Phototrophicity and Genomic Composition in Plant-Associated Sphingomonas faeni Strains

Batuhan Dogan
University of Jyväskylä
Janne A. Ihalainen
janne.ihalainen@jyu.fi

University of Jyväskylä
Yonghui Zeng
University of Copenhagen
Riitta Nissinen
University of Turku

Research Article

Keywords:

Posted Date: March 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9081327/v1

Additional Declarations: No competing interests reported.

Abstract

Solar radiation impacts most life forms on Earth by acting as a energy source or a regulatory signal. However, relatively little is still known about phototrophic potential and strategies of environmental bacteria beyond Cyanobacteria. This study explores the phototrophy related genomic diversity of *Sphingomonas faeni* strains from Arctic and sub-Arctic regions. We analyzed the genomes of 25 plant-associated *S.faeni* strains isolated from *Vaccinium myrtillus*, *Oxyria digyna*, *Vaccinium vitis-idaea*, and *Bistorta vivipara*, along with a reference *S. faeni* genome MA-olki. The strains showed diversity both in overall genome level but also in phototrophic capabilities: Seven strains were identified as aerobic anoxygenic phototrophic (AAP) bacteria with a complete photosynthetic gene cluster, 16 strains contained Xanthorhodopsin (XR) genes, and three strains were non-phototrophic, possessing no AAP or XR genes. The AAP strains were found exclusive from *Vaccinium* hosts. *O. digyna* contained only XR containing strains and *B. vivipara* showed XR genes and one non-phototrophic strain. *V. vitis-idaea* hosted strains for all three different phototrophy categories. Phylogenetic analyses using 16S rRNA gene and whole genome alignments showed AAP positive strains forming a tight phylogenetic group. We found no strong evidence of horizontal gene transfer of photosynthetic gene cluster. XR strains and non-phototrophic strains clustered into three different subgroups. Phototrophic strains had more photoreceptors, and AAP strains exhibited dual copies of the 5-aminolevulinic acid (5-ALA) gene within the photosynthetic gene cluster (PGC). Our genomic analysis suggests a relationship between phototrophic strategies and host plant specificity.

2. Introduction

The Arctic and sub-Arctic regions represent environments where low temperatures, and variable light conditions pose challenges to all life forms [1]. In these environments, microorganisms have evolved a variety of adaptive strategies to exploit available resources and cope with environmental stressors. Enabling organisms to utilize light as an energy source may confer an advantage in harsh ecosystems [2]. Photosynthetic microorganisms like cyanobacteria and algae adapt to low temperatures and light levels by optimizing their membrane lipid compositions to maintain fluidity at low temperatures and by evolving specialized pigments to maximize light absorption even in low light conditions [2, 3]. These adaptations enable them to thrive in cold environments, where seasonally low light availability and nutrient scarcity would otherwise limit growth.

Anoxygenic phototrophic bacteria use bacteriochlorophyll (BChl) based reaction centers and light harvesting antennas to capture light [4]. Some of those require oxygen rich environments for growth and electron transport reactions [5]. These organisms are known as aerobic anoxygenic phototrophic bacteria (AAPB) and thrive in aerobic settings, using various organic compounds as electron donors in their phototrophic processes [6]. Initially identified in nutrient poor marine environments [7], AAPB are now known to be widespread in both marine and freshwater ecosystems [6, 8]. In these aquatic environments, they may represent up to one-third of all bacterial populations, playing an important role in the carbon cycling processes of aquatic systems [9, 10]. AAPB have also been observed in terrestrial

habitats such as soil biocrusts [11] and Antarctic soils [12]. Additionally, AAPB have been detected within the metagenomes of plant phyllospheres and are found in both epi- and endophytic niche communities of diverse plant species in Arctic and boreal zones [13–15]. The ecology of AAPB is not well understood. Kolber et al. [16], and Koblížek [6] speculated that the ability to utilize light may be especially beneficial in nutrient poor marine environments.

Retinal based phototrophy involves microorganisms that use rhodopsins to capture light energy. Proteorhodopsin genes, found across all three domains of life, are most extensively studied in marine bacteria, and they are present in approximately half of all surface ocean heterotrophic bacteria [17, 18]. Xanthorhodopsins enable microorganisms to utilize light as an energy source, thereby conferring an adaptive advantage in nutrient poor environments [19–21]. Structurally, XRs are transmembrane proteins with seven transmembrane helices, a configuration crucial for their function in creating a proton gradient across the membrane upon light exposure [22–24]. The retinal molecule, covalently bound to a lysine residue within the protein, undergoes isomerization upon light absorption, initiating a series of conformational changes that result in proton pumping across the membrane [19, 20, 22–26]. Hence, AAPB utilize BChl *a* to absorb blue and near-infrared light, while rhodopsins utilize green light for ATP synthesis. The biosynthesis of rhodopsin based proton pumps is energetically more frugal than the complex photosynthetic machinery required by bacteria [27]. Due to larger light harvesting capability the photosynthetic machinery can function in lower light regimes than the rhodopsin systems, however.

Naturally, phototrophic bacteria require detailed sensing of light and possess a set of photoreceptor proteins [28]. All photosensors operate through a similar mechanism: photon absorption by the chromophore induces physical changes in the chromophore and its surrounding environment. Flavin containing systems like Blue Light Using Flavin domain (BLUF) complex undergo charge transfer processes following photon absorption [29]. In contrast, bacteriophytochromes (BphP) utilize a photoinduced isomerization process of bilin chromophores [30, 31]. These chromophore changes lead to alterations in the protein environment of the photosensory domain. The biochemical signaling occurs through the effector domain of the complex, which often binds to or separates from its corresponding response regulator (RR), acting as a gene expression regulator. In summary, while the architecture of the photosensory domain is largely conserved across life forms, the effector domain exhibits significant variability [32]. This variability allows effector domains to control a wide range of functions through their interactions with their respective RRs or other suppressor/effector domains. Numerous studies detail the variability and functional mechanisms of microbial photosensory systems [30, 33–37].

Sphingomonas faeni is a Gram-negative, non-spore-forming, psychrotolerant bacterium that was first described by Busse et al. [38]. having been originally isolated from straw. This species is part of the genus *Sphingomonas*, which is characterized by its distinctive sphingoglycolipids, ubiquinone Q-10, and the presence of sym-homospermidine as the predominant polyamine [39, 40].

S. faeni has been repeatedly isolated from endophytic leaf and root tissues of an arcto-alpine pioneer plant *Oxyria digyna* [41, 42], from leaves of a high altitude medicinal plant *Arnebia euchorma* [43], and it

has been shown to be consistently present in leaves and photosynthetic stems of *Vaccinium vitis-idaea* and *V. myrtillus* in several sampling locations in Finland [15]. The *S. faeni* strains from *Vaccinium* species were shown to be AAP positive by BChl *a* fluorescence imaging and targeted PCR [15, 44]. In contrast, none of the strains isolated from *O. digyna* were AAP positive ([41, 42] and Nissinen, unpublished). Likewise, genomes of *S. faeni* strain ALB2 from *A. euchorma*, or the type strain of the species, MA-Olki, indicated no presence of AAP related genes [38, 45].

Although *S. faeni* strains have previously been isolated from various regions and plant species, the genomic correlations among the strains along different phototrophic characteristics within the same bacterial species are missing, however. For this, we isolated *S. faeni* strains from *V. vitis-idaea*, and together with our strain collection from various plants from different locations (Table 1), we obtained the whole genome sequences of both AAP positive and AAP negative *S. faeni* strains. With these collected strains, we conducted a comparative genomic analysis, concentrating on their entire genomic sequence profiles and phototrophic characters. This allowed to explore the host specific commonalities among these groups and further pinpoint genomic variations on the protein composition level of the photosynthetic gene clusters and photoreceptor sites, all the way to single site information with potential structural implications of the XR proteins.

3. Materials and Methods

3.1. Selection and Isolation of studied strains

We utilized *S. faeni* strains isolated during previous studies from phyllo-, endosphere and seeds of *O. digyna*, *V. myrtillus*, *V. vitis-idaea* and *B. vivipara (*[15, 42] and Nissinen, unpublished). The isolation sites spanned across diverse geographical locations in Finland and in Svalbard (Supplementary Fig. 1). We screened the strains from *O. digyna* (37 strains) and *B vivipara* (3 strains) for presence of BChl *a* and thus for AAP type phototrophy by NIR-fluorescence imaging of bacterial colonies [44]. All strains from *O. digyna* and *B. vivipara* were negative in NIR-imaging (data not shown). Still, 11 strains from *O.digyna* and 2 strains from *B. vivipara* were selected for full genome sequencing (Table 1). Additional AAP negative *S. faeni* strains were isolated from *V. vitis-idaea* (this study) as described in Nissinen et al. [42]. Post incubation, colonies exhibiting characteristic features of *Sphingomonas* were identified and subcultured onto sterile ½ R2A agar plates. Criteria for selection included colony size exceeding 2 millimeters and a distinct orange-light red pigmentation. Polymerase Chain Reaction (PCR) was performed to identify isolates in genus *Sphingomonas*. The primers used were SA429f (5' TAAAGCTCTTTTACCCG 3'), and SA933r (5' AAACCACATGCTCCACC 3') [46]. Colonies positively identified as *Sphingomonas* were then propagated on fresh ½ R2A media and identified by near full length amplicidation and sequencing of 16S rRNA gene with primers 27f (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492r (5'-TACGGYTACCTTGTTACGACTT-3'). In addition to AAP negative strains, seven AAP positive strains from previous study [15] were selected for this study.

