Reconstruction of bacteriochlorophyll biosynthesis pathways in the filamentous anoxygenic phototrophic bacterium Oscillochloris trichoides DG-6 and evolution of anoxygenic phototrophs of the order Chloroflexales

Denis S. Grouzdev,1,2 Boris B. Kuznetsov,1 Olga I. Keppen,2 Elena N. Krasil’nikova,2 Natalia V. Lebedeva2 and Ruslan N. Ivanovsky2

Correspondence
Denis S. Grouzdev
denisgrouzdev@gmail.com

1Bioengineering Center, Russian Academy of Sciences, Moscow, Russia
2Faculty of Biology, Moscow State University, Moscow, Russia

It is commonly accepted that green filamentous anoxygenic phototrophic (FAP) bacteria are the most ancient representatives of phototrophic micro-organisms. Modern FAPs belonging to the order Chloroflexales are divided into two suborders: Chloroflexineae and Roseiflexineae. Representatives of Roseiflexineae lack chlorosomes and synthesize bacteriochlorophyll a, whereas those of Chloroflexineae synthesize bacteriochlorophylls a and c and utilize chlorosomes for light harvesting. Though they constitute a small number of species, FAPs are quite diverse in their physiology. This bacterial group includes autotrophs and heterotrophs, thermophiles and mesophiles, aerobes and anaerobes, occupying both freshwater and halophilic environments. The anaerobic mesophilic autotroph Oscillochloris trichoides DG-6 is still not well studied in its physiology, and its evolutionary origin remains unclear. The goals of this study included identification of the reaction centre type of O. trichoides DG-6, reconstruction of its bacteriochlorophyll biosynthesis pathways, and determination of its evolutionary relationships with other FAPs. By enzymic and genomic analysis, the presence of RCII in O. trichoides DG-6 was demonstrated and the complete gene set involved in biosynthesis of bacteriochlorophylls a and c was established. We found that the bacteriochlorophyll gene sets differed between aerobic and anaerobic FAPs. The aerobic FAP genomes code oxygen-dependent AcsF cyclases, but lack the bchQ/bchR genes, which have been associated with adaptation to low light conditions in the anaerobic FAPs. The aerobic FAP genomes code oxygen-dependent AcsF cyclases, but lack the bchQ/bchR genes, which have been associated with adaptation to low light conditions in the anaerobic FAPs. A scenario of evolution of FAPs belonging to the order Chloroflexales is proposed.

INTRODUCTION

Filamentous anoxygenic phototrophs (FAPs) form a deep branching lineage of the domain Bacteria. On the basis of the results of phylogenetic analysis of 16S rRNA and a wide range of proteins, FAP is proposed to be the earliest branching lineage that involves chlorophototrophs (Blankenship, 2010; Gupta, 2003; Xiong et al., 2000). At the present time, FAPs belonging to the order Chloroflexales are divided into two suborders: Chloroflexineae, which includes families Chloroflexaceae and Oscillochloridaceae, and Roseiflexineae, consisting of family Roseiflexaceae (Gupta et al., 2013). Among the genome-sequenced Chloroflexales, ‘Candidatus Chlorothrix halophila’ (Klappenbach & Pierson, 2004) should be mentioned in spite of the lack of classification of this bacterium within the order. The representatives of the family Chloroflexaceae are thermophilic photo-organotrophs capable of growing photolithothrophically. Besides this, they are able to grow in aerobic chemo-organotrophic conditions (Hanada & Pierson, 2006). Autotrophic CO₂ assimilation by the representatives of family Chloroflexaceae (Chloroflexus aurantiacus) is performed through the 3-hydroxypropionate cycle (Herter et al., 2002; Zarzycki et al., 2009). The representatives of the family Oscillochloridaceae are able to grow as photo-organotrophs and photolithotrophs, and they are strict anaerobes (Keppen et al., 1994, 2000). Autotrophic CO₂ assimilation by the representatives of family Oscillochloridaceae is performed through the reductive pentose phosphate cycle (Berg et al., 2005; Ivanovsky et al., 2006).

