Photochemical & Photobiological Sciences
https://doi.org/10.1007/s43630-026-00991-0

RESEARCH

# Phototrophicity and genomic composition in plant-associated *Sphingomonas faeni* strains

**Batuhan Dogan**[1] · **Janne A. Ihalainen**[1] · **Yonghui Zeng**[2] · **Riitta Nissinen**[3]

Received: 10 March 2026 / Accepted: 24 August 2026


**Abstract**

Solar radiation impacts most life forms on Earth as an energy source or a regulatory signal. Still, relatively little is known about phototrophic potential and strategies of environmental bacteria beyond cyanobacteria. This study explores the phototrophy related genomic diversity of *Sphingomonas faeni* strains from boreal, sub-arctic and arctic regions. We analyzed the genomes of 25 plant-associated *S. faeni* strains isolated from *Vaccinium myrtillus*, *Oxyria digyna*, *V. 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 bacteria with a complete photosynthesis gene cluster, 16 strains contained xanthorhodopsin genes, and three strains were non-phototrophic, possessing no aerobic anoxygenic phototrophic or xanthorhodopsin genes. Aerobic anoxygenic phototrophic strains were found exclusively in *Vaccinium* hosts. *O. digyna* contained only xanthorhodopsin containing strains and *B. vivipara* showed xanthorhodopsin genes and one non-phototrophic strain. *V. vitis-idaea* hosted strains for all three different phototrophy categories. Phylogenetic analyses showed aerobic anoxygenic phototrophic positive strains forming a tight phylogenetic group. Xanthorhodopsin strains and non-phototrophic strains clustered into three different subgroups. Phototrophic strains had more photoreceptors. Aerobic anoxygenic phototrophic strains encoded two 5-aminolevulinic acid synthase isoenzymes, one from a *hemT*-like gene within the photosynthesis gene cluster and one from a *hemA*-like gene elsewhere in the genome. Our genomic analysis reveals substantial diversity in phototrophic potential among strains of a single bacterial species isolated from different host plants, possibly reflecting the distinct environmental cues each strain encountered.

**Keywords** plant-associated bacteria · phototrophy · aerobic anoxygenic phototrophy · xanthorhodopsin · comparative genomics

## 1 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

✉ Janne A. Ihalainen
janne.ihalainen@jyu.fi

✉ Riitta Nissinen
riitta.nissinen@utu.fi

[1] Nanoscience Center, Department of Biological and Environmental Science, University of Jyväskylä, 40014 Jyvaskyla, Finland

[2] Department of Plant and Environmental Sciences, University of Copenhagen, 1172 Copenhagen, Denmark

[3] Department of Biology, University of Turku, 20500 Turku, Finland

Published online: 04 September 2026

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 oxic 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 seaweed [7], AAPB are now known to be widespread in both marine and freshwater ecosystems [6, 8]. In these aquatic environments, AAPB can constitute a substantial fraction of the bacterial community, 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]. Kolber et al. [16] 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. Proton-pumping rhodopsins, 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]. Proton-pumping rhodopsins form a diverse group with varying spectral tuning, for instance, proteorhodopsins include both green-absorbing and blue-absorbing variants adapted to different light environments [17, 18]. Xanthorhodopsins (XR) represent a distinct subgroup of proton-pumping rhodopsins characterized by the presence of an auxiliary carotenoid antenna, which extends their light harvesting capacity into the blue spectral region in addition to the green absorption of the retinal chromophore itself [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, 23]. 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]. On the other hand, BChl *a* in the AAPB absorbs in the UV-A (Soret band), yellow (Qx band) and near-infrared (Qy band) regions [27], largely outside the blue-green window exploited by retinal-based proton pumps [18]. The biosynthesis of rhodopsin based proton pumps is energetically more frugal than the complex photosynthetic machinery required by bacteria [28]. 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 [29]. 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 [30]. In contrast, bacteriophytochromes (BphP) utilize a photoinduced isomerization process of bilin chromophores [31, 32]. 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 [33]. 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 [31, 34–38].

*Sphingomonas faeni* is a Gram-negative, non-spore-forming, psychrotolerant bacterium that was first described by Busse et al. [39], 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 [40, 41].

*S. faeni* has been repeatedly isolated from endophytic leaf and root tissues of an arcto-alpine pioneer plant *Oxyria digyna* [42, 43], from leaves of a high altitude medicinal plant *Arnebia euchroma* [44], 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 BChl *a*-positive by NIR fluorescence imaging of colonies and to contain photosynthesis gene cluster-related genes by targeted PCR on *pufM* gene, consistent with AAP character [15, 44]. In contrast, none of the strains isolated from *O. digyna* were BChl *a*-positive ([42, 43] and Nissinen, unpublished). Likewise, genomes of *S. faeni* strain ALB2 from *A. euchroma*, or the type strain of the species, MA-Olki, indicated no presence of AAP related genes [39, 45].

*S. faeni* strains have previously been isolated from various regions and plant species. The genomic correlations between phototrophic characteristics and overall genome composition within this species have not been examined,

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 us to explore the host specific commonalities among these groups and further pinpoint genomic variations on the protein composition level of the photosynthesis gene clusters (PGC) and photoreceptor sites, all the way to single site information with potential structural implications of the XR proteins.

## 2 Materials and methods

### 2.1 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, 43] 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* (three strains) for presence of BChl *a* and thus for AAP type phototrophy by NIR-fluorescence imaging of bacterial colonies [46]. All strains from *O. digyna* and *B. vivipara* were negative in NIR-imaging (data not shown). Eleven strains from *O. digyna* and two strains from *B. vivipara* were selected for whole genome sequencing (Table 1). Additional AAP negative *S. faeni* strains were isolated from *V. vitis-idaea* (this study) as described in Nissinen et al. [43]. 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*, using the genus-specific 16S rRNA gene primers SA429f (5' TAAAGCTCTTTTACCCG 3') and SA933r (5' AAACCACATGCTCCACC 3') [47]. Colonies positively identified as *Sphingomonas* were then propagated on fresh ½ R2A media and identified by near full length amplification and sequencing of the 16 S 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.