The extraction of genomic DNA was carried out with MasterPure™ Complete DNA & RNA Purification Kit in accordance with the manufacturer's instructions.

3.2. Genome sequencing and processsing

The whole genome sequences of the strains were obtained using DNBseq technology, ensuring a read length of PR150 and allocating 2 GB per sample. The sequencing was conducted at BGI (bgi.com). The quality of the resultant sequencing reads was assessed and adapter sequences, contaminants, and low quality reads were removed using SOAPnuke [47], with the quality threshold set to Q > 20. The reads were then assembled into 7-132 contigs per genome using Unicycler v0.5.0 [48], further uploaded to the RAST Annotation Server [49] where the automated gene annotations were performed. Genomes in this study were deposited in the NCBI Genome Database under BioProject ID PRJNA1202125. NCBI's Average Nucleotide Identity (ANI) analysis confirmed that all strains belong to species *S. faeni*. The corresponding accession numbers are SAMN45951201, SAMN45951202, SAMN45951203, SAMN45951204, SAMN45951205, SAMN45951206, SAMN45951208, SAMN45951209, SAMN45951210, SAMN45951211, SAMN45951212, SAMN45951213, SAMN45951214, SAMN45951215, SAMN45951216, SAMN45951217, SAMN45951218, SAMN45951219, SAMN45951220, SAMN45951221, SAMN45951222, SAMN45951223, SAMN45951224, SAMN45951225, and SAMN45951226.

3.3. Comparative genome analyses and bioinformatics

The whole genome sequences were initially transferred to Geneious Prime v2024.0 (geneious.com) for processing, where sequence isolations were conducted. Isolated sequences were enhanced through in-house BLAST-P and the Biopython library function pairwise2.align.globalxx. These methodologies were implemented via in-house Python scripts to conduct searches on individual genomes. Phylogenetic trees based on these whole genome sequences were constructed by uploading the sequences to the Type Strain Genome Server (TYGS) [50]. Amino acid sequences for chosen proteins and protein complexes were aligned using ClustalW [51], and the aligned amino acid sequences were visualized using JalView [52]. Core genome alignment was generated using MAFFT (-e -n) [53, 54] for further analysis. The core genome refers to the set of genes shared by all strains within a given species or group, representing essential functions and common characteristics. PGC synteny was constructed using EasyFig version 2.2.5 [55].

Phylogenetic trees were assembled in Geneious Prime and MEGA v 11 [56] and subsequently refined and visualized using the Interactive Tree Of Life (iTOL) platform [57]. ANIs for the whole genome and core genome sequences were calculated using fastANI [58]. A multidimensional scaling (MDS) graph, illustrating the genetic distances among the strains based on fastANI results, was constructed using in-house Python script (Supplementary Fig. 2). Plasmid sequences within the utilized strains were extracted using the Platon Plasmid Finder tool [59]. Generally, visualizations of the results were generated with home-built Python scripts. MDS was applied to visualize the clustering of strains based on the distance matrix, and the results were plotted with annotated strain identifiers, and labels

indicating the proportion of variance explained by the principal coordinates. Statistical analyses and MSA were performed using custom Python scripts.

3.4. Structure predictions of the Xanthorhodopsin

Amino acid sequences of the XR proteins from 17 strains were aligned using ClustalW [51] to identify conserved motifs and polymorphisms. Protein structure predictions were performed using SWISS MODEL [60–64], which also provided Protein Data Bank (PDB) files for further analysis. The sequences were uploaded to UniProt to identify binding sites and functional motifs. DeepTMHMM – 1.0 [65] was used to predict the transmembrane locations of the proteins. The resulting protein structures and motifs were visualized using PyMOL 3.0 (pymol.org), highlighting key functional regions and conserved features essential for XR function.

4. Results

4.1. Genome sequence screenings

The genomes of 26 *S. faeni* strains varied in size from 4.1 to 5.4 Megabases (Mb). Although it was not possible to extract plasmid sequences from three strains, the number of plasmid sequences is estimated to range from one to seven. Supplementary Table 1 provides a detailed overview of the general characteristics of the 26 S. *faeni* strains examined in this study.

All the seven strains, which showed positive NIR-fluorescence signal in the bacterial imaging analysis (not shown), contained complete PGC sequences in their genome. This confirmed the AAP character of these strains and we named these strains as AAP positive strains. We detected XR genes in 16 of the 20 AAP negative strains. Despite lacking the AAP genes, this subgroup is also classified as phototrophic, as they have potential to utilize light-driven redox reactions through a retinal-rhodopsin-based proton pumping mechanism. Here, we call these XR strains. To summarize, in our analysis we have AAP positive group, which contain the PGC, XR group, which contain XR gene, and a “none” group that contain neither of those phototrophic gene patterns (four strains) (Table 1). Thirteen of the strains, all belonging to the “none” or XR groups, were isolated from endosphere, phyllosphere or seeds of *O. digyna* and *B. vivipara* in oroarctic and arctic regions. The seven AAP positive strains were collected from subarctic *V. myrtillus* or *V. vitis-idaea* plants, both from endosphere and phyllosphere, without any specifity of the location (Table 1). In addition one none-strain and four XR strains originated from *V. vitis-idaea* plants.

Table 1

Strains used in this study. The NCBI accession number of MA-Olki type-strain is under the "Location". The strains are ordered according to their isolation source and phototrophy. FI: Finland, SV: Svalbard. The map of locations is shown in Supplementary Fig. 1.

Strain	Host	Location	Niche	Phototrophy	Source
VOU02AS02	*V.myrtillus*	Oulu, FI	Endophytic	AAP	[15]
VTK02DS03	*V.myrtillus*	Turku, FI	Endophytic	AAP	[15]
VUT02AP02	*V.myrtillus*	Utsjoki, FI	Epiphytic	AAP	[15]
VTK02EP11	*V.myrtillus*	Turku, FI	Epiphytic	AAP	[15]
VKS02BS02	*V.myrtillus*	Kuusamo, FI	Endophytic	AAP	[15]
PTK02CS01	*V.vitis-idaea*	Turku, FI	Endophytic	AAP	[15]
PES02AP01	*V.vitis-idaea*	Espoo, FI	Epiphytic	AAP	[15]
POU06NBS01	*V.vitis-idaea*	Oulu, FI	Endophytic	XR	This study
POU06NCP01	*V.vitis-idaea*	Oulu, FI	Epiphytic	XR	This study
POU06NBP02	*V.vitis-idaea*	Oulu, FI	Epiphytic	XR	This study
POU06NCP02	*V.vitis-idaea*	Oulu, FI	Epiphytic	XR	This study
POU06NBS02	*V.vitis-idaea*	Oulu, FI	Endophytic	None	This study
L1DT4B	*O. digyna*	Longyearbyen, SV	Epiphytic	XR	This study
L1BT3	*O. digyna*	Longyearbyen, SV	Epiphytic	XR	This study
L1BD2	*O. digyna*	Longyearbyen, SV	Epiphytic	XR	This study
S2H28	*O. digyna*	Kilpisjärvi, FI	Endophytic	XR	[42]
9OR8	*O. digyna*	Kilpisjärvi, FI	Seed	XR	This study
26OR1	*O. digyna*	Kilpisjärvi, FI	Seed	XR	This study
JPL28	*O. digyna*	Kilpisjärvi, FI	Endophytic	XR	[41]
S3H21	*O. digyna*	Kilpisjärvi, FI	Endophytic	XR	[42]
JWL30	*O. digyna*	Kilpisjärvi, FI	Endophytic	XR	[41]
JWL22	*O. digyna*	Kilpisjärvi, FI	Endophytic	XR	[41]
JWL2	*O. digyna*	Kilpisjärvi, FI	Endophytic	XR	[41]
JOBD13	*B. vivipara*	Kilpisjärvi, FI	Epiphytic	XR	This study
JOBD21	*B. vivipara*	Kilpisjärvi, FI	Epiphytic	None	This study
MA-Olki	NCBI	GCA_003053745.1	N/A	None	NCBI

4.2. Phototrophicity correlates with overall genomic composition

A phylogenetic tree of the 26 *S. faeni* strains was established using the full length 16S rRNA gene (Fig. 1A). No differences were found in the 16S rRNA gene sequences among the AAP strains, which shared 100% similarity. 16S sequence differences between the AAP and none groups were 0.7 ± 0.1%, and between the AAP and XR groups, 0.47 ± 0.2%. Subsequently, we aligned the whole genomes to construct a phylogenetic tree, presented in Fig. 1B. In whole genome alignment, the maximum difference between AAP and XR strains was 15 ± 1.5%, while the difference between AAP and none strains was 10.7 ± 1%. Whole genome similarity among the AAP strains was 96 ± 0.7%. We distributed the whole genome sequences into four lineages (L1-L4, Fig. 1B). The AAP positive *S. faeni* strains (L3 lineage) cluster within the same phylogenetic group. The XR and none groups predominantly cluster according to their host. The L1 contains strains only from *V. vitis-idaea*. All epiphytic *O. digyna* strains locate in lineage L2. The L4 is the most heterogeneous linage in terms of the host or niche having strains from *B.vivipara*, *O. digyna* and *V. vitis-idaea*, and both XR and none phototrophic groups, including the model strain of the study (Fig. 1B).