Abbreviations: DCPIP, 2,6-dichlorophenolindophenol; FAP, filamentous anoxygenic phototroph(ic); GSB, green sulfur bacterium (bacteria); MV, methyl viologen.

Four supplementary figures and one supplementary table are available with the online Supplementary Material.
1999; Turova et al., 2006). FAPs belonging to the family Roseiflexaceae are photo-organotrophs, and they do not grow photolithotrophically as pure cultures (Hanada et al., 2002), but it has been suggested that Roseiflexus is capable of growing photo-autotrophically in hot spring microbial mats (van der Meer et al., 2005, 2010).

The structures of the photochemical apparatus in FAPs from suborders Chloroflexineae and Roseiflexineae differ from each other. Bacteriochlorophyll c of the light-harvesting antenna complex in the species belonging to Chloroflexineae is located in chlorosomes, which are similar to those of green sulfur bacteria (GSB) (Hanada & Pierson, 2006). The chlorosome in GSB is bound to the reaction centre in the cytoplasmic membrane through the basal plate and the Fenna–Matthews–Olson protein (FMO). The photochemical apparatus of Chloroflexineae lacks the FMO, and the bacteriochlorophyll in the antenna complex is connected to the reaction centre through the basal plate only (Pedersen et al., 2008, 2010). The basal plate contains CsmA protein, which is found in all chlorosome-containing representatives of GSB and FAPs. Each CsmA molecule is bound to one bacteriochlorophyll molecule, and energy transfer is mediated from bacteriochlorophyll to the reaction centre. The molecular masses of CsmA of Chlorobium tepidum, Cfl. aurantiacus and O. trichoides are 6.2, 5.7 and 6.0 kDa, respectively (Bryant et al., 2002, 2012; Frigaard et al., 2003; Frigaard & Bryant, 2004; Montaño et al., 2003; Zobova et al., 2011). The CsmA polypeptide is typically the most abundant chlorosome protein in Cfl. aurantiacus. Roseiflexus castenholtzii, a representative of the Roseiflexineae, lacks chlorosomes, and its photochemical apparatus contains bacteriochlorophyll a only (Collins et al., 2009, 2010; Hanada et al., 2002).

In the quinone reaction centre (RCII) of purple bacteria, an integral membrane protein complex is composed of three proteins: the L, M and H subunits. RCII contains two bacteriochlorophyll a molecules that form a dimer (P), two bacteriochlorophyll a monomers (Bₐ and Bₐₙ), two bacteriopheophytin a molecules (Hₐ and Hₐₙ), two ubiquinone molecules (Qₐ and Qₐₙ) and an iron ion (Blankenship, 2014). In Cfl. aurantiacus, the genes coding for the reaction centre subunits and peripheral light-harvesting polypeptides are located in two puff operons. One operon includes puffLM, coding for the L and M subunits of the reaction centre complex, respectively. The second operon contains puffBAC, coding for the β and α subunits of the light-harvesting complex and the cytochrome subunit of the reaction centre complex (Tang et al., 2011).

Chlorophylls are essential pigments for all phototrophic organisms. Chlorophylls are the product of a long evolutionary development, and proteins involved in bacteriochlorophyll biosynthesis are attractive objects for the reconstruction of the evolution of FAPs. The mechanism of bacteriochlorophyll biosynthesis in GSB has been studied in detail and described in a number of brilliant reviews (Bryant et al., 2012; Chew & Bryant, 2007; Frigaard & Bryant, 2004; Sousa et al., 2013). Various species of GSB synthesize bacteriochlorophylls a, b, c, d, e and g (Chew & Bryant, 2007). FAP species belonging to the suborder Chloroflexineae possess bacteriochlorophylls a and c, whereas species of the Roseiflexineae synthesize bacteriochlorophyll a only (Hanada & Pierson, 2006).