**Table 1** Strains used in this study

| 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 | [43] |
| 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 | [42] |
| S3H21 | *O. digyna* | Kilpisjärvi, FI | Endophytic | XR | [43] |
| JWL30 | *O. digyna* | Kilpisjärvi, FI | Endophytic | XR | [42] |
| JWL22 | *O. digyna* | Kilpisjärvi, FI | Endophytic | XR | [42] |
| JWL2 | *O. digyna* | Kilpisjärvi, FI | Endophytic | XR | [42] |
| 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 |

The NCBI accession number of MA-Olki type-strain is under the "Location". The strains are ordered according to their isolation source and phototrophy. The map of locations is shown in Supplementary Fig. 1

*FI* Finland, *SV* Svalbard

### 2.2 Genome sequencing and processing

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 [48], with the quality threshold set to Q > 20. The reads were then assembled into 7-132 contigs per genome using Unicycler v0.5.0 [49], further uploaded to the RAST Annotation Server [50] where the automated gene annotations were performed. Genome completeness and contamination were assessed using CheckM [51]. 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.

### 2.3 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) [52]. Amino acid sequences for chosen proteins and protein complexes were aligned using ClustalW [53], and the aligned amino acid sequences were visualized using JalView [54]. Core genome alignment was generated using MAFFT (-e -n) [55, 56] 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 [57].

Phylogenetic trees were assembled in Geneious Prime and MEGA v 11 [58] and subsequently refined and visualized using the Interactive Tree Of Life (iTOL) platform [59]. ANIs for the whole genome and core genome sequences were calculated using fastANI [60]. 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 [61]. 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.

### 2.4 Structure predictions of xanthorhodopsins

Amino acid sequences of the XR proteins from 16 strains were aligned using ClustalW [53] to identify conserved motifs and polymorphisms. Three-dimensional structures were predicted with AlphaFold3 [62]. The sequences were uploaded to UniProt [63] to identify binding sites and functional motifs. DeepTMHMM-1.0 [64] was used to predict the transmembrane locations of the proteins. To evaluate the potential for carotenoid-retinal coupling, the predicted structures were superposed onto the zeaxanthin-bound Kin4B8 xanthorhodopsin (Protein Data Bank: PDB 8I2Z) in PyMOL 3.1.8 (pymol.org) using the *super* command, and the all-trans-retinal was transplanted from the aligned template into each model. The presence of a fenestration was quantified as the solvent-accessible surface area (SASA) of the retinal β-ionone ring atoms (C1-C6, C16-C18), computed in PyMOL with a 1.4 Å probe (*dot_solvent* on, *dot_density* 4) on the protein-plus-retinal model, excluding all other heteroatoms. The same measurement was applied to *Salinibacter ruber* (PDB 3DDL) XR as a carotenoid-binding reference and to bacteriorhodopsin (PDB 1C3W) as a carotenoid-free reference. The resulting structures and motifs were visualized in PyMOL, highlighting key functional regions and conserved features essential for XR function.

## 3 Results

### 3.1 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 19 AAP negative strains. This subgroup is also classified as phototrophic, as these strains 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 (three 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 sub-arctic *V. myrtillus* or *V. vitis-idaea* plants, both from endosphere and phyllosphere. In addition one none-strain and four XR strains originated from *V. vitis-idaea* plants.

## 3.2 Phototrophicity correlates with overall genomic composition

Two phylogenetic trees of the 26 *S. faeni* strains were built using the full length 16 S rRNA gene, alongside six reference strains (*Rhodospirillum rubrum* ATCC 11170, *Porphyrobacter neustonensis* DSM 9434, *Sphingomonas astaxanthinifaciens* DSM 22298, *Sphingomonas kaistensis* DSM 16846, and *Sphingomonas glacialis* AAP5 and E6) included as phylogenetic context (Fig. 1A). In the 16 S tree, all reference strains formed a distinct outgroup clearly separated from the *S. faeni* strains. In our strains, no differences were found in the 16 S rRNA gene sequences among the AAP strains, which shared 100% identity. 16 S 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, again including the six reference strains as outgroup. In the whole genome tree, *S. glacialis* AAP5 and E6 placed closest to the *S. faeni* cluster among the references, consistent with their shared genus affiliation, while *R. rubrum* and *P. neustonensis* formed the most distant outgroup. Within the *S. faeni* strains, 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

