



HAL
open science

MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution

Dominique Boeuf, Stéphane Audic, Loraine Brillet-Guéguen, Christophe Caron, Christian Jeanthon

► To cite this version:

Dominique Boeuf, Stéphane Audic, Loraine Brillet-Guéguen, Christophe Caron, Christian Jeanthon. MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution. Database - The journal of Biological Databases and Curation, 2015, 2015, pp. bav080. 10.1093/database/bav080 . hal-01218501

HAL Id: hal-01218501

<https://hal.science/hal-01218501>

Submitted on 15 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Original article

MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution

Dominique Boeuf^{1,2}, Stéphane Audic^{2,3}, Loraine Brillet-Guéguen⁴,
Christophe Caron⁴ and Christian Jeanthon^{1,2,*}

¹CNRS, UMR 7144, Marine Phototrophic Prokaryotes Team, ²Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Oceanic Plankton Group, ³CNRS, UMR 7144, Team Evolution des Protistes et Ecosystèmes Pélagiques and ⁴CNRS, UPMC, FR2424, ABiMS, Station Biologique de Roscoff, F-29680 Roscoff, France

*Corresponding author: Tel: +33 298 292 563; Fax: + 33 298 292 324; E-mail: jeanthon@sb-roscoff.fr

Citation details: Boeuf,D., Audic,S., Brillet-Guéguen,L., *et al.* MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution. *Database* (2015) Vol. 2015: article ID bav080; doi:10.1093/database/bav080

Received 16 April 2015; Revised 17 July 2015; Accepted 25 July 2015

Abstract

Microbial rhodopsins are a diverse group of photoactive transmembrane proteins found in all three domains of life and in viruses. Today, microbial rhodopsin research is a flourishing research field in which new understandings of rhodopsin diversity, function and evolution are contributing to broader microbiological and molecular knowledge. Here, we describe MicRhoDE, a comprehensive, high-quality and freely accessible database that facilitates analysis of the diversity and evolution of microbial rhodopsins. Rhodopsin sequences isolated from a vast array of marine and terrestrial environments were manually collected and curated. To each rhodopsin sequence are associated related metadata, including predicted spectral tuning of the protein, putative activity and function, taxonomy for sequences that can be linked to a 16S rRNA gene, sampling date and location, and supporting literature. The database currently covers 7857 aligned sequences from more than 450 environmental samples or organisms. Based on a robust phylogenetic analysis, we introduce an operational classification system with multiple phylogenetic levels ranging from superclusters to species-level operational taxonomic units. An integrated pipeline for online sequence alignment and phylogenetic tree construction is also provided. With a user-friendly interface and integrated online bioinformatics tools, this unique resource should be highly valuable for upcoming studies of the biogeography, diversity, distribution and evolution of microbial rhodopsins.

Database URL: <http://micrhode.sb-roscoff.fr>.

Introduction

Rhodopsins are photochemically active membrane proteins that are composed of seven transmembrane helices with a retinal chromophore. According to their amino acid sequences, they are divided in two families known as either type-1 rhodopsins, that are all of microbial origin or type-2 rhodopsins that are animal photosensitive receptors (1). Type-1 rhodopsins include light-driven proton pumps (e.g. bacteriorhodopsins and proteorhodopsins), ion pumps and channels, and light sensors. The first identified microbial rhodopsin, bacteriorhodopsin, was discovered from the cell membrane of the halophilic archaeon *Halobacterium salinarum* more than 40 years ago (2). Rhodopsins functioning as light-driven chloride pumps (halorhodopsins) with positive and negative phototactic sensors (sensory rhodopsins I, II and III) were further found in the same organism (3–5).