4.3. Photosynthetic gene cluster sequence and Xanthorhodopsin structures

The phototrophic strains have either a PGC or an XR gene. The PGC is very conserved in gene syntheny in the all studied *S. faeni* AAP strains and only minor changes in the sequences of unknown hypotetical proteins exist (Supplementary Fig. 3). Due to the large size of the PGC, we did not perform a single amino acid level analysis for this region. Such detailed analysis was conducted, however, for the 17 obtained XR sequences. The amino acid similarity within the XR sequences ranged from a minimum of 79.6% to a maximum of 100% between strains. The phylogenetic tree presented in Fig. 2 shows that, with the exception of strain S3H21, the XR sequences group according to the same lineages as observed in the whole genome analysis presented in Fig. 1B.

To explore potential structural differences, XR sequences from each lineages were selected for detailed structural analysis. Amino acid sequences were aligned to well known XR sequences from *Salinibacter ruber* [25, 26, 66], Kin4B8 from *Bdellovibrionota bacterium* [67], and AAP5 from *S. glacialis* [68, 69]. Alignment is shown in Supplementary Fig. 4. The secondary structure predictions indicated that the all XR proteins consist of seven alpha-helices (Fig. 3) and show very similar structural motifs with seven transmembrane regions, marked from 1 to 7.

Key conserved motifs essential for XR function were identified. The primary proton acceptor Asp96 in *S. ruber* [66] corresponds the Asp92 in *S. faeni* strains (Fig. 3A and Supplementary Fig. 4). In *S. ruber*, the amino acid functioning as the proton donor at position Glu107 [66] corresponds to Glu103 in *S. faeni* strains. The Schiff base binding residue, located at Lys240 in *S. ruber* [66], is observed at Lys228 in *S.*

faeni. Carotenoid binding sites in *S. faeni* strains exhibit substitutions compared to those in *S. ruber* which binds salinixanthin as a carotenoid, but show closer correspondence with those of Kin4B8 (*B. bacterium*) and AAP5 (*S. glacialis*) which bind zeaxanthin and nostoxanthin, respectively [67, 69]. The Gly153 in Kin4B8 (Gly156 in *S. ruber*), which facilitates a so called fenestration and therefore energy transfer from a carotenoid to retinal, can be found in all studied *S. faeni* strains as Gly149. Other common residues in the carotenoid binding sites are listed in Supplementary Table 2 and show closer correspondence to those of the Kin4B8 and AAP5 sequences, particularly to a tyrosine residue (Y209) in Kin4B8, which has been shown to be important for carotenoid hydroxyl group binding [67], than that of *S. ruber*. It is also interesting to note that in our protein models, together with the alignment of Kin4B8 (with PDB code 8I2Z), a Lys184 site in the studied *S. faeni* strains resembles, structurally, the Asn188 position in Kin4B8 among several other amino acids along the carotenoid molecule (Fig. 3C), suggesting that *S. faeni* XRs bind a hydroxyl containing carotenoid molecule, like zeaxanthin or nostoxanthin, instead of a 4-keto group carotenoid, such as salinixanthin or echinenone. A few carotenoid binding sites show variation among the *S. faeni* groups; the L1 (POU06-group) lineage has a Val145 instead of Ala145 in other lineages, but also lineage independent variation at the positions Arg184 versus Lys184 and Ser153 versus Ala153 can be pinpointed (Supplementary Fig. 4 and Supplementary table 2). It will be interesting to investigate in future studies whether these amino acids play a critical role in determining the carotenoid type or the energy transfer efficiency in these strains and how the carotenoid is located in these XR proteins, as the SWISS-MODEL structural predictions lack the positioning of the cofactors of the proteins.

The conserved 3-omega motif, a characteristic of various eubacterial rhodopsins, is partially present in the *S. faeni* strains. In *S. ruber*, the first aromatic amino acid of this motif is Tyr13 in helix A [66], while in *S. faeni*, it corresponds to Tyr9 in POU06NCP02, and Phe9 in JWL30 strains (Supplementary Fig. 4). The second aromatic component, Trp70 in helix B of *S. ruber*, is Trp66 in *S. faeni*. However, instead of the third aromatic component in the B-C loop, Tyr81 in *S. ruber*, non-aromatic residues Val77 and Ile77 exists in L1 and L4 strains, respectively.

In summary, the analysis identified key conserved motifs, primary proton donor, the retinal binding pocket with the critical proton acceptor and another proton donor, the proton release group, and the Schiff base for which carotenoid can transfer excitation energy. The carotenoid binding site suggests for a hydroxyl grouped carotenoid and some amino acid variations among and between the lineages can be observed, which may indicate different type of carotenoid binding in each of the *S. faeni* strains.

4.4. 5-ALA Sequences

5-ALA is a crucial precursor in the biosynthesis of tetrapyrroles, which include essential molecules such as chlorophyll, heme, and vitamin B12 [70]. In phototrophic bacteria, 5-ALA is synthesized through two main pathways: the C4 pathway (also known as the Shemin pathway) and the C5 pathway [71]. The C4 pathway involves the condensation of glycine and succinyl-CoA, catalyzed by the enzyme 5-ALA synthase, which is encoded by the *hemA* gene. During the analysis of PGC's of AAP strains, 5-ALA, a precursor in heme synthesis [72] was found. As expected, 5-ALA sequences were identified in all strains,

regardless of the phototrophic group. Intriguingly, the AAP positive strains exhibited the presence of two copies of the gene involved in 5-ALA synthesis, with one copy consistently located within the PGC, while the other is situated outside of the PGC. In contrast, AAP negative strains (both ‘XR’ and ‘none’ group) contained only a single copy. We extracted the 5-ALA sequences from our strains, and then performed an alignment. In the alignment results, the amino acid sequences of 5-ALA across phototrophic groups demonstrate a minimum similarity of 87%. However, within the PGC, the 5-ALA sequences exhibit a maximum similarity of only 67% when compared with 5-ALA sequences from outside the PGC. Conversely, non-PGC 5-ALA sequences from AAP strains show up to 96% similarity with those from XR and other none group 5-ALA sequences.This difference can be observed in the phylogenetic tree constructed based on the alignment results (Fig. 4), which also shows that the 5-ALA phylogeny follows the host species distribution, marked as different lineages, as described in the Fig. 1B.

4.5. Photosensors in different phototrophic groups

We analyzed the photosensor capacity of the strains in two fold. First the number of Blue Light Using Flavin (BLUF), bacteriophytochromes (BphP), sensory rhodopsins (SR), and photoactive yellow proteins (PYP) for each strain, are identified as presented in Table 2. The photosensor content followed largerly the lineages based on whole genome similarity, presented in Fig. 1B. The strains in lineage L1 contained BLUFs and BPhPs, but none of these strain contained PYP or SR. On the other hand, all BLUF sensors were absent in the lineage L2, but multiple PYPs, BphPs and SR were present in this lineage. The L3 strains, which all contain PGC, had on average the highest number and the most versatile set of photosensors. The lineage L4 had the largest range in number of photoreceptors per strain, from the least amount but also one of the highest amount of photosensors of our strain collection. This lineage shows also that the non-phototrophic strains had generally the least amount of photosensors, as also observed in Ihalainen et al [28].

