidate phylum named *Candidatus* Eremiobacterota[42]. However, phylogenetic analysis of the genomes previously assigned to *Ca.* Eremiobacterota and additional genomes described here via concatenated ribosomal protein phylogenies (Figure 1, Figure 3) and taxonomic assignment with GTDB-Tk[27] (Supplemental Table S1) suggest that the broad WPS-2 clade encompasses at least three distinct phyla. For the WPS-2 clade that includes all aerobic and phototrophic lineages described here we propose the candidate phylum *Candidatus* Palusbacterota, from the Latin “palus” for bog or swamp, reflecting the typical environment for known members of this clade.

Most lineages of anoxygenic phototrophs can be preferentially found in characteristic environments, such as anoxic regions of stratified water columns for Chlorobi[76], low-oxygen hot springs for Chloracidobacteria and phototrophic Chloroflexi[8,9,77-79], and soils for Heliobacteria[80], though exceptions do occur (e.g. Chloroflexi in carbonate tidal flats[81] and Heliobacteria in hot springs[82] or soda lakes[83,84]). In contrast, the phototrophic Proteobacteria and Gemmatimonadetes appear to have a more cosmopolitan distribution, including freshwater, marine, and soil environments[85,86]. The preferred environment for *Ca.* Palusbacterota appears to be cold, acidic, and aerobic environments with at least some members of the group forming a close association with plants, particularly mosses. This niche overlaps with that of plant-associated phototrophic Proteobacteria[87], and may be due to the relatively high oxygen tolerance of these phototrophic lineages as compared to typically more oxygen-sensitive phototrophs such as Heliobacteria and Chlorobi. The apparent close and specific association of phototrophic *Ca.* Palusbacterota with plants may reflect a long-term evolutionary association, which could further suggest that this group has radiated alongside plants over a timescale of <0.5 billion years, though this hypothesis will require molecular clocks or other analyses to test.

The majority of *Ca.* Palusbacterota genomes described here (~85% of all genomes, and ~97% of >80% completeness) encode aerobic respiration using closely related A-family heme-copper oxidoreductases and *bc* complex III proteins (Supplemental Figures S3 and S4). This is in contrast to the closely related phylum *Ca.* Eremiobacter which appears to be composed of obligate anaerobes, with no respiration genes encoded in any of the available genomes in this phylum, and aerobic Chloroflexi which do not encode closely related complex III or complex IV proteins. This, together with the broad congruence of *Ca.* Palusbacterota organismal phylogenies with complex III and complex IV protein phylogenies, implies that aerobic respiration was acquired via horizontal gene transfer into stem group *Ca.* Palusbacterota after their divergence from *Ca.* Eremiobacterota but before the radiation of crown group Palusbacterota. It is likely that anaerobic ancestors of *Ca.* Palusbacterota and related phyla diverged during Archean time, with the radiation of crown group *Ca.* Palusbacterota occurring after the GOE ~2.3 billion years ago which led to the expansion of aerobic metabolisms[51]. This hypothesis will require molecular clock estimates for the divergence and radiation of *Ca.* Palusbacterota or other tests to verify, but similar evolutionary trends have been seen in other groups, including the Chloroflexi and Cyanobacteria[50,51,88]. Following the acquisition of aerobic respiration by the last common ancestor of crown group *Ca.* Palusbacterota, aerobic respiration via an A-family HCO and a *bc* complex III appears to have been largely vertically inherited, with additional respiratory proteins (e.g. second copies of *bc* complex and A-family HCOs, and the B-family HCOs associated with phototrophy) acquired via later HGT.

It is well established that all reaction centers originated as homodimers, with electron transfer occurring symmetrically on both side of the reaction center as it is still seen in Type I reaction centers[89]. It is thought that heterodimeric Type II reaction centers evolved from two independent duplication events, one leading to L and M in anoxygenic Type II and another leading to D1 and D2 in cyanobacterial PSII[12,13,90]. The heterodimerization process led to the evolution of asymmetric electron transfer, which occurs exclusively via the L branch in anoxygenic Type II and via D1 in PSII, while the M and D2 branches became inactive, respectively. That the three different anoxygenic Type II reaction centers show distinct mechanisms of redox tuning on key positions around Bch_M and other photochemical pigments could indicate that these lineages radiated soon after the L and M duplication. Furthermore, given that anoxygenic Type II reaction centers show up to five times faster rates of evolution than cyanobacterial PSII[30] and that the emergence of at least some of major clades of phototrophs with anoxygenic Type II reaction centers postdate the GEO[50], it is possible that the duplication of L and M subunits, and the radiation of the known anoxygenic Type II reaction centers occurred after the origin of oxygenic photosynthesis[30]. Therefore, the idea that extant forms of anoxygenic phototrophy powered by heterodimeric Type II reaction centers represent a “primitive” form of photosynthesis is not supported by the available data. Instead, it should be considered a relatively recent, highly specialized, and mobile evolutionary innovation within the larger context of the evolution of anoxygenic photosynthesis, which reaches into Paleoarchean time[91,92]. If so it would imply that anoxygenic photosynthesis early in Earth’s history may have been driven by now-extinct groups of bacteria using variations on phototropic pathways not yet observed today[93].

The known phylogenetic diversity of anoxygenic phototrophs has increased substantially in recent years, thanks largely to genome-resolved metagenomic sequencing of diverse environments[8,9,11]. Better coverage of the extant diversity of phototrophs and their relatives allows us to better query the evolutionary history of both organisms and their metabolisms by providing opportunities for comparative genomic, phylogenetic, and molecular clock analyses. Previous attempts to interpret the early evolution of photosynthesis

relied on the narrower set of phototrophic lineages that were known at the time (e.g.,[94,95], whose analysis predated the discovery of phototrophic Chloracidobacteria, Gemmatimonadetes, Thermofonsia, and Palusbacterota). The discovery of phototrophic Palusbacterota and other lineages will provide crucial data for future reconsideration of the evolutionary relationships and history of phototrophy, improving our understanding of one of Earth's most important metabolic pathways.

Competing interests

The authors declare no competing interests.

Author's contributions

L.M.W., T.C. and H.H.-M. designed the study, analyzed the data, and wrote the manuscript.

Acknowledgements

T.C. is supported by a Leverhulme Trust grant (RPG-2017-223). L.M.W. is supported by an Agouron Institute postdoctoral fellowship. H. H. -M. is supported by NSF Dimensions of Biodiversity grant (DEB 1542609).