In 2000, a survey of total community DNA from Monterey Bay surface waters led to the discovery of a novel type of bacterial rhodopsin found in an uncultured marine gammaproteobacterium (6). Proteorhodopsin-mediated phototrophy is now known in a large variety of Bacteria and Archaea from diverse environments and lateral gene transfer most likely played an important role in their wide distribution across marine prokaryotes (7). Proteorhodopsin-containing microorganisms are widespread in terrestrial (soils, crusts, phyllosphere), freshwater (lakes, rivers, ponds, ice) and marine (including sea ice, hypersaline and brackish) photic environments (8–11). Recently, proteorhodopsin homologs were also detected in giant viruses that infect unicellular aquatic eukaryotes (12, 13). Proteorhodopsin acts as a proton pump (6, 14, 15) and could be involved as a secondary source of energy in the metabolism of heterotrophic prokaryotes through ATP generation (16, 17). Based on the analysis of marine and terrestrial metagenomic data, Finkel *et al.* (18) suggested that microbial rhodopsins are the prominent phototrophic mechanism on Earth. However, more investigations are needed to understand the physiological functions and fitness benefits of these proteins and their actual role in microbial ecology and in energetic balance of ecosystems.

In recent years, environmental genomics surveys have increasingly demonstrated the remarkable diversity of microbial rhodopsins in diverse aquatic and terrestrial environments (8–10, 12, 13, 19–31). Most of these studies have been performed by using the proteorhodopsin gene as molecular marker. Analyses of microbial gene sequences that serve as markers are facilitated by the availability of annotated databases of aligned sequences. Aligned sequences are required for diversity and phylogenetic analyses and for the design and evaluation of polymerase chain reaction

(PCR) primers and probes. Group-specific PCR primers are used in quantitative real-time PCR for the quantification of gene copy numbers in the environment and for expression studies.

Here, we present MicRhoDE, a comprehensive, high-quality and freely accessible resource of nucleic acid sequences coding for microbial rhodopsins. The database and its associated description will be useful for studying the diversity, phylogeny and evolution of rhodopsin-containing microorganisms.

Data collection and curation

The MicRhoDE database was initially constructed by extracting reference proteorhodopsin sequences from GenBank (32), Global Ocean Sampling (GOS) database obtained from the CAMERA website (<http://camera.crbs.ucsd.edu/>) and from the literature (Figure 1). This initial set was further complemented with other type-1 rhodopsins (actinorhodopsins, xanthorhodopsins, bacteriorhodopsins, halorhodopsins and sensory rhodopsins) and newly discovered types (33, 34). To this initial set of sequences was added an original dataset (ProteoRhodopsin Global Diversity, PRGD) of marine proteorhodopsin genes obtained by Illumina sequencing of amplicons from diverse marine regions. The whole dataset was then used as a diversified seed to perform exhaustive similarity searches using BLAST in GenBank (as of March 2013) or GOS databases. BLAST results were de-duplicated and manually checked for quality. Finally, all reference nucleic acid sequences were manually curated and modified from their original deposits in GenBank and GOS databases, when necessary, to be all in the same open reading frame. Because all MicRhoDE sequences are also stored in GenBank and GOS databases, NCBI or JCVI record IDs are also available in MicRhoDE to keep track of the original data source.

To date, 7857 type-1 rhodopsin sequences are stored in MicRhoDE, most of which (7193 sequences) represent proteorhodopsins. Although the majority of sequences are derived from environmental surveys of proteorhodopsin genes, the database also contains sequences obtained from a range of isolates or large genomic DNA fragments bearing a 16S rRNA gene copy. Among the 295 sequences whose taxonomic affiliation can be inferred from a 16S rRNA gene, 186 sequences come from cultivated organisms. Most sequences come from marine (93%) or freshwater environments (6%).