In the second analysis, the neighbouring genes of the identified photosensors associated with each lineage were investigated (Supplementary Table 3). Several patterns of functional groups could be identified in the analysis. The BLUF photoreceptors were flanked by Superoxide dismutase on the 5' side in the L1BT3 strain and 9OR8, belonging to the lineage L4. The BLUFs in strains of the lineage L1 all had a N5-carboxyaminoimidazole ribonucleotide synthase on the 3' side. In terms of BphP genes, we identified that in all studied strains one BphP is related with a sodium/bile acid transporter in the 5’ neighboring site. With the same BphP gene the 3’ side varied according to the lineages. The L1 showed a hypotetical protein, the L2 lineage indicated an estrase/lipase gene, the L3 a M38 beta-aspartyl dipeptidase gene and L4 an uncharacterized MFS-transporter or a ferrichrome-iron receptor. An exception was JOBD21 strain which also showed a M38 gene on the 3’ side from the BphP gene. The second dominating BphP related 5’ gene was a cAMP-binding proteins - catabolite gene activator gene, which could be found from strains from the L1, L3 and L4 lineages. The L2 lineage lacked these BphP related cAMP-binding protein genes but contained mannonate dehydratase genes on 3’ side in the few strains. An oxidoreductase gene on the 3’ side was found from the few strains of L4 lineage and transposase genes could be also identified from L1 and L4 strains. In the analysis of PYP-flanking genes

we found that some strains of the L2 lineage showed a chloride channel protein as a 5’ neighbor whereas AAP or XR containing strains from L3 and L4 lineages typically have GAF domain/GGDEF domain/EAL domain proteins in their 5' neighborhood. We found a PYP related or BphP related transcriptional regulator in one L2 strain (5’ side) and in two L3 strains (3’ side), respectively. All in all, multitude of potentially important gene products related to photosensor proteins could be identified, and it is highly interesting to test their functionality in biochemical and biophysical terms in the future.

Table 2

Types of light-harvesting systems, rhodopsins, and the number of photoreceptor proteins in the studied S. faeni strains. Rho: Rhodopsin Type, SR: Sensory Rhodopsin, XR: Xanthorhodopsin, LH: Light-Harvesting System, BLUF: BLUF Domain (Blue Light Using Flavin), BphP: Bacteriophytochrome, PYP: Photoactive Yellow Protein

Strain	LH	Rho	BLUF	BphP	PYP	Phototrophy	Lineage
PES02AP01 (P)	LH1	-	1	3	1	AAP	3
VTK02DS03 (P)	LH1	-	1	3	1	AAP	3
VTK02EP11 (M)	LH1	SR	-	2	1	AAP	3
VUT02AP02 (M)	LH1	SR	1	2	1	AAP	3
VOU02AS02 (M)	LH1	SR	1	2	1	AAP	3
PTK02CS01 (P)	LH1	-	1	2	1	AAP	3
VKS02BS02 (M)	LH1	-	2	3	1	AAP	3
POU06NCP02 (P)	-	XR	1	3	-	Rho	1
POU06NBS01 (P)	-	XR	1	1	-	Rho	1
POU06NCP01 (P)	-	XR	1	3	-	Rho	1
POU06NBP02 (P)	-	XR	1	2	-	Rho	1
L1BD2 (O)	-	XR	-	2	3	Rho	2
L1DT4B(O)	-	XR	-	2	2	Rho	2
L1BT3 (O)	-	SR & XR	2	1	1	Rho	-
S2H28 (O)	-	XR	1	4	1	Rho	4
JOBD13 (B)	-	XR	-	5	1	Rho	4
9OR8 (O)	-	XR	1	3	2	Rho	4
26OR1 (O)	-	SR & XR	-	2	1	Rho	4
JPL28 (O)	-	XR	-	2	1	Rho	4
S3H21 (O)	-	XR	-	1	1	Rho	4
JWL30 (O)	-	XR	-	1	1	Rho	4
JWL22 (O)	-	XR	-	1	1	Rho	4
JWL2 (O)	-	XR	-	1	1	Rho	4
JOBD21 (B)	-	-	1	2	1	None	4
MA-Olki (NCBI)	-	-	1	2	-	None	4

Strain	LH	Rho	BLUF	BphP	PYP	Phototrophy	Lineage
POU06NBS02 (P)	-	-	1	3	-	None	4

5. Discussion

The genetic composition of individual bacterial strains among the same species can be quite divergent. Here, we demonstrate this with 25 *S. faeni* strains isolated from various plant hosts, *O. digyna, V. myrtillus*, *V. vitis-idaea, and B. vivipara*, from boreal, sub-arctic and arctic regions. The strains could be divided into four different lineages based on whole genome alignment, which was also reflected in more detailed genomic analysis along this study. The lineages based on whole genome alignment correlated with three phototrophic strategies of these strains: AAP positive, XR containing, and non-phototrophic bacterial strains. Interestingly, their overall phylogeny together with their phototrophy correlated also with the plant from which these strains were isolated. Our AAP positive strains were all from *Vaccinium* species, from phyllo- or endosphere of *V. myrtillus* or from *V. vitis-idaea* plants. Both of these plant species are perennial shrubs dominating in boreal forests and common also in arctic fell flora, and thus, offer a long term habitat for their microbial companions. In our previous study, we detected a high abundance of AAPB - in particular endospheric AAPB - in perennial plants that retain their photosynthetic tissues also through winter months [15]. In that same study, we isolated diverse AAP positive Methylobacteria from these *Vaccinium* species [15]. What might be the role, if any, of AAP in the interaction with *Vaccinium sp.* and AAPB, remains to be investigated. Our preliminary analysis of the genes other than PGC specific to AAP positive, Vaccinium associated strains did not offer any obvious clues to this, but are an obvious target for further studies.

Curiously, throughout our campaigns we never found any AAP positive *S. faeni* strains from *O. digyna* or from *B. vivipara*, although *S. faeni* has been shown to be abundant in these plants [[15] and Nissinen, unpublished]. The endophytic strains from *O. digyna* formed their own lineage in whole genome alignment. Instead we found that the strains from *O. digyna* contained an XR protein indicating therefore a potential for phototrophy but without “an expensive” photosynthetic machinery. Unlike *V. myrtillus* and *V. vitis-idaea*, *O. digyna* and *B. vivipara* overwinter as rhizomes, with decidious above-ground stems and leaves.

Curiously, most of the AAP negative XR containing *S. faeni* strains formed an own genomic linage, clearly distinct from AAP positive *Vaccinium*-associated strains. The genomic composition of this linage (L1 in Fig. 1B) is particularly similar in the multidimensional scaling, XR sequence, and photosensor composition.

In the predicted XR structures, the 3-omega motif of a Xanthorhodopsin containing AAP Rhodobacter strain, the third position is substituted by a non-aromatic Ile77 residue [24, 69], replacing the Tyr81 residue observed in *S. ruber* XR. Similarly, substitutions in the 3-omega motifs of *S. faeni* strains, involving non-aromatic residues such as valine and isoleucine, highlight the plasticity of this motif in XR

proteins. The amino acid substitutions observed in *S. faeni* XR proteins suggest both conservation of function and potential adaptive modifications. Key proton transfer residues, including Asp92 (Asp96 in *S. ruber*), Glu103 (Glu107 in *S. ruber*), and Lys228 (Lys240 in *S. ruber*), remain unchanged, indicating that the retinal Schiff base binding and proton pumping mechanism are likely preserved. However, notable substitutions in the carotenoid binding region (Fig. 3, Supplementary Fig. 4, Supplementary Table 2) suggest a nostoxanthin or zeaxanthin binding, similar to a comparable sequence of AAP5 [69], which shows even a dual phototrophic character. Notably, the C-terminal structure of *S. faeni* strains resembles to some extend that of light-gated cation channel rhodopsins [23].

Across all *S. faeni* strains in this study regardless of phototrophic group, the core genome, which consists of genes shared among all strains, exhibits minimal nucleotide variation (~ 1%), highlighting a high degree of genetic consistency across the species. Furthermore, genome analytics such as ANI analysis and MDS based on ANI values highlight genetic consistency within phototrophic groups. These results align with our findings, as the phylogenetic trees in this study also exhibit a similar topology, suggesting a long term evolution of the PGC with the overall genome of these strains. This pattern is consistent with observations in other species, such as *Citromicrobium sp.*, which harbors two distinct phototrophic groups [73], and haloalkaliphilic Rhodobacterales [74] which suggests that AAP strains have undergone stronger genetic conservation compared to other phototrophic or non-phototrophic groups.

The plant-associated *S. faeni* strains share their ecological niche with many other bacteria, making horizontal gene transfer plausible. However, the absence of significant differences in GC content between plasmids, PGC, and chromosomal DNA in this study indicates that the PGC has been co-evolving with the core genome for an extended period, with no strong evidence of recent horizontal gene transfer. Rather, it highlights that their genetic characteristics are closely integrated with the chromosomal genome.

5-ALA sequences were found in all *S. faeni* strains used in the study. However, while the XR and none strains have only one 5-ALA gene, the AAP strains have two 5-ALA genes, one of which is always in the PGC. This situation resembles a common feature in AAP species [74]. Heme, being an essential component in various proteins, including those integral to photosynthesis and electron transport [75], implicates the necessity for more robust biosynthetic pathways in these phototrophs. The duplication of the 5-ALA gene in AAP positive strains, especially the positioning of one copy within the PGC, might reflect an evolutionary adaptation, providing a means for more nuanced regulation of heme synthesis. This positioning could cater specifically to the photosynthetic machinery, while the other gene copy addresses general cellular functions that require heme. Conversely, the single gene copy in AAP negative strains suggests a simpler heme biosynthesis pathway, sufficient for their non-photosynthetic metabolic demands. Furthermore, 5-ALA is not just a precursor in heme synthesis but also plays a role in the biosynthesis of various tetrapyrroles, like bacteriochlorophylls [76].