References

1 Raven, J. A. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56**, 177-192, doi:10.3354/ame01315 (2009).

2 Kharecha, P., Kasting, J. & Siefert, J. A coupled atmosphere-ecosystem model of the early Archean Earth. *Geobiology* **3**, 53-76, doi:10.1111/j.1472-4669.2005.00049.x (2005).

3 Canfield, D. E., Rosing, M. T. & Bjerrum, C. Early anaerobic metabolisms. *Philos. Trans. Royal Soc. B* **361**, 1819-1834; discussion 1835-1816, doi:10.1098/rstb.2006.1906 (2006).

4 Ward, L. M., Rasmussen, B. & Fischer, W. W. Primary productivity was limited by electron donors prior to the advent of oxygenic photosynthesis. *J. Geophys. Res. (G Biogeosci.)*, doi:10.1029/2018JG004679 (2019).

5 Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515-548, doi:10.1146/annurev-arplant-042110-103811 (2011).

6 Thiel, V., Tank, M. & Bryant, D. A. Diversity of chlorophototrophic bacteria revealed in the omics era. *Annu. Rev. Plant Biol.* **69**, 21-49, doi:10.1146/annurev-arplant-042817-040500 (2018).

7 Gest, H. & Blankenship, R. E. in *Discoveries in Photosynthesis* (eds Govindjee, J. Thomas Beatty, Howard Gest, & John F. Allen) 51-62 (Springer Netherlands, 2005).

8 Bryant, D. A. *et al.* Candidatus Chloracidobacterium thermophilum: An aerobic phototrophic acidobacterium. *Science* **317**, 523-526, doi:10.1126/science.1143236 (2007).

9 Ward, L. M., Hemp, J., Shih, P. M., McGlynn, S. E. & Fischer, W. W. Evolution of phototrophy in the Chloroflexi phylum driven by horizontal gene transfer. *Front. Microbiol.* **9**, 260, doi:10.3389/fmicb.2018.00260 (2018).

10 Zeng, Y. H., Feng, F. Y., Medova, H., Dean, J. & Koblizek, M. Functional Type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7795-7800, doi:10.1073/pnas.1400295111 (2014).

11 Holland-Moritz, H. *et al.* Novel bacterial lineages associated with boreal moss species. *Environ. Microbiol.* **20**, 2625-2638, doi:10.1111/1462-2920.14288 (2018).

12 Beanland, T. J. Evolutionary relationships between Q-Type photosynthetic reaction centers - Hypothesis-testing using parsimony. *J. Theor. Biol.* **145**, 535-545, doi:10.1016/S0022-5193(05)80487-4 (1990).

13 Cardona, T. A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynth. Res.* **126**, 111-134, doi:10.1007/s11120-014-0065-x (2015).

14 Yu, L. J., Suga, M., Wang-Otomo, Z. Y. & Shen, J. R. Structure of photosynthetic LH1-RC supercomplex at 1.9 Å resolution. *Nature* **556**, 209-213, doi:10.1038/s41586-018-0002-9 (2018).

15 Xin, Y. Y. *et al.* Cryo-EM structure of the RC-LH core complex from an early branching photosynthetic prokaryote. *Nat. Commun.* **9**, 1568, doi:10.1038/s41467-018-03881-x (2018).

16 Pierson, B. K. & Castenholz, R. W. Bacteriochlorophylls in gliding filamentous prokaryotes from Hot Springs. *Nature-New Biol* **233**, 25-27, doi:10.1038/newbio233025a0 (1971).

17 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

18 Camacho, C. *et al.* BLAST+: architecture and applications. *Bmc Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

19 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

20 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).

21 Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533-1542, doi:10.1038/s41564-017-0012-7 (2017).

22 Ward, L. M., Shih, P. M. & Fischer, W. W. MetaPOAP: presence or absence of metabolic pathways in metagenome-assembled genomes. *Bioinformatics* **34**, 4284-4289, doi:10.1093/bioinformatics/bty510 (2018).

23 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

24 Miller, M. A., Pfeiffer, W. & Schwartz, T. in *2010 Gateway Computing Environments Workshop (GCE).* 1-8.

25 Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452-456, doi:10.1038/s41586-018-0043-0 (2018).

26 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-W245, doi:10.1093/nar/gkw290 (2016).

27 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004 doi:10.1038/nbt.4229 (2018).

28 Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 16048, doi:10.1038/nmicrobiol.2016.48 (2016).

29 Doolittle, R. F. *Of urfs and orfs: A primer on how to analyze derived amino acid sequences.* (University Science Books, 1986).

30 Cardona, T., Sanchez-Baracaldo, P., Rutherford, A. W. & Larkum, A. W. D. Early Archean origin of Photosystem II. *Geobiology*, doi:10.1111/gbi.12322 (2018).

31 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539, doi:10.1038/msb.2011.75 (2011).

32 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

33 Lefort, V., Longueville, J. E. & Gascuel, O. SMS: Smart model selection in PhyML. *Mol. Biol. Evol.* **34**, 2422-2424, doi:10.1093/molbev/msx149 (2017).

34 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567-580, doi:10.1006/jmbi.2000.4315 (2001).

35 Hessa, T. *et al.* Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030, doi:10.1038/Nature06387 (2007).

36 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

37 Nogales, B. *et al.* Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl. Environ. Microbiol.* **67**, 1874-1884, doi:10.1128/Aem.67.4.1874-1884.2001 (2001).

38 Brantner, J. S. *et al.* Depth-dependent geochemical and microbiological gradients in Fe(III) deposits resulting from coal mine-derived acid mine drainage. *Front. Microbiol.* **5**, 215, doi:10.3389/fmicb.2014.00215 (2014).

39 Trexler, R. *et al.* Assessing impacts of unconventional natural gas extraction on microbial communities in headwater stream ecosystems in Northwestern Pennsylvania. *Front. Microbiol.* **5**, 522, doi:10.3389/fmicb.2014.00522 (2014).

40 Costello, E. K. *Molecular phylogenetic characterization of high altitude soil microbial communities and novel, uncultivated bacterial lineages*, University of Colorado, (2007).

41 Grasby, S. E. *et al*. The Paint Pots, Kootenay National Park, Canada - a natural acid spring analogue for Mars. *Can. J. Earth Sci./Rev. Can. Sci. Terre* **50**, 94-108, doi:10.1139/e2012-060 (2013).