Alignment and phylogenetic affiliation

Although lateral gene transfers and duplication events are prominent processes for the diversification of microbial

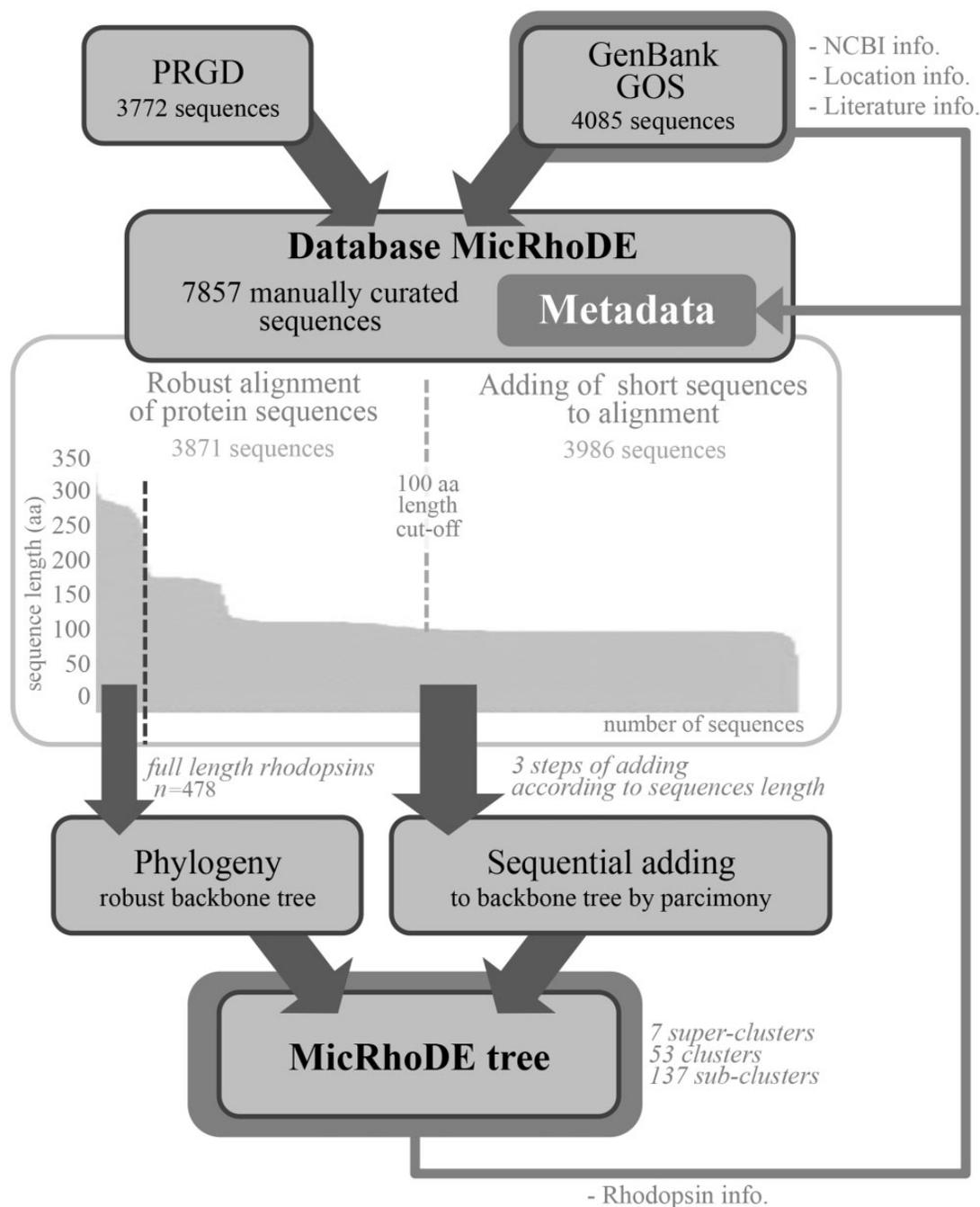


Figure 1. Flowchart of data in the MicRhoDE database. Arrows indicate sequence and metadata flows.