Our photosensor analysis revealed here and in our previous study the wider need for photosensing among phototrophs [28]. However, the limited number of strains from the non-phototrophic group analyzed poses challenges in drawing statistically significant conclusions about this group. Still, the photosensor composition followed the linages in phylogeny grouping, but direct functional properties for particualr photosensors is challenging to propose. For example, we could not pinpoint any particular BphP genes near PGC regulating its function, similar as found for microsymbiotic nitrogen fixing bacteria *Bradyrhizobium japonicum* [77].

On the other hand, one BphP was associated with sodium/bile acid symporters in every strain in this study (Supplementary Table 3). Together with presence of transcriptional regulators of the LysR family, this suggests a network where light sensing proteins influence membrane transport and gene expression, aligning with previous findings that highlight the role of photoreceptors in environmental adaptation [78]. Similarly, the proximity of PYP photoreceptors to GAF domain proteins and chloride channel proteins suggests their potential role in signal transduction and ion transport. This photoreceptor has already been shown to regulate biofilm formation in *Idiomarina loihiensis* [79]. The frequent co-localization of BLUF photoreceptors with superoxide dismutase genes suggests a role in regulating the oxidative stress response. For instance, in *Acinetobacter baumannii*, BlsA, a BLUF-type photoreceptor, has been shown to effectively stimulate catalase activity, trehalose production, fluoroquinolone antibiotic tolerance, and the Type VI secretion system (T6SS) a macromolecular secretion machinery used by many Gram-negative bacteria to eliminate competitors [80, 81]. Recent studies further show the role of blue light photoreceptors in regulating photosynthetic and stress adaptive responses. In *Dinoroseobacter shibae*, the LOV-domain-containing protein LdaP was identified as a light dependent antirepressor that interacts with PpsR to regulate the photosynthetic gene cluster, demonstrating how photoreceptors modulate light driven gene expression at a transcriptional level [82].

To conclude, the genomic sequencing of 25 *S. faeni* strains, along with the type strain MA-Olki from the NCBI Genome Database, revealed genome sizes ranging from 4.1 to 5.4 Mb. Members of the XR group were the most prevalent, suggesting a widespread distribution of this photoreceptor among the strains. Our study enhances the understanding of microbial adaptation and the ecological dynamics of *S. faeni* in Arctic and boreal environments. The findings highlight the complex relationship between genetic diversity and environmental adaptability, emphasizing the need for further investigation into the ecological impacts and evolutionary pathways of phototrophic strains, including both AAP and XR containing strains.

Declarations

Used Python Scripts

All custom Python scripts can be found in this repository

github.com/batuhanthebioinformatician/Plant_Microbiome_Phototrophy

Competing Interests:

The authors declare that they have no relevant financial or non-financial interests to disclose.

Author Contribution

J.A.I and R.N. conceptualized the work. R.N. and B.D performed the experimental work. B.D. wrote the genome analysis scripts and together with R.N. and Y.Z performed the genomic analysis. B.D., J.A.I,R.N.,Y.Z. analyzed the data and wrote the manuscript. B.D. prepared all figures and tables of the manuscript.

Acknowledgement

The financial support for the study originates from KONE foundation (grant “Jaettu Valo”) and Science Council Finland (grant #259180 for Riitta Nissinen). Puhti supercomputer at the CSC - IT Center for Science in Finland was utilized to handle the demanding data processing and analysis tasks efficiently. This infrastructure provided the necessary computational power to conduct intensive genomic analyses and support the advanced bioinformatics processing required for this research. Ole Franz is acknowledged for instructive comments on the manuscript. The authors wish to thank Jani Hohti and Joonas Ikävalko for isolating the strains from O. digyna seeds.

Data Availability

Genomes in this study were deposited in the NCBI Genome Database under BioProject ID PRJNA1202125. NCBI’s Average Nucleotide Identity (ANI) analysis confirmed that all strains belong to species S. faeni. The corresponding accession numbers are SAMN45951201, SAMN45951202, SAMN45951203, SAMN45951204, SAMN45951205, SAMN45951206, SAMN45951208, SAMN45951209, SAMN45951210, SAMN45951211, SAMN45951212, SAMN45951213, SAMN45951214, SAMN45951215, SAMN45951216, SAMN45951217, SAMN45951218, SAMN45951219, SAMN45951220, SAMN45951221, SAMN45951222, SAMN45951223, SAMN45951224, SAMN45951225, and SAMN45951226.All custom Python scripts can be found in this repository github.com/batuhanthebioinformatician/Plant_Microbiome_Phototrophy

References

1. Holmberg, S. M., & Jørgensen, N. O. G. (2023). Insights into abundance, adaptation and activity of prokaryotes in arctic and Antarctic environments. *Polar Biology*, *46*(5), 381–396.
2. Morgan-Kiss, R. M., et al. (2006). Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiology And Molecular Biology Reviews*, *70*(1), 222–252.

3. Green, B. R., & Durnford, D. G. (1996). THE CHLOROPHYLL-CAROTENOID PROTEINS OF OXYGENIC PHOTOSYNTHESIS. *Annual Review Of Plant Physiology And Plant Molecular Biology*, *47*, 685–714.
4. George, D. M., Vincent, A. S., & Mackey, H. R. (2020). An overview of anoxygenic phototrophic bacteria and their applications in environmental biotechnology for sustainable Resource recovery. *Biotechnol Rep (Amst)*, *28*, e00563.
5. Thiel, V., Tank, M., & Bryant, D. A. (2018). *Diversity of Chlorophototrophic Bacteria Revealed in the Omics Era.* Annual Review of Plant Biology, 69(Volume 69, 2018): pp. 21–49.
6. Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews*, *39*(6), 854–870.
7. Xu, L., et al. (2024). *Aerobic Anoxygenic Phototrophic Bacteria in the Marine Environments Revealed by Raman/Fluorescence-Guided Single-Cell Sorting and Targeted Metagenomics*. Environmental Science & Technology.
8. Waidner, L. A., & Kirchman, D. L. (2008). Diversity and distribution of ecotypes of the aerobic anoxygenic phototrophy gene pufM in the Delaware estuary. *Applied And Environment Microbiology*, *74*(13), 4012–4021.
9. Fauteux, L., et al. (2015). Patterns in Abundance, Cell Size and Pigment Content of Aerobic Anoxygenic Phototrophic Bacteria along Environmental Gradients in Northern Lakes. *PLoS One*, *10*(4), e0124035.
10. Kolber, Z. S., et al. (2001). Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science*, *292*(5526), 2492–2495.
11. Tang, K. (2018). *Aerobic Anoxygenic Phototrophic Bacteria Promote the Development of Biological Soil Crusts*. *Frontiers in Microbiology*, 9.
12. Tahon, G., & Willems, A. (2017). Isolation and characterization of aerobic anoxygenic phototrophs from exposed soils from the Sør Rondane Mountains, East Antarctica. *Systematic And Applied Microbiology*, *40*(6), 357–369.
13. Aguirre-von-Wobeser, E. (2021). *Type II Photosynthetic Reaction Center Genes of Avocado (Persea americana Mill.) Bark Microbial Communities are Dominated by Aerobic Anoxygenic Alphaproteobacteria*. *Current Microbiology*, 78.
14. Atamna-Ismaeel, N., et al. (2012). Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environmental Microbiology Reports*, *4*(2), 209–216.
15. Nissinen, R. (2023). *Aerobic anoxygenic phototrophic bacteria are ubiquitous in phyllo- and endosphere microbiomes of boreal and subarctic plants.* bioRxiv, : p. 2023.02.19.529139.
16. Kolber, Z. S., et al. (2000). Bacterial photosynthesis in surface waters of the open ocean. *Nature*, *407*(6801), 177–179.
17. Béjà, O., et al. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, *411*(6839), 786–789.
18. Pinhassi, J., et al. (2016). Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiology And Molecular Biology Reviews*, *80*(4), 929–954.