42 Ji, M. *et al*. Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**, 400-403, doi:10.1038/nature25014 (2017).

43 Ji, M. *et al*. Microbial diversity at Mitchell Peninsula, Eastern Antarctica: a potential biodiversity "hotspot". *Polar Biol.* **39**, 237-249, doi:10.1007/s00300-015-1776-y (2016).

44 Stibal, M. *et al.* Different bulk and active bacterial communities in cryoconite from the margin and interior of the Greenland ice sheet. *Environ. Microbiol. Rep.* **7**, 293-300, doi:10.1111/1758-2229.12246 (2015).

45 Camanocha, A. & Dewhirst, F. E. Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/candidate divisions. *J. Oral Microbiol.* **6**, 25468, doi:10.3402/jom.v6.25468 (2014).

46 Serkebaeva, Y. M., Kim, Y., Liesack, W. & Dedysh, S. N. Pyrosequencing-based assessment of the bacteria diversity in surface and subsurface peat layers of a Northern Wetland, with focus on poorly studied phyla and candidate divisions. *PloS one* **8**, e63994, doi:10.1371/journal.pone.0063994 (2013).

47 Bragina, A., Berg, C. & Berg, G. The core microbiome bonds the Alpine bog vegetation to a transkingdom metacommunity. *Mol. Ecol.* **24**, 4795-4807, doi:10.1111/mec.13342 (2015).

48 Bertin, P. N. *et al.* Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J.* **5**, 1735-1747, doi:10.1038/ismej.2011.51 (2011).

49 Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49-54, doi:10.1038/s41586-018-0338-1 (2018).

50 Shih, P. M., Ward, L. M. & Fischer, W. W. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10749-10754, doi:10.1073/pnas.1710798114 (2017).

51 Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647-683, doi:10.1146/annurev-earth-060313-054810 (2016).

52 Brinkmann, H., Goker, M., Koblizek, M., Wagner-Dobler, I. & Petersen, J. Horizontal operon transfer, plasmids, and the evolution of photosynthesis in Rhodobacteraceae. *ISME J.* **12**, 1994-2010, doi:10.1038/s41396-018-0150-9 (2018).

53 Bryant, D. *et al.* in *Functional Genomics and Evolution of Photosynthetic Systems* Vol. 33 *Advances in Photosynthesis and Respiration* (eds R. L. Burnap & W. Vermaas) 47-102 (Springer, 2012).

54 Sousa, F. L., Shavit-Grievink, L., Allen, J. F. & Martin, W. F. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol. Evol.* **5**, 200-216, doi:10.1093/Gbe/Evs127 (2013).

55 Costas, A. M. G. *et al.* Ultrastructural analysis and identification of envelope proteins of "*Candidatus* Chloracidobacterium thermophilum" chlorosomes. *J. Bacteriol.* **193**, 6701-6711, doi:10.1128/Jb.06124-11 (2011).

56 Hanada, S., Takaichi, S., Matsuura, K. & Nakamura, K. *Roseiflexus castenholzii* gen. nov., sp nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *Int. J. Syst. Evol. Microbiol.* **52**, 187-193, doi:10.1099/00207713-52-1-187 (2002).

57 Han, H. *et al.* Adaptation of aerobic respiration to low O_2 environments. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14109-14114, doi:10.1073/pnas.1018958108 (2011).

58 Sutcliffe, I. C. A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol.* **18**, 464-470, doi:10.1016/j.tim.2010.06.005 (2010).

59 Sutcliffe, I. C. Cell envelope architecture in the Chloroflexi: a shifting frontline in a phylogenetic turf war. *Environ. Microbiol.* **13**, 279-282, doi:10.1111/j.1462-2920.2010.02339.x (2011).

60 Ward, L. M., McGlynn, S. E. & Fischer, W. W. Draft genome sequences of a novel lineage of *Armatimonadetes* recovered from Japanese hot springs. *Genome Announc.* **5**, e00820-00817, doi:10.1128/genomeA.00820-17 (2017).

61 Ward, L. M., McGlynn, S. E. & Fischer, W. W. Draft genome sequence of a divergent anaerobic member of the *Chloroflexi* class *Ardenticatenia* from a sulfidic hot spring. *Genome Announc.* **6**, e00571-00518, doi:10.1128/genomeA.00571-18 (2018).

62 Ward, L. M., McGlynn, S. E. & Fischer, W. W. Draft genome sequences of two basal members of the *Anaerolineae* class of *Chloroflexi* from a sulfidic hot spring. *Genome Announc.* **6**, e00570-00518, doi:10.1128/genomeA.00570-18 (2018).

63 Pierson, B. K., Thornber, J. P. & Seftor, R. E. B. Partial purification, subunit structure and thermal stability of the photochemical reaction center of the thermophilic green bacterium *Chloroflexus aurantiacus*. *Biochim. Biophys. Acta* **723**, 322-326, doi:10.1016/0005-2728(83)90132-9 (1983).

64 Heimdal, J., Jensen, K. P., Devarajan, A. & Ryde, U. The role of axial ligands for the structure and function of chlorophylls. *J. Biol. Inorg. Chem.* **12**, 49-61, doi:10.1007/s00775-006-0164-z (2007).

65 Lin, X. *et al.* Specific alteration of the oxidation potential of the electron donor in reaction centers from *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10265-10269, doi:10.1073/pnas.91.22.10265 (1994).

66 Bruce, B. D., Fuller, R. C. & Blankenship, R. E. Primary photochemistry in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6532-6536, doi:10.1073/pnas.79.21.6532 (1982).

67 Ivancich, A., Feick, R., Ertlmaier, A. & Mattioli, T. A. Structure and protein binding interactions of the primary donor of the *Chloroflexus aurantiacus* reaction center. *Biochemistry* **35**, 6126-6135, doi:10.1021/Bi952772r (1996).

68 Collins, A. M., Kirmaier, C., Holten, D. & Blankenship, R. E. Kinetics and energetics of electron transfer in reaction centers of the photosynthetic bacterium Roseiflexus castenholzii. *Biochim. Biophys. Acta* **1807**, 262-269, doi:10.1016/j.bbabio.2010.11.011 (2011).