genes among the three domains of life, we constructed a reference phylogenetic tree of microbial rhodopsins to allow a presumptive classification. Despite type-1 and type-2 rhodopsins share structural and functional similarities, there is a very low sequence identity between these two families (1). The seven transmembrane α -helices form a pocket in which the retinal, a vitamin-A aldehyde chromophore, is bound to a lysine residue by a *Schiff*-base linkage (35). This structure implies that evolutionary

constraints vary according to the protein region. As a consequence, putative structure of the protein has to be considered when nucleic acid sequences are aligned. Since aligning 7857 nucleic acid sequences according to the secondary structure of the corresponding proteins is time-consuming, the sequence dataset was split in two parts (Figure 1). The 3871 longest amino acid sequences (>100 amino acid residues) were aligned according to the protein secondary structure using MAFFT eINSi strategy (36). The

shorter ones were added to the robust alignment using MAFFT FFT strategy with the ‘-addfragments’ option that conserves the original alignment. The 478 full-length type-1 rhodopsin sequences of the database, including 86 strains and 73 different species, were used to construct a robust backbone tree (Figure 1) by Bayesian inference (4 Markov Chain Monte Carlo chains of 150 million generations) using PhyloBayes software (37). Shorter sequences were then sequentially inserted into the backbone tree by using the parsimony add option of the ARB software (38). The resulting tree (Figure 2) allowed us to establish a comprehensive classification system consisting in 5 superclusters, 53 clusters and 137 subclusters.

Available metadata

The associated metadata of each sequence such as the sampling date, location, biome of origin, oceanic province were extracted from the related literature, GenBank and GOS records, checked manually, and reconciled before importation in the database. For each sequence are also provided the position in the phylogenetic tree, its NCBI taxonomy when available, the type of rhodopsin, its predicted spectral tuning according to the amino acid residue at position 105 for proteorhodopsins (15, 39) (Supplementary Figure S1). Putative activity and function according to the residues at position 97 and 108, respectively (40) and residue 101 for flavobacterial NQ rhodopsins (34) are also indicated. Altogether, the diversity of metadata associated to aligned and unaligned nucleic acid and protein sequences allows a variety of search options and data outputs.

MicRhoDE web interface

MicRhoDE is a freely accessible public database (<http://micrhode.sb-roscoff.fr>) implemented using the perl Catalyst web framework (<http://www.catalystframework.org/>) backed by a PostgreSQL database (<http://www.postgresql.org/>). MicRhoDE will be updated annually by adding new type-1 rhodopsin gene sequences. In addition to a short introduction to microbial rhodopsins, the homepage provides a brief description of the database content and clickable icons with direct links to the major utilities of the database, including database and sequence similarity searching, and phylogeny (Figure 3a). MicRhoDE provides a powerful search module that accepts complex searches of given taxonomy, predicted protein features (such as activity, function and spectral tuning) and of a range of features (Figure 3b). Using a menu list, filters are also accessible for combining searches at the different cluster levels and features such as e.g. taxonomy, marine province of origin and

predicted spectral tuning. Sequence similarity searches within the database are available using BLAST (version 2.2.26+) submission form in the BLAST page (Figure 3c). Metadata available in other public databases (accession ID, NCBI taxonomy, location, biome, marine province, date of isolation and related literature) and other restricted to MicRhoDE (rhodopsin affiliation according to phylogeny, rhodopsin type, predicted spectral tuning, putative activity and function) are optionally accessible in the data outputs of both search and BLAST modules (Figure 3d). Available data outputs include visualization of results on a map (Figure 3e and f).

The phylogeny page provides three different items: (i) a software pipeline proposing the user to place its own sequences into the type-1 rhodopsin reference phylogenetic tree, (ii) a schematic representation of the reference phylogenetic tree, highlighting the classification in clusters and superclusters and (iii) a detailed reference phylogenetic tree, displayed using the Archaeopteryx phylogenetic tree viewer Java applet (41). To place query amino acid sequences in the reference phylogeny, the user is redirected to a dedicated Galaxy instance (42–44) where the MicRhoDE workflow performs phylogenetic placement using Bayesian inference as implemented in the pplacer software (45). The Galaxy instance (Figure 3g) is available at http://webtools.sb-roscoff.fr/root?tool_id=abims_micrhode_workflow. Output files are visualized using the guppy program (a companion program of pplacer). Guppy (http://matsen.github.io/pplacer/generated_rst/guppy.html#) generates the phylogenetic tree showing either the probability of placements (fat visualization) or the best placements (tog visualization) of query sequences. The Galaxy framework provides interoperability mechanisms to dynamically call external viewer. Trees are generated in the phyloXML format and displayed using the Archaeopteryx phylogenetic tree viewer java applet (41) (Figure 3h).