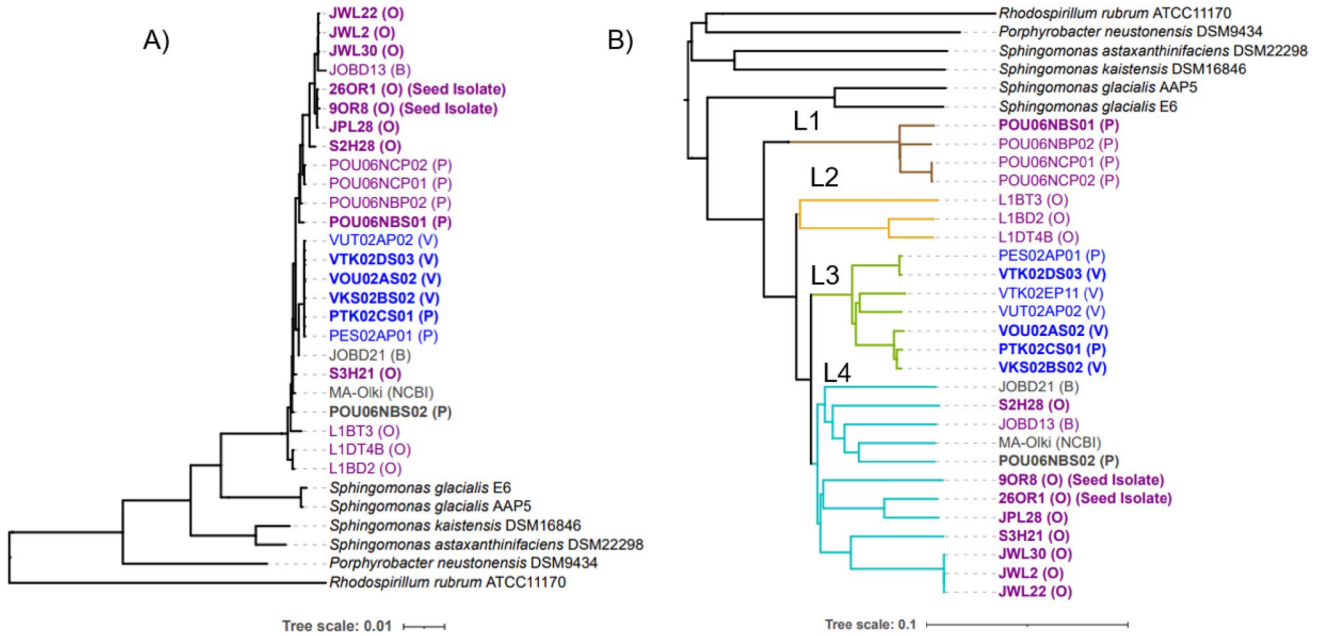


**Fig. 1** Phylogenetic trees of 26 *Sphingomonas faeni* strains built with the Neighbor-Joining method. **A** Tree based on 16 S rRNA gene alignment. **B** Tree based on whole genome alignment. In both trees, six reference strains (*Rhodospirillum rubrum* ATCC11170, *Porphyrobacter neustonensis* DSM9434, *Sphingomonas astaxanthinifaciens* DSM22298, *Sphingomonas kaistensis* DSM16846, and *Sphingomonas glacialis* AAP5 and E6) are included as outgroup and shown in black italic text. The *Sphingomonas faeni* strains are 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 *Oxyria digyna*, P for *Vaccinium vitis-idaea*, V for *Vaccinium myrtillus*, and B for *Bistorta vivipara*. Seed isolates are marked in parentheses. 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)

10.7 ± 1%. Whole genome identity 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 lineage 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).

### 3.3 Photosynthesis gene cluster sequence and xanthorhodopsin structures

The potential phototrophic strains have either a PGC or an XR gene. The PGC is very conserved in gene synteny in all studied *S. faeni* AAP strains and only minor changes in the sequences of unknown hypothetical proteins exist (Supplementary Fig. 3). To examine the phylogenetic placement of the *S. faeni* PGC within the broader AAP diversity, a *pufM*-based phylogenetic tree was constructed including the seven AAP-positive *S. faeni* strains alongside reference anoxygenic photosynthetic strains (*R. rubrum* ATCC11170, *P. neustonensis* DSM9434, *S. astaxanthinifaciens* DSM22298, *S. kaistensis* DSM16846, and *S. glacialis* AAP5 and E6) (Fig. 2). The *S. faeni* AAP strains formed a tight monophyletic cluster grouping most closely with *S. glacialis* AAP5 and E6, and clearly separated from more distantly related AAP taxa, consistent with stable lineage-specific maintenance of the PGC. 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 16 obtained XR sequences.

The amino acid identity within the XR sequences ranged from a minimum of 79.6% to a maximum of 100% between strains. The phylogenetic tree presented in Fig. 3 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. When placed in context with reference XR sequences from *Salinibacter ruber*, *Roseiflexus* sp. RS-1, uncultured *Bdellovibrionales* bacterium (Kin4B8), and *S. glacialis* AAP5, the *S. faeni* XR sequences form a distinct cluster separate from the halophilic *S. ruber* lineage.

To explore potential structural differences, XR sequences from each lineage were selected for detailed structural analysis. Amino acid sequences were aligned to well known XR sequences from *S. ruber* [25, 26, 65], Kin4B8 from *Bdellovibrionales* bacterium [66], and AAP5 from *S. glacialis* [67, 68]. Alignment is shown in Supplementary Fig. 4. The secondary structure predictions indicated that all XR proteins consist of seven alpha-helices (Fig. 4) 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* [65] corresponds to Asp92 in *S. faeni* strains (Fig. 4A and Supplementary Fig. 4). In *S. ruber*, the amino acid functioning as the proton donor at position Glu107 [65] corresponds to Glu103 in *S. faeni* strains. The Schiff base binding residue, located at Lys240 in *S. ruber* [65], is observed at Lys228 in *S. faeni*. The carotenoid-binding sites in *S. faeni* strains exhibit substitutions relative to *S. ruber*, which binds the 4-keto carotenoid salinixanthin, and instead align with hydroxylated-carotenoid binders such as Kin4B8, which binds zeaxanthin (Fig. 4B), and AAP5 of *S. glacialis*, which binds nostoxanthin [66–68]. The conserved fenestration glycine (Gly153 in Kin4B8 and Gly156 in *S. ruber*), which enables the opening required for carotenoid-to-retinal energy transfer, is retained in all studied *S. faeni* strains as Gly149. At the residue level, the divergent pocket positions listed in Supplementary Table 2 match the AAP5 sequence specifically rather than that of *S. ruber*. Among these positions, *S. faeni* carries Lys or Arg at 184 where both *S. ruber* and Kin4B8 carry Asn, matching the basic residue (Lys) of

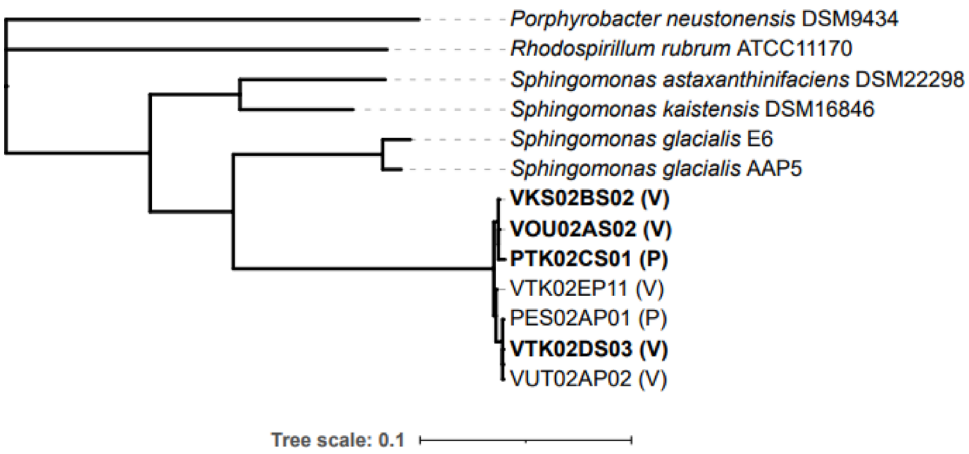


**Fig. 2** Neighbor-Joining phylogenetic tree of *pufM* sequences from the seven AAP-positive *Sphingomonas faeni* strains and reference AAP bacteria. Reference strains are shown in black italic text. *Sphingomonas faeni* strains are shown in non-italic text, with bold text representing endophytic strains and non-bold text representing epiphytic strains. Isolation sources are indicated as: P for *Vaccinium vitis-idaea* and V for *Vaccinium myrtillus*