69 Moss, D. A., Leonhard, M., Bauscher, M. & Mantele, W. Electrochemical redox titration of cofactors in the reaction center from *Rhodobacter sphaeroides*. *FEBS Lett.* **283**, 33-36, doi:10.1016/0014-5793(91)80547-G (1991).

70 Williams, J. C. *et al.* Effects of mutations near the bacteriochlorophylls in reaction centers from *Rhodobacter sphaeroides*. *Biochemistry* **31**, 11029-11037, doi:10.1021/bi00160a012 (1992).

71 Woodbury, N. W. & Allen, J. P. in *Anoxygenic photosynthetic bacteria: advances in photosynthesis* (eds R. E. Blankenship, M. T. Madigan, & C. E. Bauer) 527-557 (Kluwer Academic Publishers, 1995).

72 Jones, M. R. The petite purple photosynthetic powerpack. *Biochem. Soc. Trans.* **37**, 400-407, doi:10.1042/Bst0370400 (2009).

73 Cogdell, R. J. *et al.* How carotenoids protect bacterial photosynthesis. *Philos. Trans. Royal Soc. B* **355**, 1345-1349, doi:10.1098/rstb.2000.0696 (2000).

74 Dohse, B. *et al.* Electron transfer from the tetraheme cytochrome to the special pair in the *Rhodopseudomonas viridis* reaction center: Effect of mutations of tyrosine L162. *Biochemistry* **34**, 11335-11343, doi:10.1021/bi00036a006 (1995).

75 Wohri, A. B. *et al.* Light-induced structural changes in a photosynthetic reaction center caught by Laue diffraction. *Science* **328**, 630-633, doi:10.1126/science.1186159 (2010).

76 Imhoff, J. F. in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (eds Eugene Rosenberg *et al.*) 501-514 (Springer Berlin Heidelberg, 2014).

77 Hallenbeck, P. C., Grogger, M., Mraz, M. & Veverka, D. Draft genome sequence of the photoheterotrophic *Chloracidobacterium thermophilum* strain OC1 found in a Mat at Ojo Caliente. *Genome Announc.* **4**, e01570-01515, doi:10.1128/genomeA.01570-15 (2016).

78 Klatt, C. G. *et al.* Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J.* **5**, 1262-1278, doi:10.1038/ismej.2011.73 (2011).

79 Ward, L. M., McGlynn, S. E. & Fischer, W. W. Draft genome sequence of *Chloracidobacterium* sp. CP2_5A, a phototrophic member of the phylum *Acidobacteria* recovered from a Japanese hot spring. *Genome Announc.* **5**, e00821-00817, doi:10.1128/genomeA.00821-17 (2017).

80 Madigan, M. T. & Ormerod, J. G. in *Anoxygenic Photosynthetic Bacteria* (eds Robert E. Blankenship, Michael T. Madigan, & Carl E. Bauer) 17-30 (Springer Netherlands, 1995).

81 Trembath-Reichert, E. *et al.* Gene sequencing-based analysis of microbial-mat morphotypes, Caicos Platform, British West Indies. *J. Sediment. Res.* **86**, 629-636, doi:10.2110/jsr.2016.40 (2016).

82 Kimble, L. K., Mandelco, L., Woese, C. R. & Madigan, M. T. *Heliobacterium modesticaldum*, sp. nov., a thermophilic heliobacterium of hot springs and volcanic soils. *Arch. Microbiol.* **163**, 259-267, doi:10.1007/bf00393378 (1995).

83 Bryantseva, I. A., Gorlenko, V. M., Kompantseva, E. I., Achenbach, L. A. & Madigan, M. T. Heliorestis daurensis, gen. nov. sp. nov., an alkaliphilic rod-to-coiled-shaped phototrophic heliobacterium from a Siberian soda lake. *Arch. Microbiol.* **172**, 167-174, doi:10.1007/s002030050756 (1999).

84 Asao, M., Jung, D. O., Achenbach, L. A. & Madigan, M. T. Heliorestis convoluta sp. nov., a coiled, alkaliphilic heliobacterium from the Wadi El Natroun, Egypt. *Extremophiles* **10**, 403-410, doi:10.1007/s00792-006-0513-4 (2006).

85 Imhoff, J. F. in *Modern topics in the phototrophic prokaryotes: Environmental and applied aspects* (ed Patrick C. Hallenbeck) 47-85 (Springer International Publishing, 2017).

86 Zeng, Y. *et al.* Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environ. Microbiol. Rep.* **8**, 139-149, doi:10.1111/1758-2229.12363 (2016).

87 Atamna-Ismaeel, N. *et al.* Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environ. Microbiol. Rep.* **4**, 209-216, doi:10.1111/j.1758-2229.2011.00323.x (2012).

88 Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J. & Fischer, W. W. Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* **15**, 19-29, doi:10.1111/gbi.12200 (2017).

89 Hohmann-Marriott, M. F. & Blankenship, R. E. in *Photosynthetic Protein Complexes* (ed P Fromme) 295-324 (Wiley-VCH Verlag GmbH & Co. KGaA, 2008).
90 Sadekar, S., Raymond, J. & Blankenship, R. E. Conservation of distantly related membrane proteins: Photosynthetic reaction centers share a common structural core. *Mol. Biol. Evol.* **23**, 2001-2007, doi:10.1093/molbev/msl079 (2006).
91 Tice, M. M. & Lowe, D. R. Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* **431**, 549-552, doi:10.1038/nature02888 (2004).
92 Tice, M. M. & Lowe, D. R. Hydrogen-based carbon fixation in the earliest known photosynthetic organisms. *Geology* **34**, 37-40, doi:10.1130/G22012.1 (2006).
93 Ward, L. M. & Shih, P. M. The evolution and productivity of carbon fixation pathways in response to changes in oxygen concentration over geological time. *Free Radical Biol. Med.*, doi:Uder review (2019).
94 Xiong, J., Fischer, W. M., Inoue, K., Nakahara, M. & Bauer, C. E. Molecular evidence for the early evolution of photosynthesis. *Science* **289**, 1724-1730, doi:10.1126/science.289.5485.1724 (2000).
95 Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y. & Blankenship, R. E. Whole-genome analysis of photosynthetic prokaryotes. *Science* **298**, 1616-1620, doi:10.1126/science.1075558 (2002).