To provide an intuitive overview of the geographic distribution of current data, the Map page displays for each location, the number of sequences available in MicRhoDE, the actual number of superclusters and clusters according to phylogeny and the dominant ones as well as the proportion of predicted spectral variants. The download page allows the download of the raw and aligned sequences of the database, their associated metadata, phylogenetic trees as well as a complete version of MicRhoDE formatted for the ARB software.

Conclusion

MicRhoDE is a specialized database devoted to the study of microbial rhodopsins, which are functionally versatile proteins of crucial importance in the ecology of terrestrial

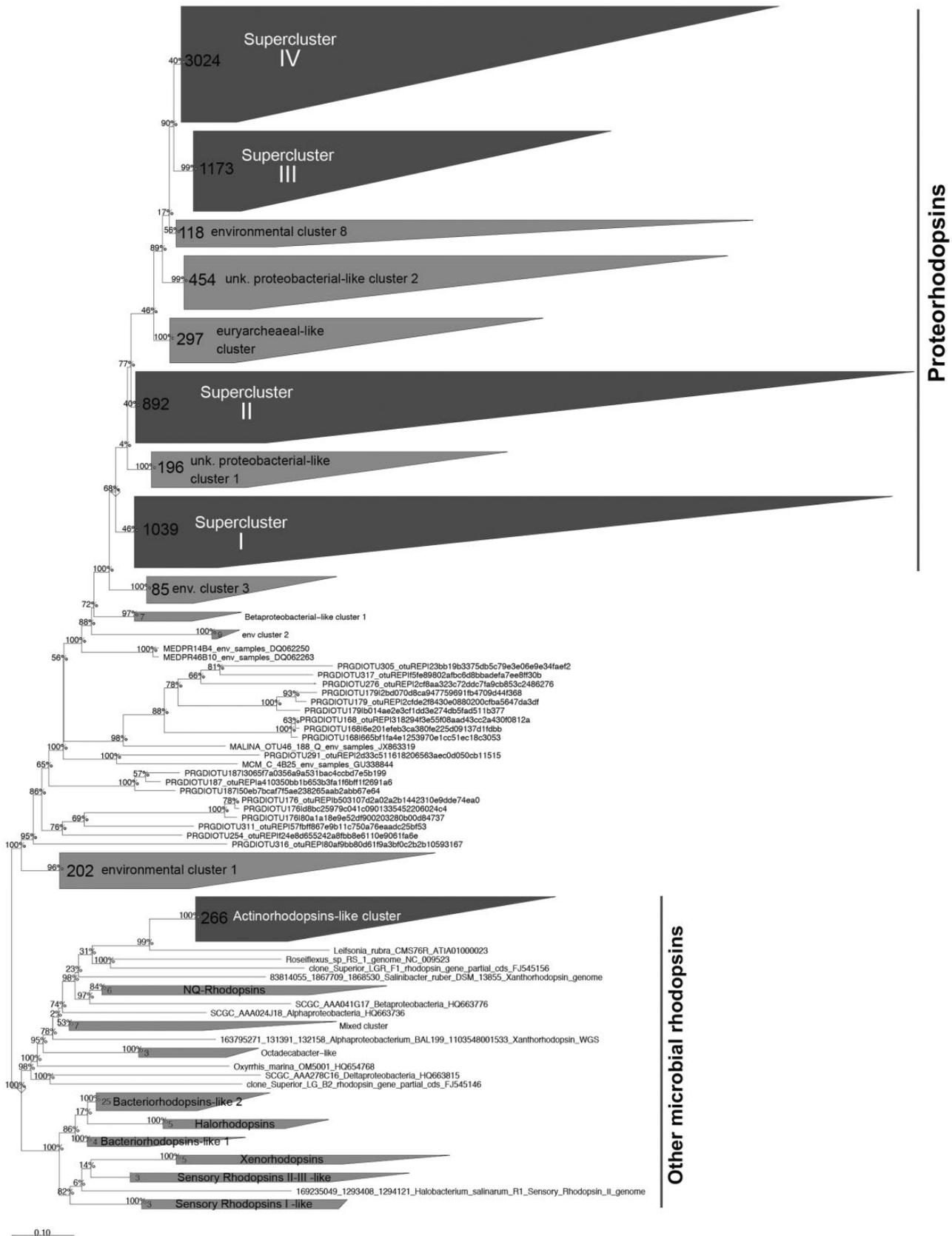


Figure 2. Phylogenetic relationships between the microbial rhodopsins stored in the MicRhoDe database. Numbers at the nodes are bootstrap values obtained by maximum parsimony. Numbers in clusters indicate the number of affiliated sequences.

a MICRHODE
Microbial Rhodopsin database for Diversity and Evolution studies

b AND OR + Add rule + Add group
supercluster equal proteorhodopsin super-cluster IV Delete

c AND OR + Add rule + Add group + Add group + Add group Delete
predicted_spectral_tuning equal Blue (Q105) Delete

d Enter query sequence (fasta) or accession number:
>Bacterium HTCC7216_EF616635
RYIDWLITVPLQMVFEFYLIILSAVGKANSQMFWRLLLSVVMVLVGGYLGEAGYINATLGFIIIGMAGWVYIYE
VFSGEAGKAAAKSGNKALVTAFGAMRMIVTVGWAIIYPLGYVFGYLTGGVDAES

e blastN (nucl Query vs nucl DB) All Micrhode Only cultivated strains
METADATA (optional):
 Location informations (GPS, province, biome, and isolation source)
 Rhodopsin informations (rhodopsin clade, predicted spectral tuning)
 NCBI informations (taxonomy & dbxref, mol type, dates of publication & isolation, and note)