The aim of this study was to analyse the photosynthesis-related genes of the previously sequenced FAP O. trichoides DG-6 (Kuznetsov et al., 2011) followed by reconstruction of the evolutionary relationships between O. trichoides and other FAPs within the order Chloroflexales. The main steps were the following: (1) biochemical identification of the type of reaction centre (RCI or RCII) in O. trichoides DG6; (2) determination of the spatial localization of genes involved in synthesis of the core domain of the reaction centre in the genome of O. trichoides DG6; (3) determination of the spatial localization of genes involved in synthesis of bacteriochlorophylls a and c in the genome of O. trichoides DG6; (4) reconstruction of biosynthesis pathways of bacteriochlorophylls a and c in O. trichoides DG6; and (5) reconstruction of evolutionary relationships between O. trichoides DG6 and other FAPs belonging to the order Chloroflexales based on data from analysis of the spatial localization of photosynthesis-related genes in their genomes and comparative analysis of phylogenetic relationships of corresponding proteins and 16S rRNA genes.

**METHODS**

**Bacterial strains and growth conditions.** The following strains were used: filamentous anoxygenic phototrophic bacterium Cfl. aurantiacus OK-70fl, GSB Chlorobaculum macestae, heliobacterium H. daurensis BT-H1, purple sulfur bacterium Thiocapsa roseopersicina BBS, cyanobacterium Anabaena variabilis Kutz no. 458 and a representative of the family Oscillochloridaceae: O. trichoides DG6. These bacteria were obtained from the Culture Collection of the Department of Microbiology of Moscow State University. The following phototrophic bacteria were grown under light anaerobic conditions in media as described elsewhere: T. roseopersicina (Bogorov, 1974); Chlorobaculum macestae, Chlorobium vibrioforme (Larsen, 1952); Cfl. aurantiacus (Madigan et al., 1974); O. trichoides DG-6 (Keppen et al., 1994); H. daurensis (Bryantseva et al., 1999). A. variabilis was grown aerobically in the light (Kratz & Myers, 1955).

**Studies on whole cells.** Cultures from the late exponential growth phase were harvested, washed with potassium phosphate buffer (50 mM, pH 7.5), and resuspended in the same buffer at a density of 0.1–0.3 mg protein ml⁻¹. The resulting cell suspensions were used in whole-cell experiments.

**Preparation of cell extracts.** Cells were collected and washed and finally resuspended in potassium phosphate buffer (50 mM, pH 7.5) at a density of 5–10 mg protein ml⁻¹. Cell extracts were prepared by sonication at 22 kHZ for 3 min at 4 °C. Debris was removed by centrifugation at 36,000 g for 30 min (4 °C), and the supernatant was used for further experiments.

**Detection of reaction centre activity.** RCI-mediated oxygen uptake was measured polarographically in the presence of 2,6-dichlorophenolindophenol (DCPIP) reduced by sodium ascorbate as electron donor and methyl viologen (MV) as electron acceptor. The assay mixture (1 ml) contained 50 mM potassium phosphate buffer
RESULTS AND DISCUSSION

Spectral analysis of O. trichoides DG-6 bacteriochlorophylls

The photochemical apparatus of all anoxygenic phototrophic bacteria including GSB and FAPs contains bacteriochlorophyll a in the reaction centres. But the structure of their light-harvesting complexes differs among phototrophic bacteria. In various GSB the light-harvesting complex can include bacteriochlorophylls b, c or e (Frigaard & Bryant, 2004; Orf & Blankenship, 2013). Members of the Chloroflexales also differ from each other in the structure of their light-harvesting complexes. In Roseiflexaceae species the antenna complex contains bacteriochlorophyll a only, whereas one of the Chloroflexaceae representatives contains bacteriochlorophyll c (Hanada & Pierson, 2006). Spectral analysis of cell-free extracts revealed that the reaction centre of O. trichoides DG6 contains bacteriochlorophyll a, which has maximal absorbance at 860 nm. The light-harvesting complex of O. trichoides DG6 contains bacteriochlorophyll c, with maximal absorbance at 746 nm. The presence of these pigments was also demonstrated by spectral analysis of methanol extracts. Two absorbance peaks were revealed: at 666 and 770 nm, which are associated with bacteriochlorophylls c and a, respectively (Fig. 1). The molar ratio of bacteriochlorophylls c:a in O. trichoides cells was shown to be ~22:1 for cells grown at medium light intensity (1000 lx) (Taisova et al., 2002).