19. Boichenko, V. A., et al. (2006). Functions of carotenoids in xanthorhodopsin and archaerhodopsin, from action spectra of photoinhibition of cell respiration. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1757*(12), 1649–1656.
20. Thiel, V. (2017). *The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. II. Metabolic Functions of Abundant Community Members Predicted from Metagenomic Analyses*. *Frontiers in Microbiology*, 8.
21. Ma, M., et al. (2024). *Proton-pumping rhodopsin of the coral symbiont Breviolum minutum and its potential role in coping with phosphorus deficiency in future warmer ocean*. Coral Reefs.
22. Chuon, K., et al. (2022). Carotenoid binding in Gloeobacteria rhodopsin provides insights into divergent evolution of xanthorhodopsin types. *Communications Biology*, *5*(1), 512.
23. Ernst, O. P., et al. (2014). Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews*, *114*(1), 126–163.
24. Kyndt, J. A. (2022). *Genome Sequence and Characterization of a Xanthorhodopsin-Containing, Aerobic Anoxygenic Phototrophic Rhodobacter Species, Isolated from Mesophilic Conditions at Yellowstone National Park*. *Microorganisms*, 10(6).
25. Hour, C., et al. (2024). Unveiling the critical role of K + for xanthorhodopsin expression in E. coli. *Journal of Photochemistry and Photobiology B: Biology*, *258*, 112976.
26. Balashov, S. P., et al. (2005). Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science*, *309*(5743), 2061–2064.
27. DeLong, E. F., & Béjà, O. (2010). The Light-Driven Proton Pump Proteorhodopsin Enhances Bacterial Survival during Tough Times. *PLOS Biology*, *8*(4), e1000359.
28. Ihalainen, J. A., et al. (2024). Multifaceted photoreceptor compositions in dual phototrophic systems – A genomic analysis. *Journal of Molecular Biology*, *436*(5), 168412.
29. Park, S. Y., & Tame, J. R. H. (2017). Seeing the light with BLUF proteins. *Biophysical Reviews*, *9*(2), 169–176.
30. van der Horst, M. A., & Hellingwerf, K. J. (2004). Photoreceptor proteins, star actors of modern times: a review of the functional dynamics in the structure of representative members of six different photoreceptor families. *Accounts Of Chemical Research*, *37*(1), 13–20.
31. Takala, H., et al. (2020). Tips and turns of bacteriophytochrome photoactivation. *Photochemical & Photobiological Sciences*, *19*(11), 1488–1510.
32. Multamäki, E., et al. (2021). Comparative analysis of two paradigm bacteriophytochromes reveals opposite functionalities in two-component signaling. *Nature Communications*, *12*(1), 4394.
33. Losi, A., & Gärtner, W. (2008). Bacterial bilin- and flavin-binding photoreceptors. *Photochemical & Photobiological Sciences*, *7*(10), 1168–1178.
34. Ziegler, T., & Möglich, A. (2015). *Photoreceptor engineering*. *Frontiers in Molecular Biosciences*, 2.
35. Losi, A., & Gärtner, W. (2021). A light life together: photosensing in the plant microbiota. *Photochemical & Photobiological Sciences*, *20*, 451–473.

36. Rockwell, N. C., & Lagarias, J. C. (2020). Phytochrome evolution in 3D: deletion, duplication, and diversification. *New Phytologist*, *225*(6), 2283–2300.
37. Auldridge, M. E., & Forest, K. T. (2011). Bacterial phytochromes: more than meets the light. *Critical Reviews In Biochemistry And Molecular Biology*, *46*(1), 67–88.
38. Busse, H. J., et al. (2003). Sphingomonas aurantiaca sp. nov., Sphingomonas aerolata sp. nov. and Sphingomonas faeni sp. nov., air- and dustborne and Antarctic, orange-pigmented, psychrotolerant bacteria, and emended description of the genus Sphingomonas. *Int J Syst Evol Microbiol*, *53*(Pt 5), 1253–1260.
39. Busse, H. J., Kämpfer, P., & Denner, E. B. (1999). Chemotaxonomic characterisation of Sphingomonas. *Journal Of Industrial Microbiology And Biotechnology*, *23*(4–5), 242–251.
40. An, H., et al. (2011). Sphingomonas xinjiangensis sp. nov., isolated from desert sand. *Int J Syst Evol Microbiol*, *61*(Pt 8), 1865–1869.
41. Given, C. (2020). *Tissue-Specific Dynamics in the Endophytic Bacterial Communities in Arctic Pioneer Plant Oxyria digyna*. *Frontiers in Plant Science*, 11.
42. Nissinen, R. M., Männistö, M. K., & van Elsas, J. D. (2012). Endophytic bacterial communities in three arctic plants from low arctic fell tundra are cold-adapted and host-plant specific. *Fems Microbiology Ecology*, *82*(2), 510–522.
43. Jain, R. (2021). *Arnebia euchroma, a Plant Species of Cold Desert in the Himalayas, Harbors Beneficial Cultivable Endophytes in Roots and Leaves*. *Frontiers in Microbiology*, 12.
44. Ole, F. (2023). *NIRis: A low-cost, versatile imaging system for NIR fluorescence detection of phototrophic cell colonies used in science and education.* bioRxiv, : p. 2023.05.31.543100.
45. Bhardwaj, P., Jain, R., & Kumar, S. (2022). Draft Genome Sequence of Endophytic Sphingomonas faeni Strain ALB2, Isolated from the Leaf of a Cold-Desert Medicinal Plant. *Microbiol Resour Announc*, *11*(11), e0068722.
46. Zhou, L., et al. (2012). Abundance and diversity of Sphingomonas in Shenfu petroleum-wastewater irrigation zone, China. *Environmental Science And Pollution Research International*, *19*(1), 282–294.
47. Chen, Y. (2017). *SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data*. *GigaScience*, 7(1).
48. Wick, R. R., et al. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *Plos Computational Biology*, *13*(6), e1005595.
49. Aziz, R. K., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *Bmc Genomics*, *9*(1), 75.
50. Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, *10*(1), 2182.
51. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680.

52. Waterhouse, A. M., et al. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.
53. Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166.
54. Kuraku, S., et al. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, *41*(W1), W22–W28.
55. Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics*, *27*(7), 1009–1010.
56. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027.
57. Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296.
58. Jain, C., et al. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114.
59. Schwengers, O. (2020). *Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores*. *Microb Genom*, 6(10).
60. Waterhouse, A., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–w303.
61. Bienert, S., et al. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Research*, *45*(D1), D313–d319.
62. Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*, *30*(Suppl 1), S162–S173.
63. Studer, G., et al. (2020). QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics*, *36*(6), 1765–1771.
64. Bertoni, M., et al. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, *7*(1), 10480.
65. Hallgren, J. (2022). *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks.* bioRxiv, : p. 2022.04.08.487609.
66. Luecke, H., et al. (2008). Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc Natl Acad Sci U S A*, *105*(43), 16561–16565.
67. Chazan, A., et al. (2023). Phototrophy by antenna-containing rhodopsin pumps in aquatic environments. *Nature*, *615*(7952), 535–540.
68. Kopejtka, K. (2021). *Characterization of the Aerobic Anoxygenic Phototrophic Bacterium Sphingomonas sp. AAP5*. *Microorganisms*, 9(4).

69. Kopejtka, K. (2022). *A bacterium from a mountain lake harvests light using both proton-pumping xanthorhodopsins and bacteriochlorophyll-based photosystems.* Proceedings of the National Academy of Sciences, 119(50): p. e2211018119.
70. Ilag, L. L., Kumar, A. M., & Söll, D. (1994). Light regulation of chlorophyll biosynthesis at the level of 5-aminolevulinate formation in Arabidopsis. *The Plant Cell*, *6*(2), 265–275.
71. Sasaki, K., et al. (2002). Biosynthesis, biotechnological production and applications of 5-aminolevulinic acid. *Applied Microbiology And Biotechnology*, *58*(1), 23–29.
72. Jiang, M., et al. (2022). *Natural 5-Aminolevulinic Acid: Sources, Biosynthesis, Detection and Applications* (p. 10). Frontiers in Bioengineering and Biotechnology.
73. Zheng, Q., et al. (2012). Gain and loss of phototrophic genes revealed by comparison of two Citromicrobium bacterial genomes. *PLoS One*, *7*(4), e35790.
74. Kopejtka, K., et al. (2017). Genomic Analysis of the Evolution of Phototrophy among Haloalkaliphilic Rhodobacterales. *Genome Biology and Evolution*, *9*(7), 1950–1962.
75. Layer, G., et al. (2010). Structure and function of enzymes in heme biosynthesis. *Protein Science*, *19*(6), 1137–1161.
76. Jiang, M., et al. (2022). Natural 5-Aminolevulinic Acid: Sources, Biosynthesis, Detection and Applications. *Frontiers In Bioengineering And Biotechnology*, *10*, 841443.
77. Giraud, E., et al. (2002). Bacteriophytochrome controls photosystem synthesis in anoxygenic bacteria. *Nature*, *417*(6885), 202–205.
78. Gomelsky, M., & Kaplan, S. (1998). AppA, a Redox Regulator of Photosystem Formation inRhodobacter sphaeroides 2.4.1, Is a Flavoprotein: IDENTIFICATION OF A NOVEL FAD BINDING DOMAIN*. *Journal of Biological Chemistry*, *273*(52), 35319–35325.
79. van der Horst, M. A., et al. (2009). Locked chromophore analogs reveal that photoactive yellow protein regulates biofilm formation in the deep sea bacterium Idiomarina loihiensis. *Journal Of The American Chemical Society*, *131*(47), 17443–17451.
80. Müller, G. L. (2017). *Light Modulates Metabolic Pathways and Other Novel Physiological Traits in the Human Pathogen Acinetobacter baumannii*. *Journal Of Bacteriology*, 199(10).
81. Squire, M. S., Townsend, H. A., & Actis, L. A. (2022). *The Influence of Blue Light and the BlsA Photoreceptor on the Oxidative Stress Resistance Mechanisms of Acinetobacter baumannii*. *Frontiers in Cellular and Infection Microbiology*, 12.
82. Pucelik, S. (2024). *The blue light-dependent LOV-protein LdaP of Dinoroseobacter shibae acts as antirepressor of the PpsR repressor, regulating photosynthetic gene cluster expression*. *Frontiers in Microbiology*, 15.