 Literature informations (title, authors, journal, and pubmed id)
 Sequences (DNA & protein, raw & aligned)
To generate a fasta file, please select type of sequence
 DNA Raw Short name
 Protein Aligned Full name (NCBI)
Blast output in browser tab Blast output in tab separated file Get matching fasta sequences Map sequences

f This is the MAP module
The map shows all localities where sequences corresponding to your criteria are found.

| Taxid | Count |
|---------------------------|-------|
| [III-Alpha2-SAR11/Himb59] | 2 |
| [IV-Alpha8-18B2] | 1 |
| [IV-Alpha8-a] | 1 |
| [IV-Gamma5-SAR86a] | 1 |

g Galaxy / ABIMS public
Tools
search tools
Get Data
MICRHODE WORKFLOW
MicRhoDE workflow
STEP 1
Upload File from your computer
STEP 2
MicRhoDE workflow Places query sequences on a fixed reference phylogenetic tree according to a

h A job has been successfully added to the queue – resulting in the following datasets:
6: guppy fat tree
7: guppy tog tree
8: visualize guppy trees
You can check the status of queued jobs and view the resulting data by refreshing the History. The job has been run the status will change from 'running' to 'finished' if completed successfully if problems were encountered.

Figure 3. Screenshots of the MicRhoDE web interface showing the main content panel (a), the search (b) and (c) forms, the metadata (d) and output (e) options, a view of the map output option (f), the Galaxy instance for phylogenetic analysis (g) and an example of phylogenetic tree output (h).

and aquatic photic environments. As microbiologists from all fields use molecular, genomic and metagenomic methods to look at microbial diversity in the biosphere in more breadth and depth, we anticipate that the release of MicRhoDE will help comprehensive ecological and evolutionary analyses of these cosmopolitan genes.