Identification of the reaction centre of O. trichoides DG-6

Two types of reaction centres are known: type I (RCI) and type II (RCII). RCI is used by green plants, cyanobacteria, Gram-positive heliobacteria and GSB. RCI contains two [4Fe-4S] clusters, which serve as acceptors of electrons. RCII contains quinone as a two-electron acceptor of electrons (Blankenship, 2014; Hillier & Babcock, 2001). RCII is common among oxygenic phototrophs, anoxygenic purple bacteria and FAPs.

The reaction centre type of the FAP O. trichoides (belonging to the family Oscillochloridaceae) has remained unknown since its description in 2000. The reaction centre type of a given bacterium is usually identified by its ability for photodependent reduction of MV. This test is based on the difference in redox potentials of electron acceptors of RCI and RCII. The redox potential of the [4Fe-4S] cluster of RCI ($E_m \approx -550$ mV) is sufficient for the reduction of soluble ferredoxin ($E_m \approx -420$ mV) (Hillier & Babcock, 2001) and MV ($E_m \approx -446$ mV). The redox potential of the electron acceptor of RCII (ubiquinone or menaquinone, $E_m \approx -100$ mV) is not sufficient for the reduction of ferredoxin and MV (Hillier & Babcock, 2001). Data given in Table 1 demonstrate that the reaction centre of O. trichoides is of the quinone type (RCII). All microorganisms possessing RCI (cyanobacterium A. variabilis, heliobacterium H. daurensis and GSB) are capable of light-dependent reduction of MV. The purple bacterium T. roseopersicina and the FAP Cfl. aurantiacus OK-70fl and O. trichoides DG-6 lack this ability. This means that the photochemical apparatus of both O. trichoides and Cfl. aurantiacus contains RCII.
The gene encoding AcsF is absent in the expressed constitutively. The gene encoding AcsF is absent in the study, it was shown that BchE expression is downregulated in aerobic growth conditions (Tang et al., 2004). In the aerobic growth conditions, BchE is responsible for the synthesis of bacteriochlorophyll (Reid & Hunter, 2004). In the aerobic group, the presence of the bchQ/bchR genes is considered as an adaptation of this bacterium to the low illumination in its habitats (Ganapathy et al., 2012; Chew et al., 2007).

Table 1. Light-dependent oxygen consumption by different phototrophic bacteria in the presence of reduced DCPIP as electron donor and MV as electron acceptor

<table>
<thead>
<tr>
<th>Micro-organism</th>
<th>Rate of oxygen consumption [nmol O₂ min⁻¹ (mg protein)⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cell suspensions</td>
</tr>
<tr>
<td>A. variabilis</td>
<td>75.0</td>
</tr>
<tr>
<td>H. daeensis</td>
<td>114.0</td>
</tr>
<tr>
<td>Chlorobium vibrioforme</td>
<td>37</td>
</tr>
<tr>
<td>Chlorobaculum macacae</td>
<td>20.5</td>
</tr>
<tr>
<td>T. roseopersicina</td>
<td>0.0</td>
</tr>
<tr>
<td>Cfl. aurantiacus OK-700</td>
<td>0.0</td>
</tr>
<tr>
<td>O. trichoides DG-6</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Topology analysis of spatial distribution of genes involved in biosynthesis of proteins of the photosynthetic apparatus in the genome of O. trichoides DG-6

Fig. 2 illustrates the spatial distribution of the photosynthesis genes in the genomic DNA of O. trichoides DG-6 and other FAPs of the order Chloroflexales. Two main groups could be distinguished within the genomes tested: the ‘rose’ group and the ‘green’ group. These groups clearly differ from each other in (1) presence/absence of a gene set responsible for the synthesis of bacteriochlorophyll c and (2) patterns of spatial distribution of other photosynthesis genes. The latter shows that two main sets of photosynthesis genes are represented in the genome of Roseiflexus species. The first set includes genes involved in synthesis of chlorophyllide a from protoporphyrin IX and consists of genes bchN-bchB-bchL-bchM-acsF-bchH-bchl. The second gene set includes two gene blocks: one (pufA-pufB-pufLM-pufC) is responsible for the synthesis of light-harvesting antennae and reaction centres, and the second (bchG-bchP-bchF-bchC-bchX-bchY-bchZ) for synthesis of bacteriochlorophyll a from chlorophyllide a. In the ‘green’ group, the corresponding genes are spread between several smaller blocks. The spread of FAP genomes into two groups is supported also by the fact that in the ‘rose’ group PufL and PufM proteins are encoded by a single pufLM gene, whereas in the ‘green’ group PufL and PufM are encoded by separate pufL and pufM genes.