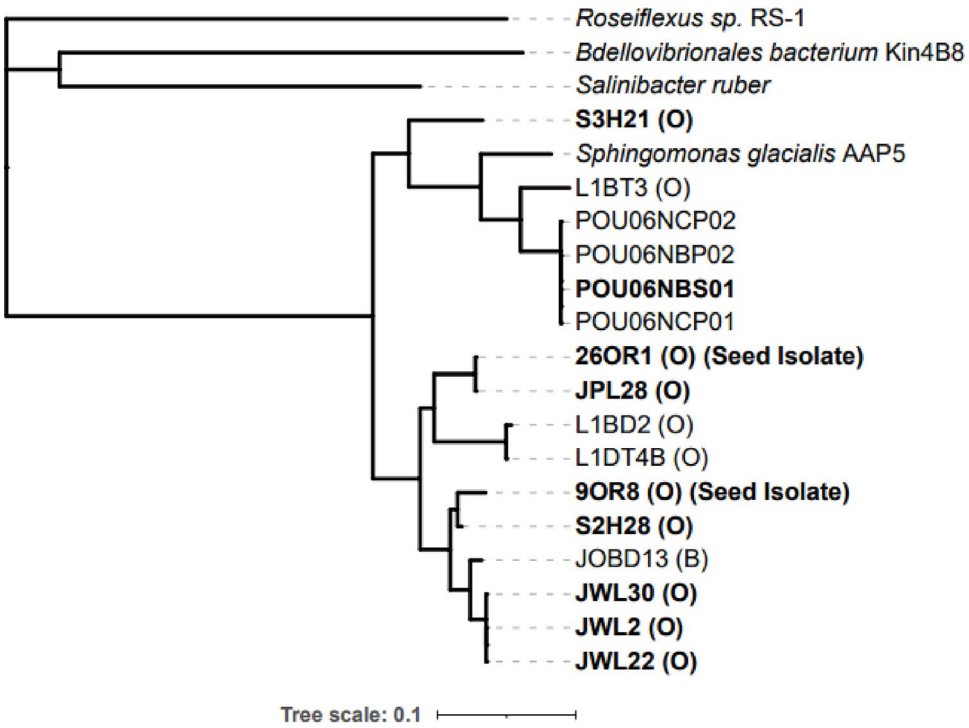


**Fig. 3** Neighbor-Joining phylogenetic tree of XR sequences from *S. faeni* strains and reference XR sequences. Reference strains (*Roseiflexus* sp. RS-1, *Bdellovibrionales* bacterium Kin4B8, *Salinibacter ruber*, and *Sphingomonas glacialis* AAP5) are shown in black italic text. *Sphingomonas faeni* strains are shown in non-italic text, with bold text representing endophytic strains and non-bold text representing epiphytic strains. Isolation sources are indicated as: O for *Oxyria digyna*, P for *Vaccinium vitis-idaea*, and B for *Bistorta vivipara*. Seed isolates are marked in parentheses

AAP5; in our models this residue is positioned to donate a hydrogen bond toward a carotenoid ring hydroxyl (Fig. 4C). Taken together, the sequence signature of the *S. faeni* carotenoid pocket tracks the nostoxanthin-binding AAP5 rather than the salinixanthin-binding *S. ruber*, consistent with binding of a hydroxylated carotenoid such as zeaxanthin or nostoxanthin instead of a 4-keto carotenoid such as salinixanthin or echinenone. A few pocket positions vary among the *S. faeni* strains: the L1 (POU06) lineage carries Val146 in place of the Ala146 found in AAP5 and the other *S. faeni* lineages, and lineage-independent variation is seen at position 184 (Lys versus Arg, both basic) and position 153 (Ser versus Ala) (Supplementary Fig. 4 and Supplementary Table 2). It will be interesting to investigate in future studies whether these residues influence the carotenoid type bound or the energy-transfer efficiency in these strains. To assess whether the predicted structures could in principle accommodate a carotenoid, we examined the fenestration adjacent to the retinal β-ionone ring in the AlphaFold3 models (Fig. 4D). In the POU06NBP02 model, the retinal β-ionone ring is exposed at the protein surface through this fenestration (solvent-accessible surface area 2.47 Å²), comparable to the carotenoid-binding xanthorhodopsin of *S. ruber* (2.08 Å²; PDB 3DDL) and markedly greater than bacteriorhodopsin, which lacks a carotenoid antenna (0.09 Å²; PDB 1C3W). This is consistent with a fenestration geometrically compatible with carotenoid-to-retinal excitation transfer, although carotenoid binding was not experimentally tested.

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 [65], while in *S. faeni*, it corresponds to Tyr9 in POU06NBP02, 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 exist 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 a hydroxylated 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.

### 3.4 5-Aminolevulinic acid synthase gene sequences

5-aminolevulinic acid (5-ALA) is a crucial precursor in the biosynthesis of tetrapyrroles, which include essential molecules such as chlorophyll, heme, and vitamin B12 [69]. In Proteobacteria, including most phototrophic members of this phylum, 5-ALA is synthesized through the C4 pathway (also known as the Shemin pathway), which involves the condensation of glycine and succinyl-CoA [70]. In *Cereibacter sphaeroides* 2.4.1, this activity is provided by two isoenzymes encoded by the separate genes *hemA* and *hemT* [71]. During the analysis of PGCs of AAP strains,