Supplementary Data

Supplementary data are available at *Database Online*.

Acknowledgements

The authors thank Gregory Farrant and Frédéric Mahé for their help.

Funding

This work was supported by grants from the Agence Nationale de la Recherche [grant no. ANR 11 BSV7 021 02] and from the European Union's Seventh Framework Programme [grant no. 287589]. Funding for open access charge: European Union's Seventh Framework Programme [grant no. 287589].

Conflict of interest. None declared.

References

- Spudich, J.L., Yang, C.S., Jung, K.H. *et al.* (2000) Retinylidene proteins: structures and functions from archaea to humans. *Ann. Rev. Cell Dev. Biol.*, **16**, 365–392.
- Oesterhelt, D. and Stoekenius, W. (1971) Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature*, **233**, 149–152.
- Matsuno-Yagi, A. and Mukohata, Y. (1977) Two possible roles of bacteriorhodopsin; a comparative study of strains of *Halobacterium halobium* differing in pigmentation. *Biochem. Biophys. Res. Comm.*, **78**, 237–243.
- Bogomolni, R.A. and Spudich, J.L. (1982) Identification of a third rhodopsin-like pigment in phototactic *Halobacterium halobium*. *Proc. Natl Acad. Sci. USA*, **79**, 6250–6254.
- Takahashi, T., Yan, B., Mazur, P. *et al.* (1990) Color regulation in the archaeobacterial phototaxis receptor phoborhodopsin (sensory rhodopsin II). *Biochemistry*, **29**, 8467–8474.
- Béjà, O., Aravind, L., Koonin, E.V. *et al.* (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, **289**, 1902–1906.
- Frigaard, N.U., Martinez, A., Mincer, T.J. *et al.* (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature*, **439**, 847–850.
- Atamna-Ismaeel, N., Sabehi, G., Sharon, I. *et al.* (2008) Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. *ISME J.*, **2**, 656–662.
- Atamna-Ismaeel, N., Finkel, O.M., Glaser, F. *et al.* (2012) Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environ. Microbiol.*, **14**, 140–146.
- Koh, E.Y., Atamna-Ismaeel, N., Martin, A. *et al.* (2010) Proteorhodopsin-bearing bacteria in Antarctic sea ice. *Appl. Environ. Microbiol.*, **76**, 5918–5925.
- Sabehi, G., Massana, R., Bielawski, J.P. *et al.* (2003) Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.*, **5**, 842–849.
- Yutin, N. and Koonin, E. (2012) Proteorhodopsin genes in giant viruses. *Biol. Direct*, **7**, 34–40.
- Philosof, A. and Béjà, O. (2013) Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environ. Microbiol. Rep.*, **5**, 475–482.
- Friedrich, T., Geibel, S., Kalmbach, R. *et al.* (2002) Proteorhodopsin is a light-driven proton pump with variable vectoriality. *J. Mol. Biol.*, **321**, 821–838.
- Man, D., Wang, W., Sabehi, G. *et al.* (2003) Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.*, **22**, 1725–1731.
- Fuhrman, J.A., Schwabach, M.S. and Stingl, U. (2008) Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.*, **6**, 488–494.
- Martinez, A., Bradley, A.S., Waldbauer, J.R. *et al.* (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc. Natl Acad. Sci. USA*, **104**, 5590–5595.
- Finkel, O.M., Béjà, O. and Belkin, S. (2012) Global abundance of microbial rhodopsins. *ISME J.*, **7**, 448–451.
- de la Torre, J.R., Christianson, L.M., Béjà, O. *et al.* (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA*, **100**, 12830–12835.
- Sabehi, G., Béjà, O., Suzuki, M.T. *et al.* (2004) Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.*, **6**, 903–910.
- Venter, J.C., Remington, K., Heidelberg, J.F. *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66–74.
- Sabehi, G., Loy, A., Jung, K.H. *et al.* (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol.*, **3**, e273.
- Rusch, D.B., Halpern, A.L., Sutton, G. *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.*, **5**, e77.
- Campbell, B.J., Waidner, L.A., Cottrell, M.T. *et al.* (2008) Abundant proteorhodopsin genes in the North Atlantic Ocean. *Environ. Microbiol.*, **10**, 99–109.
- Sharma, A.K., Zhaxybayeva, O., Papke, R.T. *et al.* (2008) Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environ. Microbiol.*, **10**, 1039–1056.
- Sharma, A.K., Sommerfeld, K., Bullerjahn, G.S. *et al.* (2009) Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater *Actinobacteria*. *ISME J.*, **3**, 726–737.
- Cottrell, M.T. and Kirchman, D.L. (2009) Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Appl. Environ. Microbiol.*, **75**, 4958–4966.
- Riedel, T., Tomasch, J., Buchholz, I. *et al.* (2010) Constitutive expression of the proteorhodopsin gene by a *Flavobacterium* strain representative of the proteorhodopsin-producing microbial community in the North Sea. *Appl. Environ. Microbiol.*, **76**, 3187–3197.
- Sineschekov, O.A., Jung, K.-H. and Spudich, J.L. (2002) Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA*, **99**, 8689–8694.
- Brown, L.S. (2004) Fungal rhodopsins and opsin-related proteins: eukaryotic homologues of bacteriorhodopsin with unknown functions. *Photochem. Photobiol. Sci.*, **3**, 555–565.
- Saranak, J. and Foster, K.W. (1997) Rhodopsin guides fungal phototaxis. *Nature*, **387**, 465–466.
- Lartillot, N., Rodrigue, N., Stubbs, D. *et al.* (2013) PhyloBayes MPI. Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.*, **62**, 611–615.