Figures

Figure 1. (A) Phylogenetic tree of bacteria showing position of WPS-2, based on concatenated ribosomal protein sequences following Hug et al. 2016, with the distribution of photosynthetic reaction centers color-coded. In case where only some members of a phylum are capable of phototrophy (e.g. Heliobacteria in the much larger Firmicutes phylum), the entire phylum is highlighted for clarity. (B) Zoomed-in relationships of the WPS-2 clade (here subdivided into the candidate phyla *Candidatus* Eremiobacterota, QHBK01, *Ca*. Palusbacterota), Chloroflexi, and other closely related phyla.

Figure 2. Phylogeny of Type II reaction center proteins. (A) A Tree calculated using only L and M subunits including only anoxygenic phototrophs. (B) A Tree including D1 and D2 proteins of Cyanobacteria.

Figure 3. Phylogeny of WPS-2 built with concatenated ribosomal protein sequences, annotated with distribution of traits including phototrophy and aerobic respiration. Green branches encode both phototrophy and aerobic respiration; blue lineages encode aerobic respiration but not phototrophy; yellow lineages encode phototrophy but not aerobic respiration; black lineages encode neither.

Figure 4. Structural comparisons of Type II reaction centers. (A) Schematic representation of the transmembrane helices. (B) The crystal structure of the PbRC from *Thermocromatium tepidum* highlighting all redox cofactors. (C) Comparison in the ligands to Bch_L and Bch_M. Residues in grey are those from the PbRC, in blue those from the Cryo-EM structure of *Roseiflexus*, and in orange those from the homology model of the WPS-2 reaction center proteins. (D) Changes around hydrogen-bonding partners to P_L. (E) Changes around L-Y161, a residue important for electron transfer from the tetraheme cytochrome into the oxidized photochemical pigments.

Supplementary figures

Figure S1. Protein phylogeny of PufL and PufM subunits of the Type II reaction center of WPS-2 bacteria. Protein sequences that were incomplete due to metagenome assembly (<80% of full length) omitted.

Figure S2. Protein phylogeny of BchL proteins from phototrophic bacteria. Phyla labeled, and WPS-2 highlighted in red. Protein sequences that were incomplete due to metagenome assembly (<80% of full length) omitted.

Figure S3. Protein phylogeny of A- and B-family heme-copper oxidoreductases from WPS-2, Chloroflexi, and other related phyla. WPS-2 highlighted in red. Protein sequences that were incomplete due to metagenome assembly (<80% of full length) omitted.

Figure S4. Protein phylogeny of *bc* complex III from WPS-2, Chloroflexi, and other related phyla. WPS-2 highlighted in red. Protein sequences that were incomplete due to metagenome assembly (<80% of full length) omitted.

Figure S5. Protein phylogeny of the large subunit of rubisco from WPS-2, Chloroflexi, and other related phyla. Major clades outside of the WPS-2 collapsed for clarity. Protein sequences that were incomplete due to metagenome assembly (<80% of full length) omitted.

Figure S6. Protein phylogeny of reaction center proteins on a subset of 176 L and M sequences (A) and the same dataset but removing poorly aligned regions (B). Poorly aligned regions were removed with Gblocks allowing for smaller final blocks, gap positions within the final blocks, and less strict flanking positions.

Supplemental tables:

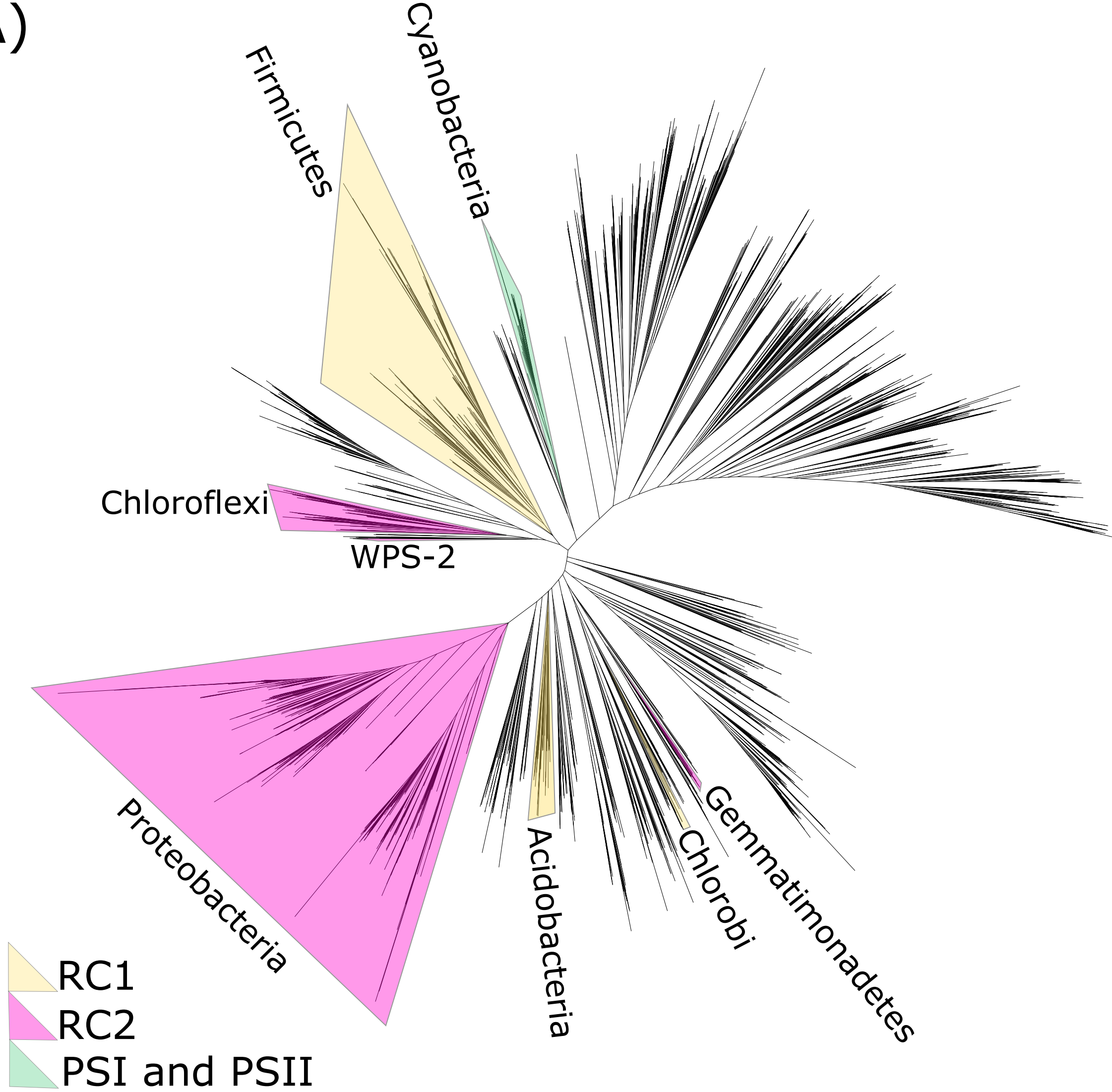
Table S1. Table of available WPS-2 genomes, including completeness, contamination, presence of respiration and phototrophy genes, candidate phylum assignment, metagenome source, and reference information.