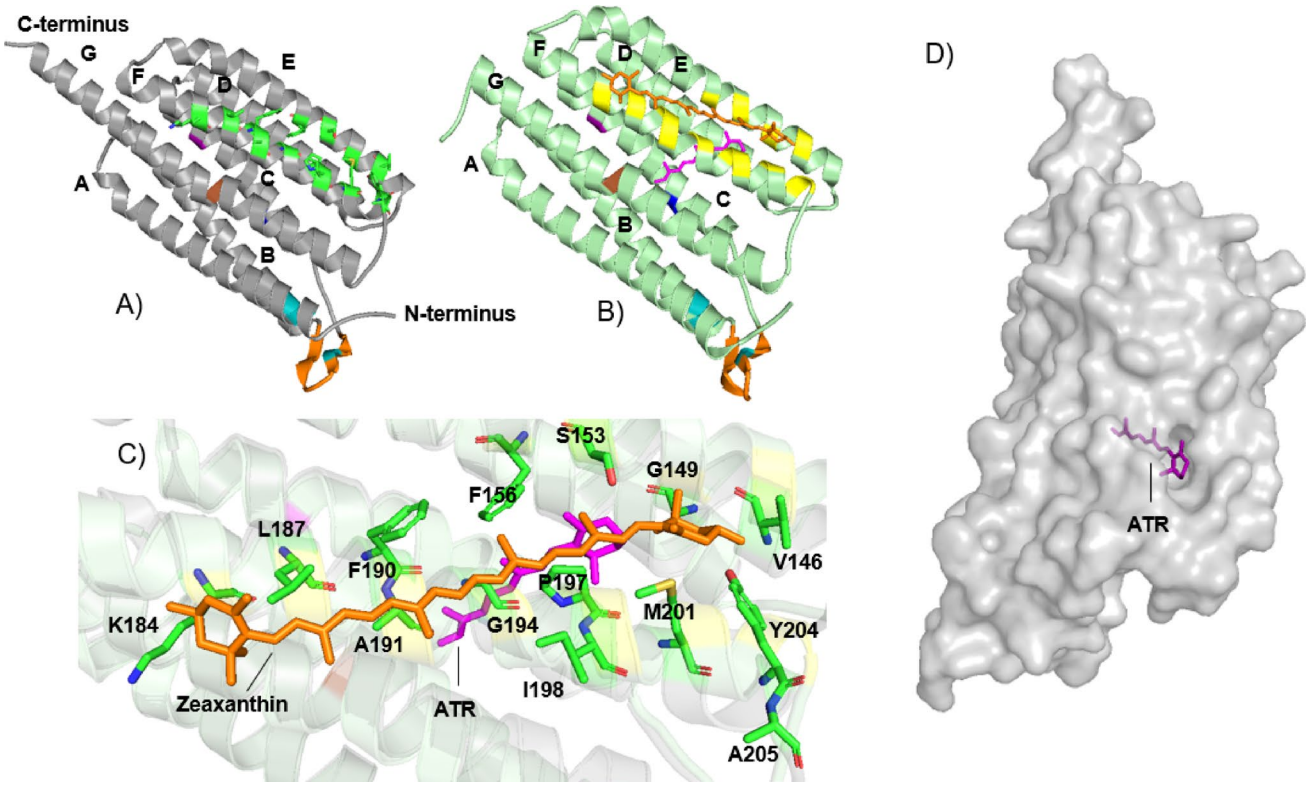


**Fig. 4** **A** AlphaFold3 model of *Sphingomonas faeni* strain POU06NBP02 (gray). Carotenoid-binding-region residues F156, K184 (or R184 in some strains; see Supplementary Table 2), L187, F190, A191, G194, P197, and I198 are highlighted in yellow; fenestration-associated residues V146 (or A146 in other strains; see Supplementary Table 2), G149, S153 (or A153 in some strains; see Supplementary Table 2), Y200, M201, Y204, and A205 in pink; the proton donor region in purple, proton acceptor region in blue, Schiff-base residues in brown, and the Omega motif in cyan. The B-C loop is colored orange. **B** The same residue positions mapped onto the cryo-EM structure of Kin4B8 XR (pale green), with zeaxanthin (orange) and all-trans-retinal (ATR, purple) resolved in the experimental structure. **C** Structural superposition of the POU06NBP02 AlphaFold3 model (pale gray) and Kin4B8 (pale green); the zeaxanthin (orange) is shown in the position obtained from the superposed Kin4B8 structure and is not present in the *S. faeni* model, indicating where a carotenoid ring would sit relative to the retinal-binding pocket. **D** Solvent-accessible surface of the POU06NBP02 model with all-trans-retinal transplanted from Kin4B8. The retinal β-ionone ring reaches the protein surface through a fenestration. The solvent-accessible surface area of the β-ionone ring atoms was 2.47 Å² in POU06NBP02, comparable to the carotenoid-binding xanthorhodopsin of *Salinibacter ruber* (2.08 Å²) and markedly greater than bacteriorhodopsin, which lacks a carotenoid antenna (0.09 Å²), indicating a fenestration geometrically consistent with carotenoid-retinal contact. All structural interpretations are based on computational predictions; carotenoid binding was not experimentally validated

a 5-ALA synthase gene was found. As expected, 5-ALA synthase genes were identified in all strains, regardless of the phototrophic group. AAP positive strains encoded two 5-ALA synthase isoenzymes, one from a *hemT*-like gene located within the PGC and one from a *hemA*-like gene outside of it (Fig. 5). In contrast, AAP negative strains (both 'XR' and 'none' groups) encoded only a single *hemA*-like 5-ALA synthase gene. We extracted the 5-ALA synthase sequences from our strains and performed an alignment. Across all 26 strains, the non-PGC sequences share 94.7–100% amino acid identity, with no separation between phototrophic groups: AAP versus XR 95.7–99.0%, AAP versus none 97.6–99.3%, and XR versus none 95.5–99.8%. The PGC-encoded sequences of the seven AAP strains are nearly invariant among themselves (99.5–100%) but share only 57.2–57.9% identity with any non-PGC sequence, including the non-PGC sequence of the same genome. No pairwise value falls between these two ranges.

To further characterize the 5-ALA synthase sequences, the *hemA* and *hemT* sequences from *C. sphaeroides* 2.4.1 [71] were included in the phylogenetic analysis (Fig. 5). The PGC-encoded 5-ALA synthase sequences from all seven AAP-positive *S. faeni* strains grouped with the *hemT* sequence of *C. sphaeroides*, while the non-PGC 5-ALA synthase sequences grouped with *hemA*.