Figures

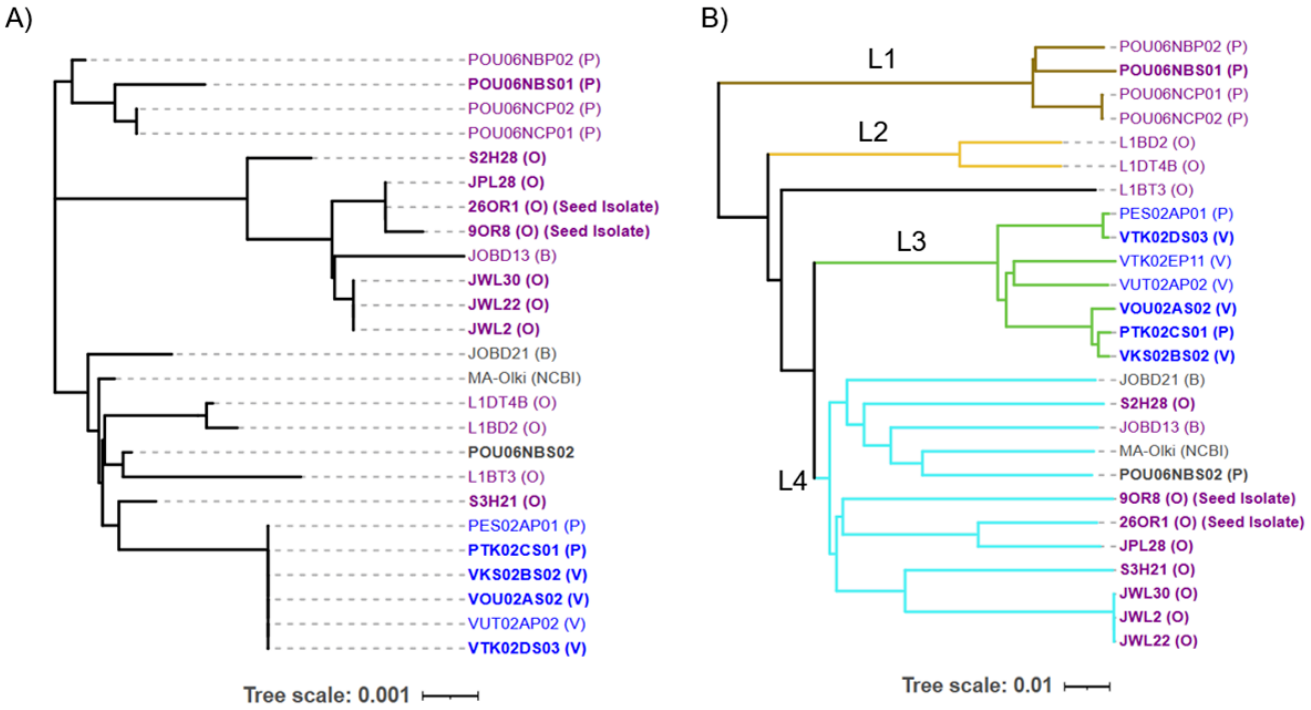

Figure 1

Phylogeny tree of 26 *S. faeni* strains built with Neighbour-Joining method A) Tree based on 16S rRNA gene alignment B) Tree of whole genome alignment. The phylogenetic tree features strains color-coded as follows: purple for XR strains, dark blue for AAP strains, and gray for non-phototrophic strains. Strains are distinguished by text formatting, with bold text representing endophytic strains and non-bold text for epiphytic strains. Isolation sources are indicated as: O for *O. digyna*, P for *V. vitis-idaea*, V for *V. myrtillus*, and B for *B. vivipara*. Seed isolated are marked in parenthesis. The lineages (L1-L4) used to categorize the strains in this study are color-coded as follows: L1 (brown), L2 (yellow), L3 (green), and L4 (cyan). Note that the strain L1BT3 is uncategorized and its phylogenetic branch is shown as a black line.

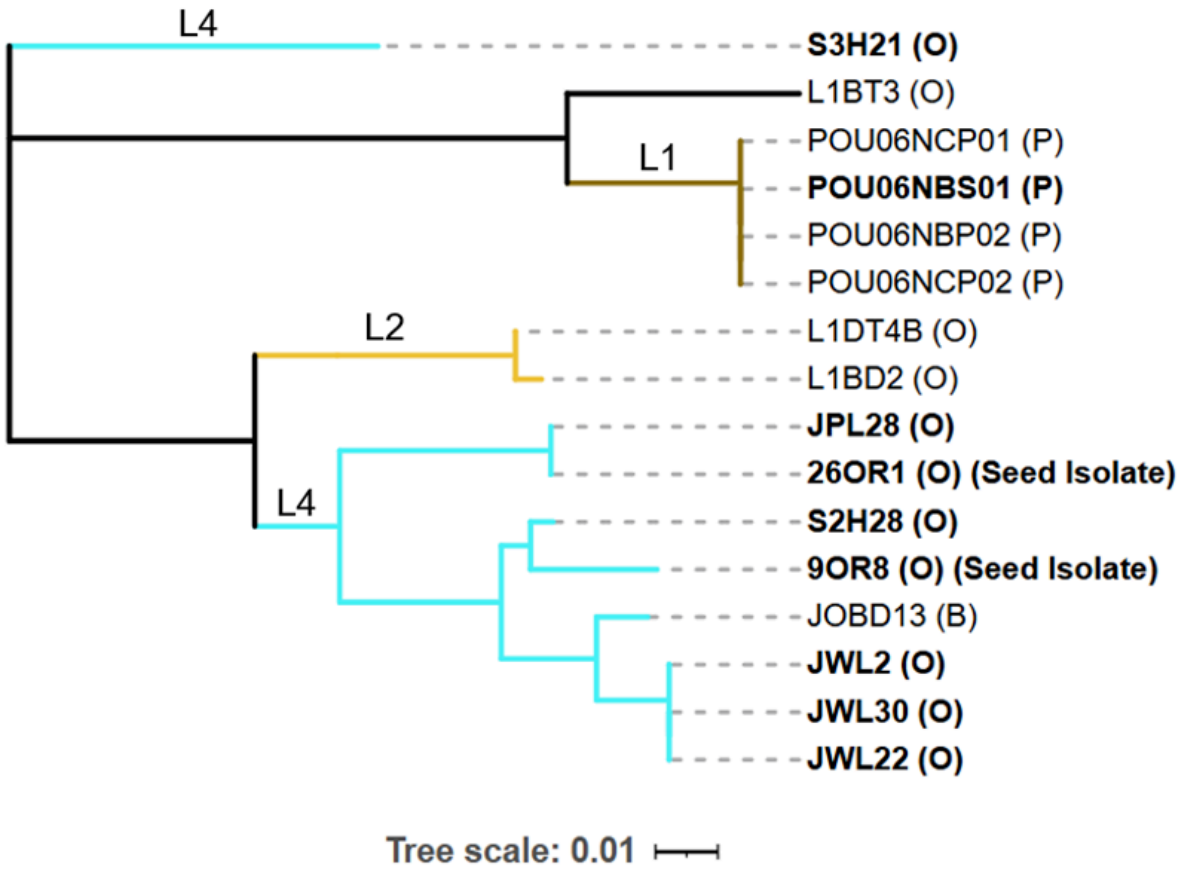

Figure 2

Phylogenetic tree of XR sequences. Bold text indicates XR from endophytic strains, while plain text shows epiphytic strains. Isolation sources are specified as follows: O for *Oxyria digyna*, P for *V. vitis-idaea*, and B for *B. vivipara*. Seed isolates are listed in parentheses.