Additionally, the FAPs studied could be separated into an aerobic and an anaerobic group. In the aerobic group, which includes Chloroflexus and Roseiflexus species, the reaction of Mg-protoporphyrin monomethyl ester cyclization is performed by two different cyclases: BchE, which was proposed to operate in anaerobic conditions, and AcsF, operating in aerobic growth conditions (Tang et al., 2009). In the same study, it was shown that BchE expression is downregulated under semi-aerobic conditions, whereas AcsF seems to be expressed constitutively. The gene encoding AcsF is absent in the O. trichoides and ‘Candidatus Chlorothrix halophila’ genomes. The presence of AcsF in the genome correlates with the oxygen requirement of these organisms: O. trichoides and ‘Candidatus Chlorothrix halophila’ are strict anaerobes, whereas Chloroflexus and Roseiflexus species are able to grow under both semi-aerobic and anaerobic conditions.

An additional feature of anaerobic bacteria belonging to the ‘green’ group is the presence in the genome of the bchQ/bchR genes, which are lacking in the Chloroflexus and Roseiflexus genomes. It was shown that knockout of the homologous genes in the GSB Chlorobaculum tepidum resulted in decrease in light harvesting caused by alteration of the orientation of the bacteriochlorophyll molecule along the chlorosome axis. Thus, the presence of the bchQ/bchR genes is considered as an adaptation of this bacterium to the low illumination in its habitats (Ganapathy et al., 2012; Chew et al., 2007).

Reconstruction of the biosynthesis of bacteriochlorophylls a and c in O. trichoides

Chlorophyllide a synthesis. The metabolic precursor for all bacteriochlorophylls is protoporphyrin IX. The reaction chain resulting in bacteriochlorophyll a and c starts with the incorporation of Mg into the protoporphyrin IX molecule. This reaction is catalysed by an ATP-dependent Mg-chelatase. This enzyme consists of three subunits, which are encoded by the bchl, bchD and bchH genes (Table S1, available in the online Supplementary Material; Fig. 3). Here, the bchl and bchD genes encode two small chelatase subunits, and bchH encodes a large one. The chelating reaction is performed in two ATP-dependent stages: activation and chelation. Containing ATP and Mg²⁺ binding sites, subunits I and D, in the presence of ATP and Mg²⁺, form hepta- and hexameric complexes, respectively, and these oligomers then form a multi-enzyme complex together with the H subunit, which contains the protoporphyrin IX binding site. The resulting complex performs the chelation. To synthesize one molecule of Mg-protoporphyrin IX, about 15 ATP molecules are required (Reid & Hunter, 2004). In the O. trichoides genome, three copies of bchH are represented (Fig. 2).

The next step is methylation of the carboxyl group of the propionate in position C-13 of the Mg-protoporphyrin, resulting in the formation of Mg-protoporphyrin monomethyl ester. This reaction is catalysed by Mg-protoporphyrin methyltransferase, encoded by the bchM gene. In the O. trichoides genome, bchM and bchH, the latter encoding the chelatase large subunit, are located in the same gene block (Fig. 2). During the next step in the synthesis, the C-13 propionic acid chain is oxidized and cyclized to form the isocyclic, fifth ring of the bacteriochlorophyll molecule. The final product of this reaction is 3,8-divinyl protochlorophyllide. In O. trichoides, this reaction is catalysed by anaerobic Mg-protoporphyrin monomethyl esterase, BchE. In Chloroflexus and Roseiflexus species, this reaction is catalysed by two different enzymes: anaerobic (BchE) and aerobic (AcsF) cyclases; the gene encoding the latter is lacking in the
Fig. 2. Spatial organization of photosynthesis genes in the genome of FAPs belonging to the order Chloroflexales. Genes involved in different synthetic processes are given in different colours.
O. trichoides DG-6 genome. This feature corresponds to the difference in physiology of these bacteria: Chloroflexus and Roseiflexus species are able to grow in both aerobic and anaerobic environments, whereas O. trichoides is a strict anaerobe. This feature, along with the lack of bchQ/bchR genes (see below), is the only difference in the bacteriochlorophyll biosynthesis pathways between aerobic and anaerobic ‘green’ FAP species. It should be mentioned that in all FAP genomes the bchM and bchH genes are co-localized in one gene block, but in the aerobic FAP genomes they are separated by the acsf gene (Fig. 2).