### 3.5 Photosensors in different phototrophic groups

We analyzed the photosensor capacity of the strains in two ways. First, we identified the number of Blue Light Using Flavin (BLUF) proteins, bacteriophytochromes (BphP), sensory rhodopsins (SR), and photoactive yellow proteins

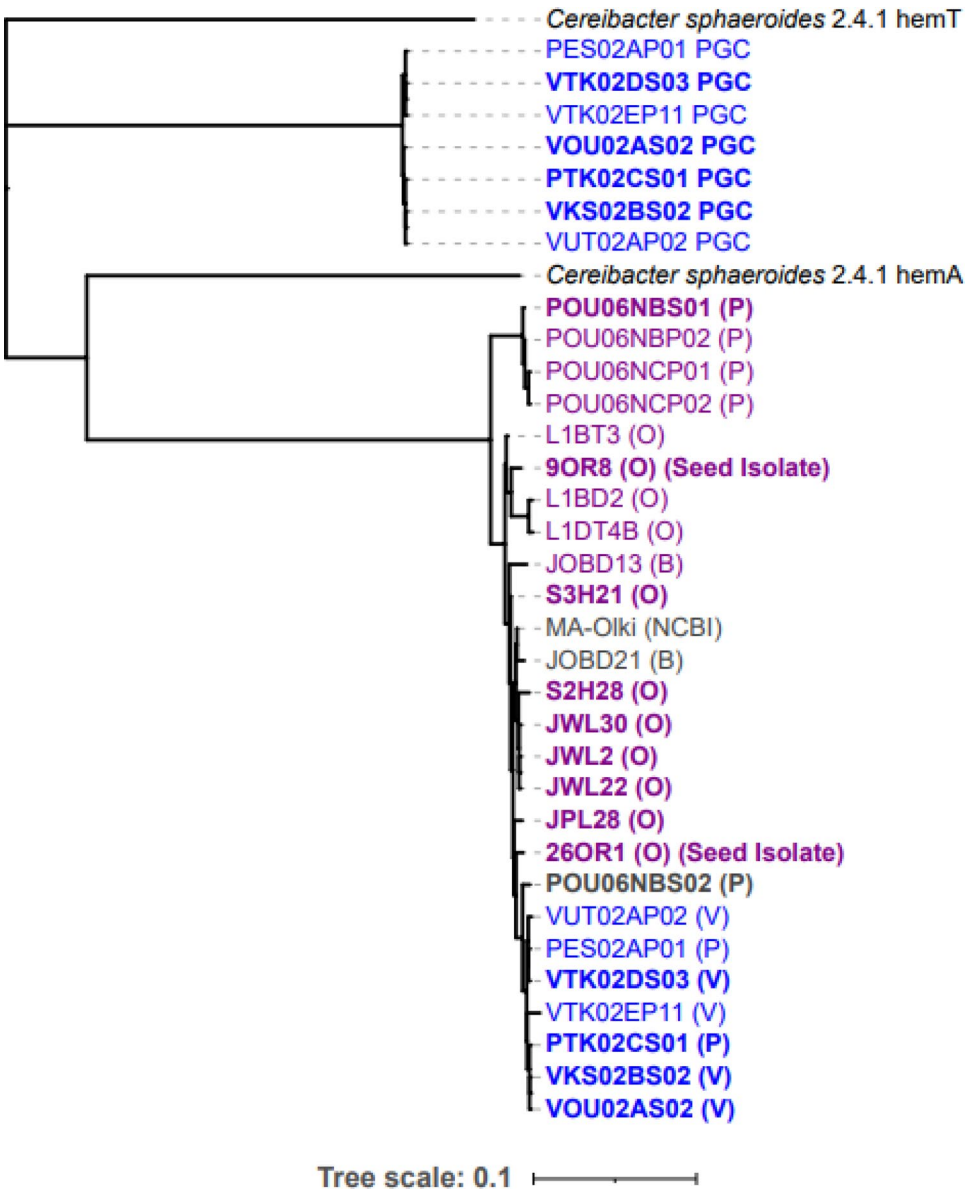


**Fig. 5** Neighbor-Joining phylogenetic tree of 5-aminolevulinic acid synthase amino acid sequences. Reference sequences *hemA* and *hemT* from *Cereibacter sphaeroides* 2.4.1 are shown in black italic text. The PGC-encoded sequences from all seven AAP-positive *Sphingomonas faeni* strains group with *Cereibacter sphaeroides hemT*, while all non-PGC sequences group with *Cereibacter sphaeroides hemA*, indicating that the two genes encode distinct isoenzymes. Strain name + PGC indicates the sequence is located within the photosynthesis gene cluster. The *Sphingomonas faeni* strains are color-coded as follows: dark blue for AAP strains, purple for XR 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. Seed isolates are indicated in parentheses. Isolation sources are indicated as: O for *Oxyria digyna*, P for *Vaccinium vitis-idaea*, V for *Vaccinium myrtillus*, and B for *Bistorta vivipara*

(PYP) in each strain (Table 2). The photosensor content followed largely the lineages based on whole genome identity, presented in Fig. 1B. The strains in lineage L1 contained BLUFs and BphPs, but none of these strains contained PYP or SR. BLUF sensors were absent from the L2 strains except L1BT3, whereas multiple PYPs, BphPs and SR were present in this lineage. The L3 strains, which all contain a PGC, had on average the highest number and the most versatile set of photosensors. Lineage L4 showed the largest range in photosensor number per strain, spanning both the lowest and among the highest counts in our strain collection.

In the second analysis, the neighboring genes of the identified photosensors associated with each lineage were investigated (Supplementary Table 3). Several patterns of functional groups could be identified. The BLUF photoreceptors were flanked by superoxide dismutase on the 5' side in strains L1BT3 (L2) and 9OR8 (L4), and all BLUFs in lineage L1 had an N5-carboxyaminoimidazole ribonucleotide synthase on the 3' side. In terms of BphP genes, we identified that in all studied strains one BphP has a sodium/bile acid transporter as its 5' neighbour, while its 3' neighbour varied according to lineage: most L1 strains showed a hypothetical protein, the L2 lineage an esterase/lipase gene, most L3 strains an M38 beta-aspartyl dipeptidase gene, and L4 strains an uncharacterized MFS-type transporter or a ferrichrome-iron receptor. An exception was strain JOBD21, which also showed an M38 gene on the 3' side of the BphP gene. The second most common BphP-related 5' gene was a cAMP-binding protein, found in strains of the L1, L3 and L4 lineages. The L2 lineage lacked these but carried mannonate dehydratase genes on the 5' side in two strains. An oxidoreductase gene on the 3' side was found in two

**Table 2** Types of light-harvesting systems, rhodopsins, and the number of photoreceptor proteins in the studied *S. faeni* strains

| Strain | LH | Rho | BLUF | BphP | PYP | Phototrophy | Lineage |
|---|---|---|---|---|---|---|---|
| PES02AP01 (P) | LH1 | - | 1 | 3 | 1 | AAP | 3 |
| **VTK02DS03 (V)** | LH1 | - | 1 | 3 | 1 | AAP | 3 |
| VTK02EP11 (V) | LH1 | SR | - | 2 | 1 | AAP | 3 |
| VUT02AP02 (V) | LH1 | SR | 1 | 2 | 1 | AAP | 3 |
| **VOU02AS02 (V)** | LH1 | SR | 1 | 2 | 1 | AAP | 3 |
| **PTK02CS01 (P)** | LH1 | - | 1 | 2 | 1 | AAP | 3 |
| **VKS02BS02 (V)** | 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 | 2 |
| 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 |
| **POU06NBS02 (P)** | - | - | 1 | 3 | - | None | 4 |