33. Ruiz-Gonzalez, M.X. and Marín, I. (2004) New insights into the evolutionary history of type 1 rhodopsins. *J. Mol. Evol.*, **58**, 348–358.
34. Kwon, S.-K., Kim, B.K., Song, J.Y. *et al.* (2013) Genomic makeup of the marine flavobacterium *Nonlabens (Donghaeana) dokdonensis* DSW-6 and identification of a novel class of rhodopsins. *Genome Biol. Evol.*, **5**, 187–199.
35. Spudich, J. and Jung, K. (2005) Microbial rhodopsins: phylogenetic and functional diversity. In: Briggs WR and Spudich JL (eds). *Handbook of Photosensory Receptors*. Wiley-VCH, Weinheim, pp. 1–24.
36. Sharma, A.K., Spudich, J.L. and Doolittle, W.F. (2006) Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol.*, **14**, 463–469.
37. Hiraishi, A. and Shimada, K. (2001) Aerobic anoxygenic photosynthetic bacteria with zinc-bacteriochlorophyll. *J. Gen. Appl. Microbiol.*, **47**, 161–180.
38. Ludwig, W., Strunk, O., Westram, R. *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.*, **32**, 1363–1371.
39. Sabehi, G., Kirkup, B.C., Rozenberg, M. *et al.* (2007) Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas. *ISME J.*, **1**, 48–55.
40. Dioumaev, A.K., Brown, L.S., Shih, J. *et al.* (2002) Proton transfers in the photochemical reaction cycle of proteorhodopsin. *Biochemistry*, **41**, 5348–5358.
41. Han, M. and Zmasek, C. (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, **10**, 356–362.
42. Blankenberg, D., Kuster, G.V., Coraor, N. *et al.* (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.*, Chapter 19: Unit 19.10.121.
43. Giardine, B., Riemer, C., Hardison, R.C. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
44. Goecks, J., Nekrutenko, A., Taylor, J. *et al.* (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
45. Matsen, F., Kodner, R. and Armbrust, E.V. (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, **11**, 538–554.