The C-8 vinyl group of 3,8-divinyl protochlorophyllide a is reduced in all ‘green’ FAPs by the reductase encoded by the bciB gene (Liu & Bryant, 2011). The product of this reaction is protochlorophyllide a.

The double bond in 3,8-divinyl protochlorophyllide a (position C-17/C-18 in porphyrin ring D) is reduced by a multi-enzyme complex encoded by the bchl, bchB and bchn genes in all FAP genomes. This reaction also involves ATP and a reducing agent (NADH or ferredoxin). The final product is chlorophyllide a. All the genes – bchl, bchB and bchn – in all FAP genomes are situated within a single gene block (Fig. 2). Chlorophyllide a is a substrate for subsequent synthesis of both bacteriochlorophylls a and c, but the synthesis pathways are separate.

**Bacteriochlorophyll a synthesis**

Bacteriochlorophyll a synthesis from chlorophyllide a starts with the reduction of the C-7/C-8 double bond in porphyrin ring B. Like the reduction of the C-17/C-18 double bond in porphyrin ring D, this reaction is catalysed by a nitrogenase (Nif)-like multi-enzyme complex chlorophyllide reductase encoded by the bchX, bchY and bchZ genes. The product of this reaction is 3-vinyl bacteriochlorophyllide a. The bchX, bchY and bchZ genes, apart from the 3-protochlorophyllide reductase-encoding genes, are spread between two gene blocks: one including bchY and bchZ and the second including bchX together with bchC and bchF (Fig. 2).

The product of the reduction reaction – 3-vinyl bacteriochlorophyllide a – is further converted to hydroxymethyl bacteriochlorophyllide a through hydroxylation of the C-3 vinyl group of ring A with the enzyme encoded by bchF. Then the oxidation of the C-3 hydroxyl group to a ketone group by the hydroxylase encoded by the bchC gene results in bacteriochlorophyllide a, the precursor of bacteriochlorophyll a. The bchF and bchC genes are located in the same gene block with the bchI and bchD genes, which encode the I and D subunits of the porphyrin IX chelatase.

The last stage in bacteriochlorophyll a synthesis is the attachment of geranyl-geranyl pyrophosphate to bacteriochlorophyllide a and the reduction of the latter to phytol. The bchG and bchP genes encoding the corresponding enzymes are co-localized in the O. trichoides genome in one gene block with the pufA and pufB genes, which encode the α- and β-subunits of the LH1 antenna complex. This gene block also includes the pufC gene encoding the C-subunit, which contains tetrahaem cytochrome c’ (Fig. 2).

**Bacteriochlorophyll c biosynthesis**

At the first stage of bacteriochlorophyll c biosynthesis, the C-13'-methylcarboxyl group is removed from chlorophyllide a, resulting in the formation of [3V,8-E,12M] bacteriochlorophyllide d (Fig. 3). This reaction is catalysed by the protein of gene bciC. Then, the C-3 vinyl group of [3V,8-E,12M] bacteriochlorophyllide d is hydrated by 3-vinyl hydratase (BchF) and methylated at positions C-82 and C-124 by proteins encoded by the bchQ and bchR genes. The bchF gene is located in the same block with bchX, bchI and bchD, whereas bchQ and bchR are located in a separate gene block in the O. trichoides and ‘Candidatus Chlorothrix halophila’ genomes (Fig. 2). R-hydroxymethyl bacteriochlorophyllide d, with the participation of C-20 methyltransferase (BchU), forms bacteriochlorophyllide c, which is the direct precursor of bacteriochlorophyll c. Bacteriochlorophyllide c is etherified with farnesol by bacteriochlorophyll c synthase (BchK). The bchK gene in the genome of the ‘green’ group FAPs is located in one
gene block with the \textit{bchU} gene (Fig. 2). In \textit{O. trichoides} there is an additional separate copy of the \textit{bchK} gene.