*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

L4 strains, and transposase genes were identified in L1 and L4 strains. In the analysis of PYP-flanking genes, some L2 strains showed a chloride channel protein as a 5' neighbour, whereas AAP and XR containing strains from the L3 and L4 lineages typically had GAF domain/GGDEF domain/EAL domain proteins in their 5' neighborhood. Notably, a sensor domain-containing diguanylate cyclase was the 3' neighbour of every PYP identified in this study, across all four lineages. BphP-related transcriptional regulators were found on the 3' side in two L2 strains (helix-turn-helix family) and two L3 strains (LysR family). Overall, a multitude of potentially important gene products related to photosensor proteins could be identified, and it will be interesting to test their functionality in biochemical and biophysical terms in the future.

## 4 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, 43] 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 gene 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 deciduous above-ground stems and leaves. Because host plant, geographic origin, and sampling niche co-vary in this dataset, the observed association between host and phototrophic strategy is correlational, and environmental drivers cannot be excluded as an alternative explanation to host specificity.

Most of the AAP negative XR containing *S. faeni* strains formed their own genomic lineage, distinct from AAP positive *Vaccinium*-associated strains. In the whole genome phylogeny, *S. glacialis* AAP5 and E6 placed closest to the *S. faeni* cluster among the reference strains, reflecting their shared genus affiliation, while *R. rubrum* and *P. neustonensis* formed the most distant outgroup, confirming the species-level coherence of the *S. faeni* strains analyzed in this study.

In an XR containing AAP Rhodobacter strain, the third aromatic position of the 3-omega motif is also substituted by a non-aromatic residue [24, 68], analogous to the situation in *S. ruber* (Tyr81) and *S. faeni*. Substitutions in the 3-omega motifs of *S. faeni* strains, involving non-aromatic residues such as valine and isoleucine, show the plasticity of this motif in XR proteins. In the XR phylogeny, the *S. faeni* sequences formed a distinct cluster clearly separated from the halophilic *S. ruber* lineage and from *Roseiflexus* sp. RS-1, grouping instead with *S. glacialis* AAP5, consistent with the observed divergence in carotenoid binding residues from the *S. ruber* archetype and the predicted preference for hydroxylated carotenoids such as zeaxanthin or nostoxanthin. 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. 4, Supplementary Fig. 4, Supplementary Table 2) suggest a nostoxanthin or zeaxanthin binding, similar to a comparable sequence of AAP5 [68], which shows even a dual phototrophic character. Notably, the C-terminal structure of *S. faeni* strains resembles to some extent that of light-gated cation channel rhodopsins [23]. It should be noted that all structural interpretations presented here are based on computational predictions and that carotenoid binding has not been experimentally validated.

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 [72], and haloalkaliphilic Rhodobacterales [73] 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 a plausible mechanism for PGC distribution, as has been proposed for other AAP lineages [72, 73]. The *pufM* phylogeny, however, shows that the *S. faeni* AAP strains form a tight monophyletic cluster grouping most closely with other *Sphingomonas* species, and clearly separated from more distantly related AAP taxa such as *P. neustonensis* and *R. rubrum*. Together with the absence of marked GC content differences between the PGC region and chromosomal DNA, and the exclusive clustering of AAP-positive strains within lineage L3 in the whole genome phylogeny, this pattern is consistent with stable long-term maintenance of the PGC within the *S. faeni* AAP lineage, though the mechanism of original acquisition cannot be determined from the available data.

5-ALA synthase genes were found in all *S. faeni* strains used in the study. However, while the XR and none strains encode a single 5-ALA synthase, the AAP strains encode two isoenzymes, from a *hemT*-like gene within the PGC and a *hemA*-like gene elsewhere in the genome. The phylogenetic analysis placed the PGC-encoded gene with *hemT* of *C. sphaeroides* 2.4.1 and the non-PGC gene with *hemA* [71]. In *C. sphaeroides*, *hemA* encodes the primary constitutively expressed ALA synthase responsible for general tetrapyrrole biosynthesis, while *hemT* encodes a secondary isoenzyme with distinct regulatory signals that is expressed only under specific conditions [71]. The 57.2–57.9% amino acid identity between the two genes in *S. faeni* falls within the 53–65% identity reported between *hemA* and *hemT* in *C. sphaeroides* [71]. By analogy, the *hemT*-like positioning

of the PGC-encoded gene in *S. faeni* suggests a dedicated role in bacteriochlorophyll biosynthesis within the photosynthetic machinery, while the non-PGC *hemA*-like gene maintains general cellular tetrapyrrole and heme biosynthesis. Heme, being an essential component in various proteins, including those integral to photosynthesis and electron transport [74], implicates the necessity for more robust biosynthetic pathways in these phototrophs. The single 5-ALA synthase gene 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 [75].

Our photosensor analysis revealed here and in our previous study the wider need for photosensing among phototrophs [29]. 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 lineages in phylogeny grouping, but direct functional properties for particular 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* [76].

On the other hand, one BphP was associated with sodium/bile acid symporters in every strain in this study (Supplementary Table 3). Together with the presence of LysR-family transcriptional regulators in these neighborhoods, this co-localization suggests a light-responsive module in which BphP signaling could couple to membrane transport and gene regulation. Such coupling would be consistent with the broader role of photoreceptors in bacterial environmental adaptation [77], including BphP-driven control of solute-transport and stress-response genes reported in the plant-associated phyllosphere bacterium *Pseudomonas syringae* [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 photosynthesis 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 proton-pumping rhodopsin among the strains. Our study enhances the understanding of microbial adaptation and the ecological dynamics of *S. faeni* in boreal, sub-arctic and arctic 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.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s43630-026-00991-0.

**Acknowledgements** The financial support for the study originates from KONE foundation (grant "Jaettu Valo") and Science Council Finland (grant #259180 for RN and # 363660 for JAI). 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.