Supplemental Discussion:

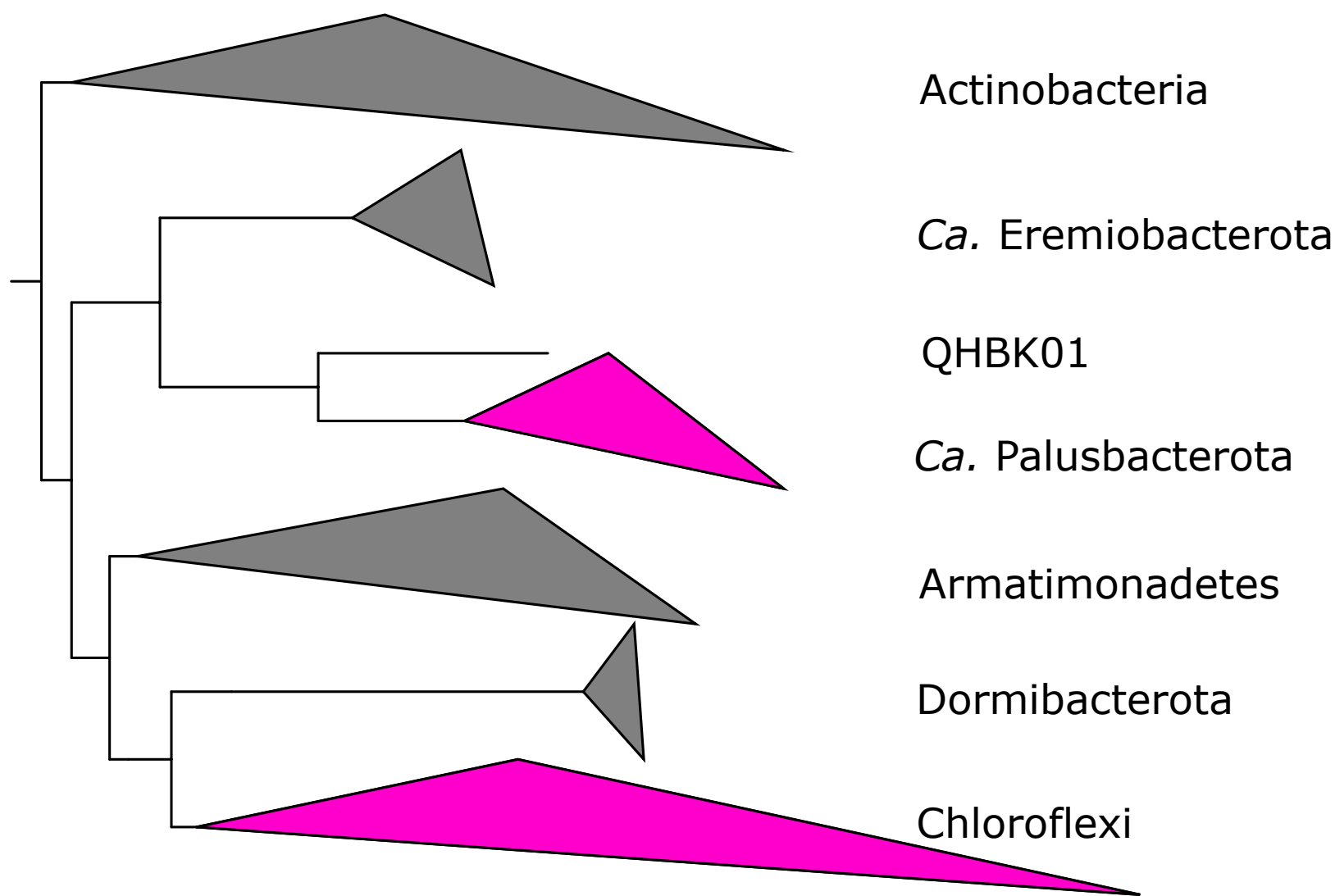
MetaPOAP False Negative estimates for the probability of phototrophy in each WPS-2 genome is found in Table 1; genomes in which some phototrophy proteins were recovered have a high probability of encoding the complete phototrophy pathway, whereas the probability that phototrophy is encoded in other genomes but was not recovered is very low, suggesting that the distribution of phototrophy shown in Figure 3 reflects the actual distribution of phototrophy in this phylum. Only the Bog1503/PKZK01 genome has a low (<0.01) MetaPOAP False Negative probability of encoding a full phototrophy pathway due to encoding only a subset of necessary proteins for phototrophy despite relatively high completeness; this organism may be an instance of transitional loss of phototrophy, but this will require more complete genome sequencing or physiological data to verify. While some *Ca.* Palusbacterota genomes did not recover O_2 reductases (PMTD01, PMUE01, PNBL01, and WPS2_33), these were primarily lower completeness (<80%), were closely related to aerobic strains, and did recover a *bc* complex, suggesting that these genomes either encode a respiratory electron transport chain that