**Phylogenetic analysis of FAPs**

We based the phylogenetic reconstruction on the following widely accepted methods (House, 2009): (1) comparison of single gene or protein sequences, for example 16S rRNA; (2) comparison of concatenated gene or protein sequences, usually those related to a specific physiological function, for example, bacteriochlorophyll biosynthesis; (3) comparison of whole genome sequences; (4) comparison of a specific gene block or operon structure in the genome.

The most common molecular marker for phylogenetic analysis is 16S rRNA. All 16S rRNA FAP genomic sequences (i.e. taken from WGS) were co-localized within two big clusters: the ‘rose’ cluster, including representatives of the family \textit{Roseiflexaceae}, and the ‘green’ cluster consisting of representatives of the families \textit{Chloroflexaceae} and \textit{Oscillochloridaceae} (Fig. 4). It should be mentioned that the ‘\textit{Candidatus Chlorothrix halophila}’ sequence clustered with \textit{O. trichoides} DG-6, but with a low bootstrap value (67%). According to this pattern, ‘\textit{Candidatus Chlorothrix halophila}’ could be assigned to the family \textit{Oscillochloridaceae}.

However, the dendrogram calculated after comparison of concatenated PufC, PufL and PufM amino acid sequences revealed an alternative clustering in which the ‘\textit{Candidatus Chlorothrix halophila}’ sequence formed a separate branch at a position intermediate between the ‘rose’ and ‘green’ groups (Fig. 5).

A similar clustering pattern for \textit{Chlorothrix} sequences was also observed in trees reconstructed on the basis of concatenated protein sequences for enzymes involved in bacteriochlorophyll synthesis and for the derived amino acid sequences of housekeeping genes.

The bacteriochlorophyll synthesis trees were reconstructed for three protein sets participating in: (1) synthesis of chlorophyllide \textit{a} from protoporphyrin IX (BchB, BchD, BchE, BchH, BchL, BchM and BchN proteins; Fig. S1); (2) synthesis of bacteriochlorophyll \textit{a} from chlorophyllide \textit{a} (BchC, BchF, BchG, BchP, BchX, BchY and BchZ proteins; Fig. S2); and (3) synthesis of bacteriochlorophyll \textit{c} from chlorophyllide \textit{a} (BchK, BchU and BciC proteins; Fig. S3).

To compare the tree topologies of photosynthesis-associated proteins and proteins responsible for alternative activity, the dendrograms based on concatenated amino acid sequences have commonly been used (Gupta \textit{et al.}, 2013). The phylogenetic tree for the housekeeping proteins (DnaB, DnaE, GyrA, GyrB and SecY) was reconstructed (Fig. S4). This tree revealed the same topology as all other protein-based dendrograms.

We propose that the dendrogram based on the 16S rRNA gene sequence comparison reflects the evolutionary processes between FAPs more adequately than the protein-based trees. The separate branching of ‘\textit{Candidatus Chlorothrix halophila}’ on protein trees is obviously the
consequence of adaptation of its protein structure to the increased salinity of the \textit{Chlorothrix} habitat. The phenomenon of adaptive variability has been described in a number of studies devoted to comparison of protein sequences of halophilic micro-organisms with freshwater ones (for example, see Paul et al., 2008). We performed an analogous comparison of the frequencies of various amino acids in FAP proteins versus \textit{E. coli}, and this parameter showed that ‘\textit{Candidatus Chlorothrix halophila}’ differs substantially from other FAPs (Fig. 6). \textit{Chlorothrix} proteins (more than 3000 proteins tested) show quite distinct usage of amino acid residues compared with non-halophilic FAPs. The frequencies of amino acid usage in \textit{Chlorothrix} were similar to those of the halophilic bacterium \textit{S. ruber}. Among the prominent trends are significant increase in Asp, Glu and Thr amino acid residues and decrease in Lys and strong hydrophobic residues (Ile, Met, Leu) are also consistent with earlier reports (Elcock & McCammon, 1998; Kennedy et al., 2001; Lanyi, 1974).