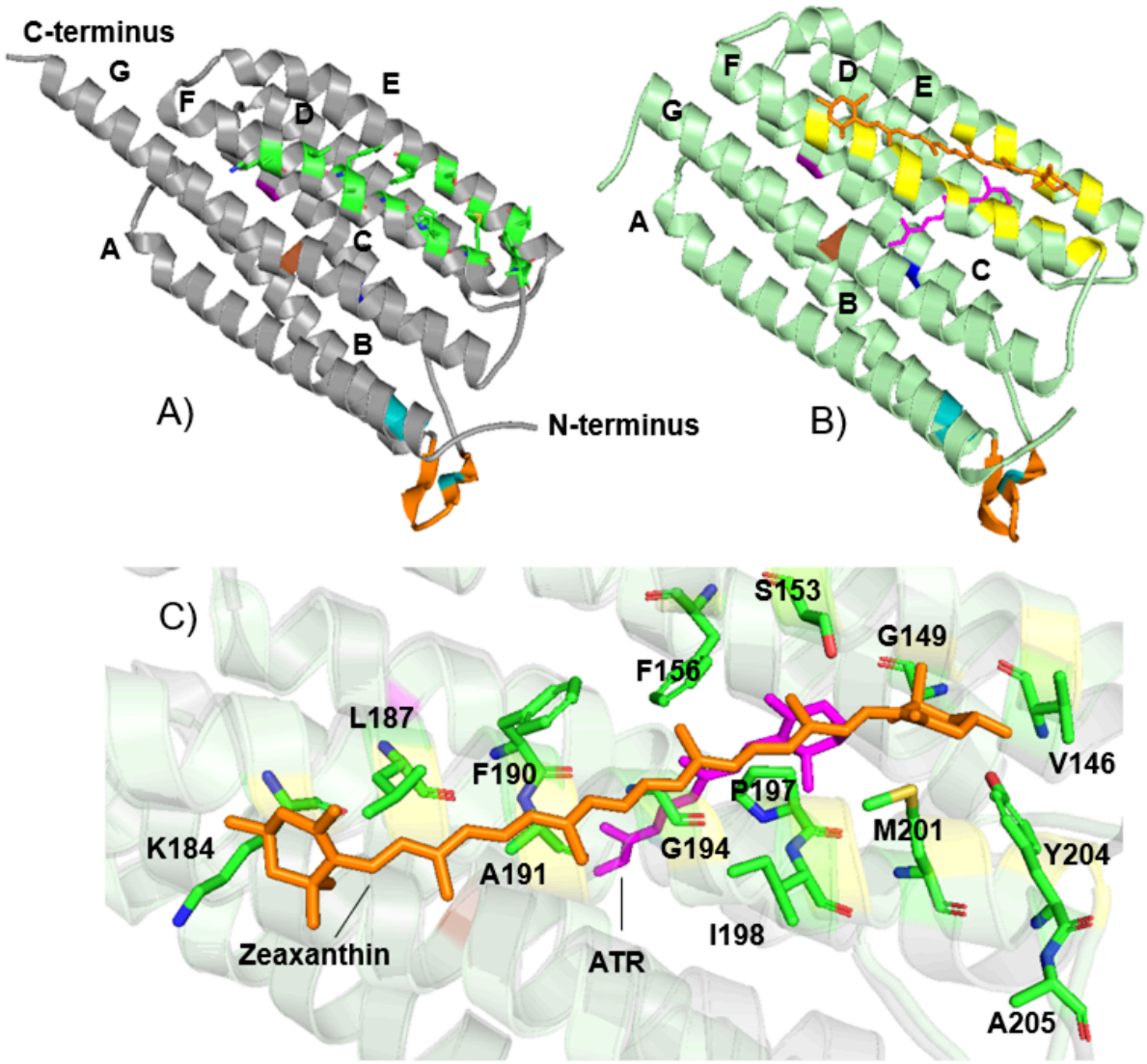

Figure 3

(A) Protein model of *S. faeni*strain POU06NBP02 generated by SWISS-MODEL. (B) The cryo-EM structure of Kin4B8 (*B.*bacterium) XR (pale green), the zeaxanthin is shown in orange and the all-trans-retinal (ATR) in purple. Carotenoid binding region in *S. faeni*strains F156, K184 (or R184 in some strains; see Supplementary Table 2), L187, F190, A191, G194, P197, and I198 are highlighted in yellow. Fenestration-associated region R146, G149, S153 (or A153 in some strains; see Supplementary Table 2), Y200, M201, Y204, and A205 are shown in pink. The B-C loop is colored orange. For *S. faeni*, proton donor region is shown in purple, proton acceptor region in blue, Schiff base residues in brown, and the Omega motif in

cyan. (C) Structural alignment of *S. faeni* POU06NBP02 (pale gray) and Kin4B8 (pale green) and its zeaxanthin binding positions in *S. faeni*. The key residues positioned according to the SWISS-MODEL are possibly involved in carotenoid binding and fenestration points in the *S. faeni* strains.

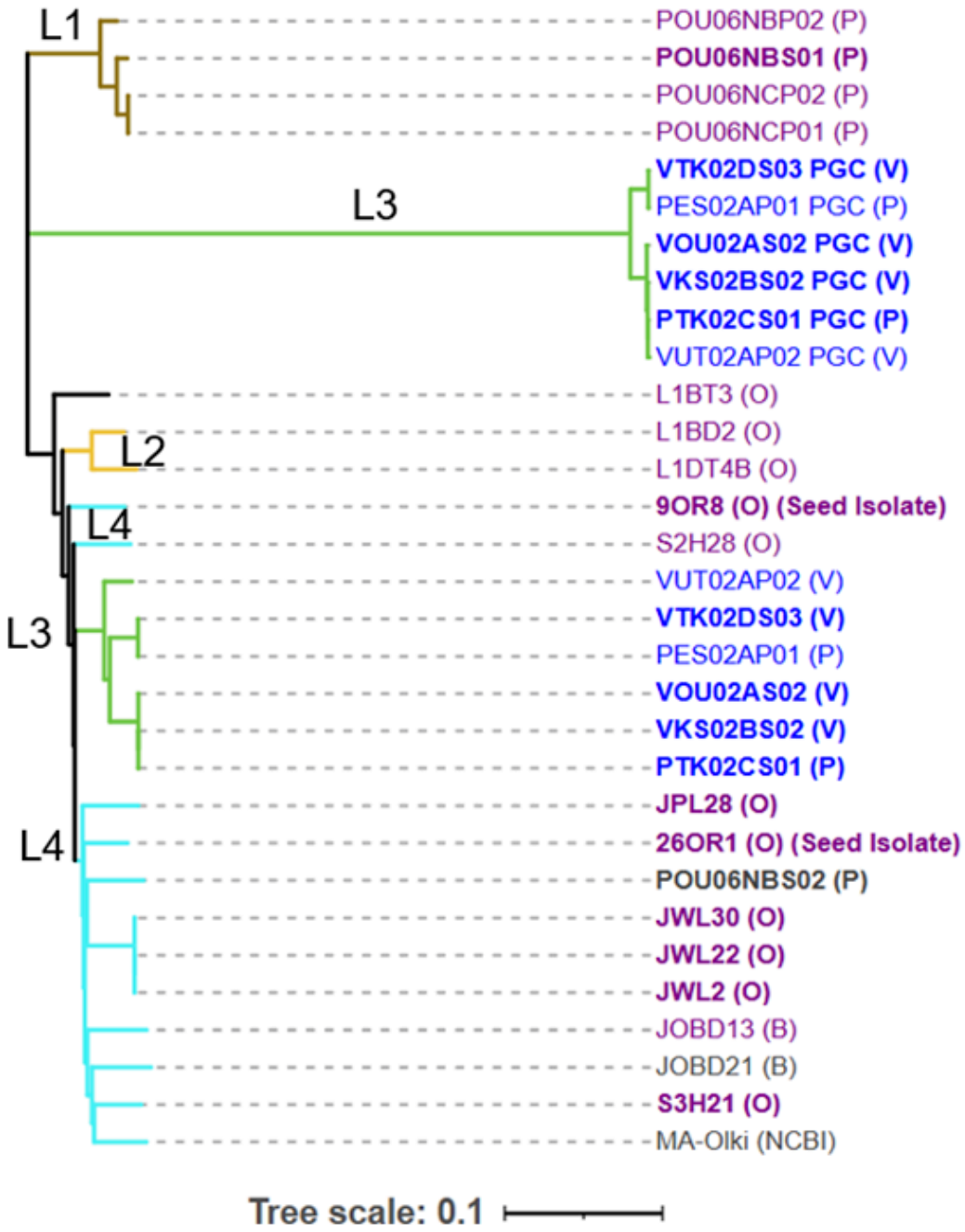

Figure 4

Phylogeny tree of 5-aminolevulinic acid nucleotide sequences built with Neighbour-Joining method. Strain name+PGC indicates the sequence is in the PGC. The phylogenetic tree features strains color-

coded as follows: purple for XR strains, dark blue for AAP strains, and gray for non-phototrophic strains. Strains are distinguished by text formatting, with bold text representing endophytic strains and plain text for epiphytic strains. The seed isolates are indicated inside parantheses. The specific isolation sources are detailed as: O for *Oxyria digyna*, P for *V. vitis-idaea*, V for *V. myrtillus*, and B for *B. vivipara*. The phylogeny tree deviates again as lineages, marked as L1 to L4, along the same strains as in Figure 1B.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFileFaeni.pdf