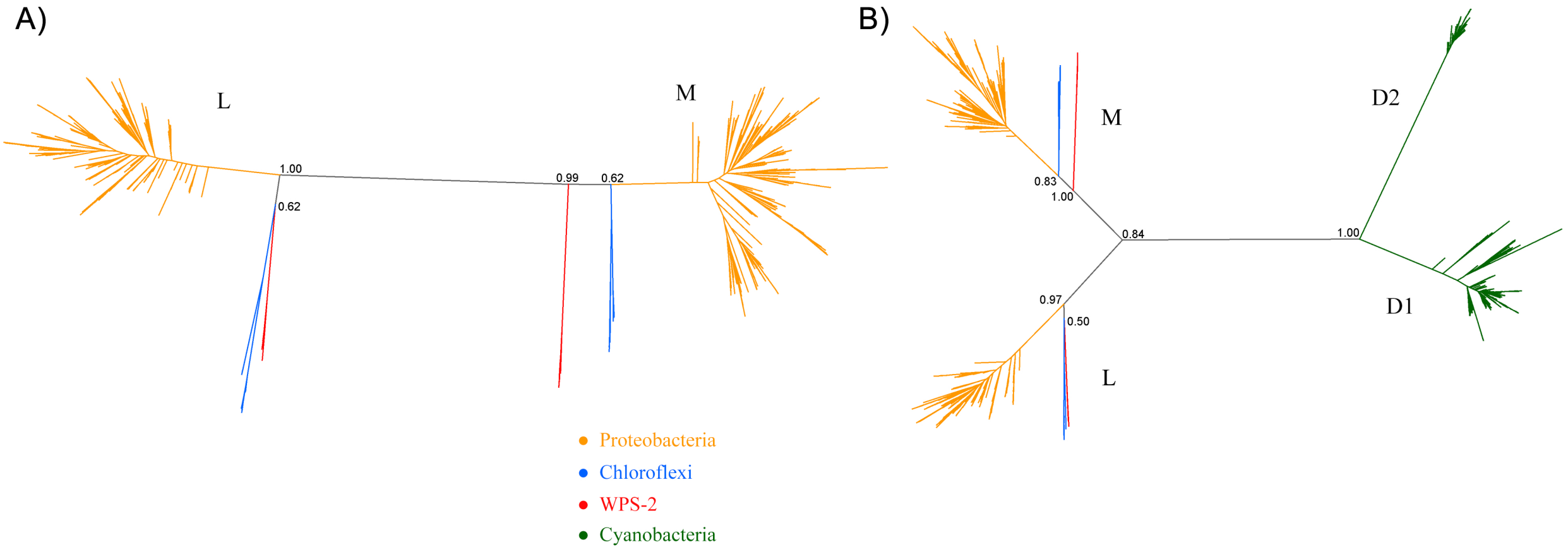
was not recovered in the MAG or that they are undergoing secondary loss of an ancestral respiratory metabolism.

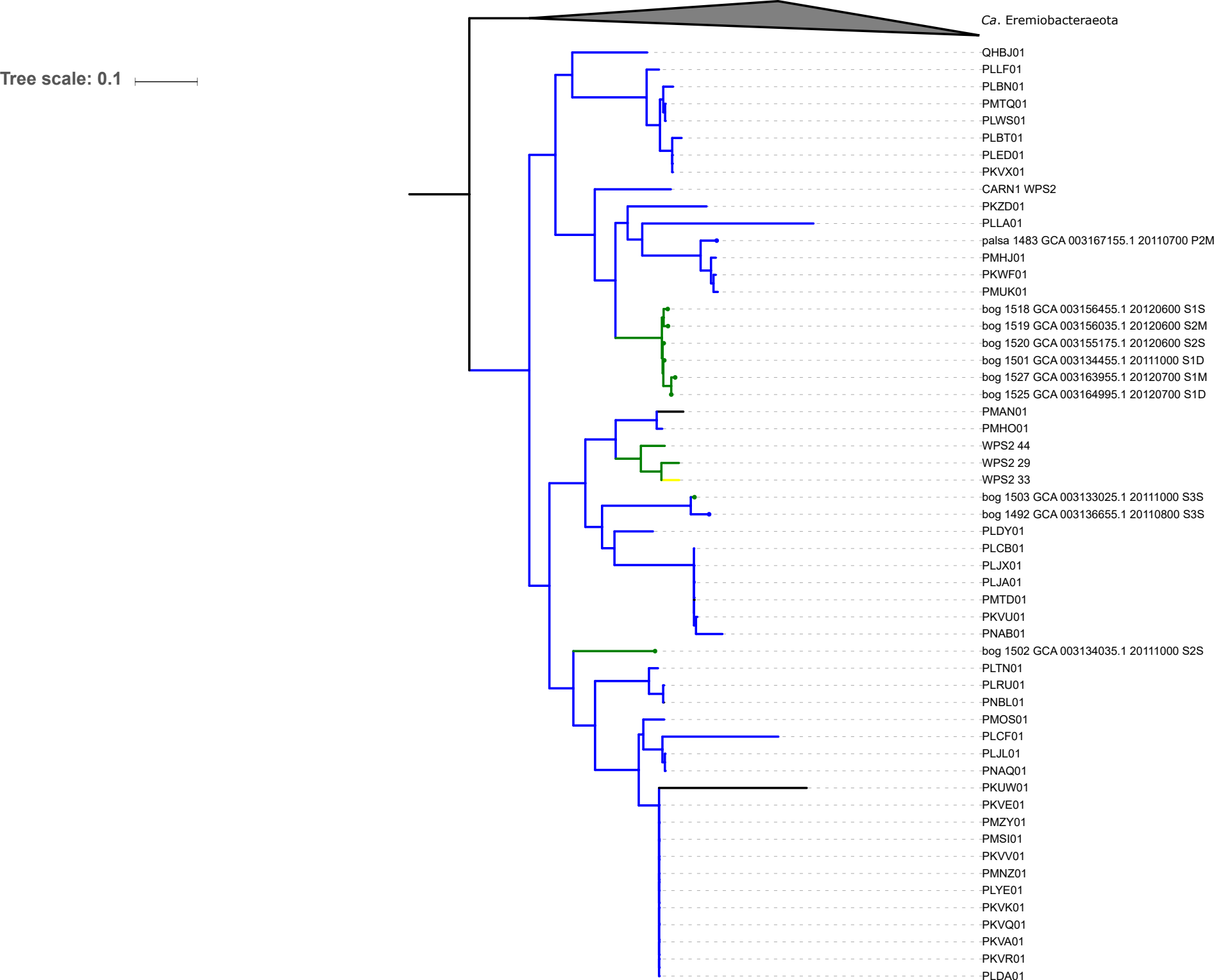
A)