**Author contributions** 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.

**Funding** Open Access funding provided by University of Jyväskylä (JYU).

**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.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Used Python scripts** All custom Python scripts can be found in this repository. https://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. *Biotechnology Reports*, *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*(69, 2018), 21–49.
6. Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews*, *39*(6), 854–870.
7. Shiba, T., & Simidu, U. (1982). *Erythrobacter longus* gen. nov., sp. nov., an aerobic bacterium which contains bacteriochlorophyll a. *International Journal of Systematic and Evolutionary Microbiology*, *32*(2), 211–217.
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*. https://doi.org/10.1101/2023.02.19.529139
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. Blankenship, R. E. (2002). Photosynthetic pigments: Structure and spectroscopy. In *Molecular mechanisms of photosynthesis* (pp. 42–60).
28. DeLong, E. F., & Béjà, O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLOS Biology*, *8*(4), e1000359.
29. Ihalainen, J. A., et al. (2024). Multifaceted photoreceptor compositions in dual phototrophic systems – a genomic analysis. *Journal of Molecular Biology*, *436*(5), 168412.
30. Park, S. Y., & Tame, J. R. H. (2017). Seeing the light with BLUF proteins. *Biophysical Reviews*, *9*(2), 169–176.
31. 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.

32. Takala, H., et al. (2020). Tips and turns of bacteriophytochrome photoactivation. *Photochemical & Photobiological Sciences*, *19*(11), 1488–1510.
33. Multamäki, E., et al. (2021). Comparative analysis of two paradigm bacteriophytochromes reveals opposite functionalities in two-component signaling. *Nature Communications*, *12*(1), 4394.
34. Losi, A., & Gärtner, W. (2008). Bacterial bilin- and flavin-binding photoreceptors. *Photochemical & Photobiological Sciences*, *7*(10), 1168–1178.
35. Ziegler, T., & Möglich, A. (2015). *Photoreceptor engineering*. *Frontiers in Molecular Biosciences*, *2*.
36. Losi, A., & Gärtner, W. (2021). A light life together: photosensing in the plant microbiota. *Photochemical & Photobiological Sciences*, *20*, 451–473.
37. Rockwell, N. C., & Lagarias, J. C. (2020). Phytochrome evolution in 3D: Deletion, duplication, and diversification. *New Phytologist*, *225*(6), 2283–2300.
38. 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.
39. 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. *International Journal of Systematic and Evolutionary Microbiology*, *53*(Pt 5), 1253–1260.
40. Busse, H. J., Kämpfer, P., & Denner, E. B. (1999). Chemotaxonomic characterisation of Sphingomonas. *Journal of Industrial Microbiology and Biotechnology*, *23*(4–5), 242–251.
41. An, H., et al. (2011). *Sphingomonas xinjiangensis* sp. nov., isolated from desert sand. *International Journal of Systematic and Evolutionary Microbiology*, *61*(Pt 8), 1865–1869.
42. Given, C. (2020). Tissue-specific dynamics in the endophytic bacterial communities in arctic pioneer plant *Oxyria digyna*. *Frontiers in Plant Science*, *11*.
43. 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.
44. 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.
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. *Microbiology Resource Announcements*, *11*(11), e0068722.
46. Ole, F. (2023). NIRis: A low-cost, versatile imaging system for NIR fluorescence detection of phototrophic cell colonies used in science and education. *bioRxiv*. https://doi.org/10.1101/2023.05.31.543100
47. 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.
48. Chen, Y., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*, *7*(1), 1–6.
49. Wick, R. R., et al. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *Plos Computational Biology*, *13*(6), e1005595.
50. Aziz, R. K., et al. (2008). The RAST server: Rapid annotations using subsystems technology. *Bmc Genomics*, *9*(1), 75.
51. Parks, D. H., et al. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055.
52. 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.
53. 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.
54. Waterhouse, A. M., et al. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.
55. 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.
56. 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.
57. Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics*, *27*(7), 1009–1010.
58. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027.
59. 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.
60. Jain, C., et al. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114.
61. Schwengers, O. (2020). Platon: Identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. *Microbial Genomics*, *6*(10).
62. Abramson, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500.
63. The UniProt, C. (2025). UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Research*, *53*(D1), D609–D617.
64. Hallgren, J. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*. https://doi.org/10.1101/2022.04.08.487609
65. Luecke, H., et al. (2008). Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proceedings of the National Academy of Sciences*, *105*(43), 16561–16565.
66. Chazan, A., et al. (2023). Phototrophy by antenna-containing rhodopsin pumps in aquatic environments. *Nature*, *615*(7952), 535–540.
67. Kopejtka, K. (2021). Characterization of the aerobic anoxygenic phototrophic bacterium *Sphingomonas* sp. AAP5. *Microorganisms*, *9*(4).
68. Kopejtka, K., et al. (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), e2211018119.
69. 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.
70. Sasaki, K., et al. (2002). Biosynthesis, biotechnological production and applications of 5-aminolevulinic acid. *Applied Microbiology and Biotechnology*, *58*(1), 23–29.
71. Neidle, E. L., & Kaplan, S. (1993). Expression of the *Rhodobacter sphaeroides* hemA and hemT genes, encoding two

5-aminolevulinic acid synthase isozymes. *Journal of Bacteriology*, *175*(8), 2292–2303.
72. Zheng, Q., et al. (2012). Gain and loss of phototrophic genes revealed by comparison of two Citromicrobium bacterial genomes. *PLoS One*, *7*(4), e35790.
73. Kopejtka, K., et al. (2017). Genomic analysis of the evolution of phototrophy among haloalkaliphilic Rhodobacterales. *Genome Biology and Evolution*, *9*(7), 1950–1962.
74. Layer, G., et al. (2010). Structure and function of enzymes in heme biosynthesis. *Protein Science*, *19*(6), 1137–1161.
75. Jiang, M., et al. (2022). Natural 5-aminolevulinic acid: Sources, biosynthesis, detection and applications. *Frontiers in Bioengineering and Biotechnology*, *10*.
76. Giraud, E., et al. (2002). Bacteriophytochrome controls photosystem synthesis in anoxygenic bacteria. *Nature*, *417*(6885), 202–205.
77. van der Horst, M. A., Key, J., & Hellingwerf, K. J. (2007). Photosensing in chemotrophic, non-phototrophic bacteria: let there be light sensing too. *Trends in Microbiology*, *15*(12), 554–562.
78. Hatfield, B. M., et al. (2023). Light cues induce protective anticipation of environmental water loss in terrestrial bacteria. *Proceedings of the National Academy of Sciences*, *120*(38), e2309632120.
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*.