This means that the rate of evolutionary changes in 16S rRNA and protein-coding genes could differ substantially in one organism. This fact is supported also by the similarity levels of corresponding sequences shown in Table S1 and correlates with the data obtained by Bromham & Penny (2003).

An additional phenomenon influencing the results of phylogenetic reconstruction is the so-called ‘attraction of long branches’ (Felsenstein, 1996), which also could disfigure the branching order of phylogenetic trees based on the comparison of nucleotide sequences of 16S rRNA genes and amino acid sequences of the proteins.

**Fig. 6.** FAP grouping according to their standardized amino acid usage. Standardized amino acid composition of halophiles and non-halophiles are grouped by unweighted pair group mean clustering. The left panel depicts the unweighted pair group mean clustering based on the relative abundances of different amino acid residues in the encoded proteins of organisms with respect to those of \textit{E. coli}. The distance in the clustering is the Euclidean distance. The right panel is a pictorial representation of relative amino acid usage in the respective organisms.

**Fig. 7.** Scheme of FAP evolution within the order \textit{Chloroflexales}. Bifurcation points where new gene sets appeared in the genome and corresponding metabolic changes occurred are marked with circles. A, FAP ancestor; 1, appearance of chlorosomes and bacteriochlorophyll \textit{c} biosynthesis pathway; 2, acquisition of \textit{bchQ/bchR} genes, \textit{acsF} gene loss, adaptation to anaerobic conditions; 3, adaptation to salinity.
Taking together all the results, the following scenario of evolution could be proposed for FAPs belonging to the order Chloroflexales (Fig. 7). The ancestral FAP species could be a photoheterotrophic micro-organism whose photosystem lacked chlorosomes and contained only bacteriochlorophyll a (Larkum, 2006). The division of FAPs into ‘rose’ and ‘green’ groups resulted from acquisition of a gene set responsible for chlorosome formation (csmA genes) and bacteriochlorophyll c biosynthesis (event 1 in Fig. 7). The species belonging to the suborder Roseiflexineae retained the ancestral features (chlorosome absence, bacteriochlorophyll a as a single pigment of light-harvesting antennae, and photoheterotrophic lifestyle), whereas representatives of the suborder Chloroflexineae changed their photochemical apparatus to that specific to modern GSB (chlorosomes, bacteriochlorophylls a and c, connection of bacteriochlorophyll c with reaction centre through the basal plate). The subsequent evolutionary events affected bacteriochlorophyll c synthesis. The next step was the appearance of mesophilic FAPs adapted to anaerobic niches. At this stage the new lineage lacked the acsF gene and acquired bchQ/bchR gene block (event 2 in Fig. 7).

Later, the Chlororhizia species emerged as the result of adaptation of mesophilic autotrophic ‘green’ FAPs to increased salinity of the environment (event 3 in Fig. 7).

**CONCLUSIONS**

The main results of this study are the following. (1) It was shown that the FAP Oscillochloris trichoides DG-6 has a reaction centre of type II. (2) The bacteriochlorophyll biosynthesis pathways in Oscillochloris trichoides DG-6 were reconstructed based on the gene set found in its genome. (3) An evolutionary scenario within FAPs belonging to the order Chloroflexales was proposed.

**ACKNOWLEDGEMENTS**

This work was supported by the Russian Foundation for Basic Research (grant no. 11-04-01184-a) and a Russian Federation President grant for the support of leading scientific schools (NSH6150.2014.4).

**REFERENCES**


Gupta, R. S., Chander, P. & George, S. (2013). Phylogenetic framework and molecular signatures for the class Chloroflexi and its different clades; proposal for division of the class Chloroflexi class. nov. [corrected] into the suborder Chloroflexineae subord. nov., consisting of the emended family Oscillochloridaceae and the family Chloroflexaceae fam. nov., and the suborder Roseiflexineae subord. nov., containing the family Roseiflexaceae fam. nov. *Antoine van Leeuwenhoek* 103, 99–119.


Edited by: A. Wilde