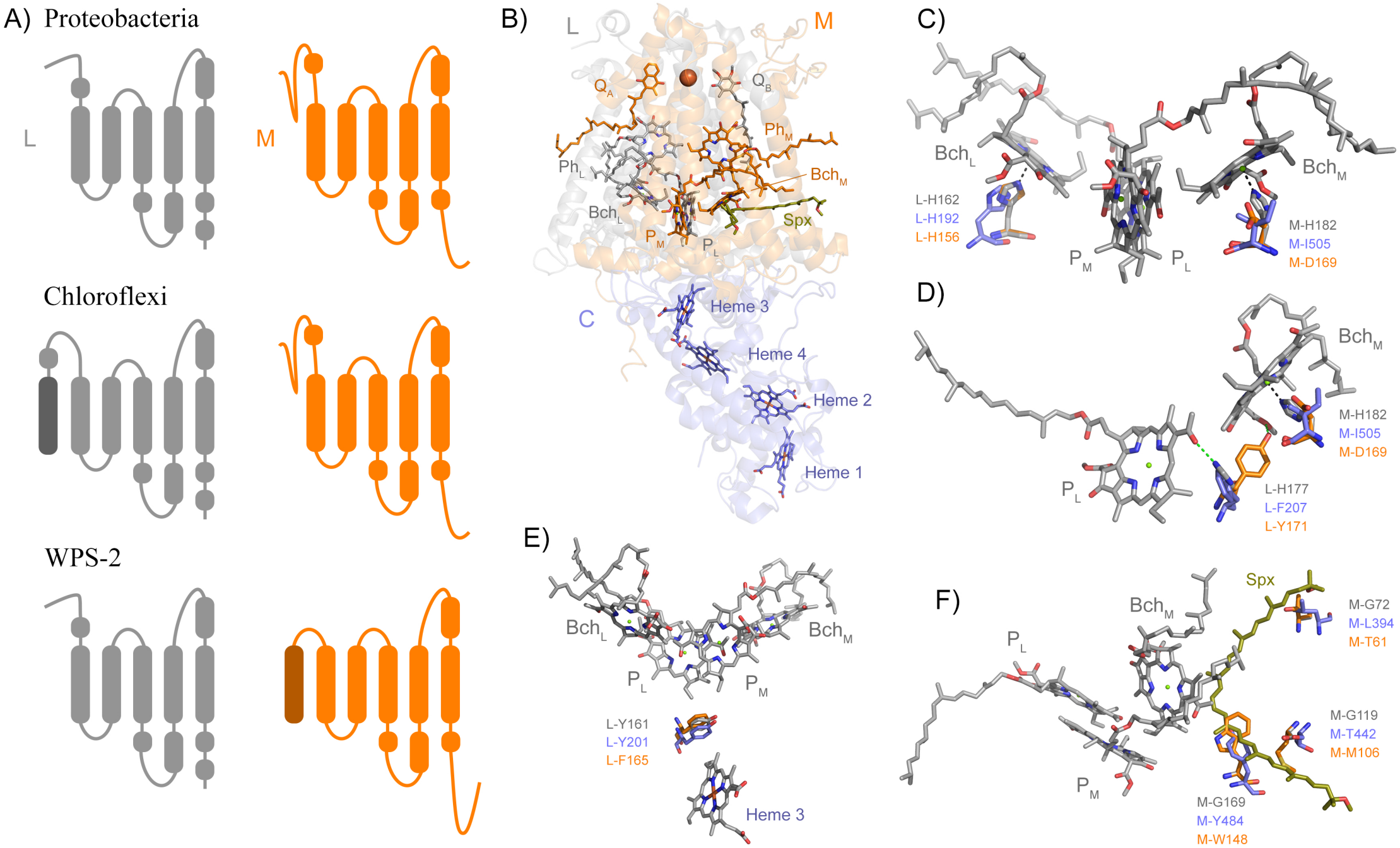

Firmicutes
Cyanobacteria
Chloroflexi
WPS-2
Proteobacteria
Acidobacteria
Chlorobi
Gemmatimonadetes
RC1
RC2
PSI and PSII

B)


Actinobacteria
Ca. Eremiobacterota
QHBK01
Ca. Palusbacterota
Armatimonadetes
Dormibacterota
Chloroflexi


A)
L
M
1.00
0.62
0.99
0.62
B)
M
D2
0.83
1.00
0.84
1.00
0.97
0.50
D1
L
Proteobacteria
Chloroflexi
WPS-2
Cyanobacteria


Tree scale: 0.1
Ca. Eremiobacteraeota
QHBJ01
PLLF01
PLBN01
PMTQ01
PLWS01
PLBT01
PLED01
PKVX01
CARN1 WPS2
PKZD01
PLLA01
palsa 1483 GCA 003167155.1 20110700 P2M
PMHJ01
PKWF01
PMUK01
bog 1518 GCA 003156455.1 20120600 S1S
bog 1519 GCA 003156035.1 20120600 S2M
bog 1520 GCA 003155175.1 20120600 S2S
bog 1501 GCA 003134455.1 20111000 S1D
bog 1527 GCA 003163955.1 20120700 S1M
bog 1525 GCA 003164995.1 20120700 S1D
PMAN01
PMHO01
WPS2 44
WPS2 29
WPS2 33
bog 1503 GCA 003133025.1 20111000 S3S
bog 1492 GCA 003136655.1 20110800 S3S
PLDY01
PLCB01
PLJX01
PLJA01
PMTD01
PKVU01
PNAB01
bog 1502 GCA 003134035.1 20111000 S2S
PLTN01
PLRU01
PNBL01
PMOS01
PLCF01
PLJL01
PNAQ01
PKUW01
PKVE01
PMZY01
PMSI01
PKVV01
PMNZ01
PLYE01
PKVK01
PKVQ01
PKVA01
PKVR01
PLDA01


A) Proteobacteria
L
M
Chloroflexi
WPS-2
B)
L
M
C
Q_A
Q_B
Ph_M
Ph_L
Bch_M
Bch_L
Spx
P_M
P_L
Heme 3
Heme 4
Heme 2
Heme 1
C)
Bch_L
Bch_M
L-H162
L-H192
L-H156
M-H182
M-I505
M-D169
P_M
P_L
D)
Bch_M
M-H182
M-I505
M-D169
P_L
L-H177
L-F207
L-Y171
E)
Bch_L
Bch_M
P_L
P_M
L-Y161
L-Y201
L-F165
Heme 3
F)
Spx
Bch_M
P_L
M-G72
M-L394
M-T61
M-G119
M-T442
M-M106
P_M
M-G169
M-Y484
M-W148