bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discovery-oriented teaching: The use of chimeric proteorhodopsins for the development of a lab curriculum in marine microbiology and for the discovery of natural red-shifted proteorhodopsins

Alina Pushkarev[1], Gur Hevroni[1], Sheila Roitman[1], Jin-gon Shim[2], Ahreum Choi[2], Kwang-Hwan Jung[2] and Oded Béjà[1]*

[1]Faculty of Biology, Technion – Israel Institute of Technology, Haifa 32000, Israel; [2]Department of Life Science and Institute of Biological Interfaces, Sogang University, 35 Baekbeom-ro, Mapo-Gu, Seoul 121-742, Republic of Korea

*Correspondence: Oded Béjà, Faculty of Biology, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel. Office: (+972)-4-8293961, Lab: (+972)-4-8293410, Fax: (+972)-4-8225153

E-mail: beja@tx.technion.ac.il

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Originality-Significance Statement

We developed an improved chimeric proteorhodopsin (PR) plasmid suitable for direct expression of environmentally retrieved PRs for visualization and functional analysis. This vector was originally envisioned for better design of lab curriculum in marine microbiology. Using this vector we were also able to detect new natural red-shifted PR variants.

Summary

Student laboratory courses in microbial ecology are conducted as condensed courses, where theory and wet lab work are combined in a very intensive short time period. During the last decades, the study of marine ecology became molecular-based, most of the research requiring sequencing that is often not available at the course facilities, and takes more time than the course allows. Therefore, students often find themselves obtaining and analyzing results weeks after the course ended. In this work, we describe a protocol combining molecular and functional methods for analyzing microbial rhodopsins, with visible results in only 4-5 days. We created a chimeric plasmid suitable for direct expression of environmentally retrieved proteorhodopsins (PRs) using PCR, and cloned it into *Escherichia coli* cells for visualization and functional analysis. Using this assay, we discovered several exceptional cases of PRs whose phenotype is different than predicted based on sequence only. We expect this assay to improve the marine

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

microbiology laboratories experience for students, gaining fast feedback and reward for their work, and to promote the use of functional assays for discovery of marine diversity that was masked by sequence-based methods.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Microbial retinal-based ion pumps were first discovered in the hypersaline dwelling archaea *Halobacterium salinarum* (Oesterhelt and Stoeckenius, 1971). Since then, rhodopsins were found in various microorganisms, spanning the three domains of the tree of life (Béjà et al., 2013; Pinhassi et al., 2016) and even detected in viruses (Yutin and Koonin, 2012; Philosof and Béjà, 2013).

The first bacterial rhodopsin was discovered in the abundant uncultured proteobacterial SAR86 group, and was therefore named proteo-rhodopsin (PR) (Béjà et al., 2000). PRs are light driven proton pumps, that absorb light in the blue or green regions of the visible light spectrum (*i.e.* spectral tuning), according to the light available in the depth in which they are isolated from (*i.e.* niche adaptation) (Béjà et al., 2001). The dominant residue responsible for spectral tuning in PRs resides in the retinal-binding pocket at position 105, with Leucine or Methionine in green absorbing PRs (GPRs) and Glutamine in blue absorbing PRs (BPRs) (Man et al., 2003; Gómez-Consarnau et al., 2007). PRs are abundant in the marine environment and a recent metagenomic survey estimated that, on average, over 60% of small microbial cells carry rhodopsin genes (Finkel et al., 2012).

The search for novel rhodopsins is mostly based on sequence homology screens, by utilizing metagenomic data collected from various environments, or by PCR performed on environmental DNA samples using degenerate primers designed for conserved regions in rhodopsin proteins. Only two functional screens were employed to search for new rhodopsins, (i) based on

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

colony colour by plating fosmid libraries on retinal containing plates (Martínez et al., 2007), and (ii) based on pH changes of fosmid clones in response to illumination (Pushkarev and Béjà, 2016).

In order to combine sequence homology and function based methods, we devised a protocol based on a previously designed chimeric PR construct (Supplementary Fig. S1). This chimeric construct was used to express individual partial clones recovered from the environment via PCR amplification, cloning and sequencing (Choi et al., 2013). Here, we improved the chimeric PR construct to enable the screening of diverse partial PR sequences directly from the environment, enabling rapid visualization of the PR activity. In this manner, we developed a simple way to demonstrate the concept of niche adaptation and spectral tuning to undergraduate and graduate students. Student labs in marine microbial ecology or marine microbiology are usually operated as condensed courses ranging between 10 to 30 days. Hence, there is a need for short experiments that can demonstrate some proof of concepts within the timeframe of the course.

In this work we present a protocol (Fig. 1) for analysing microbial samples using functional and molecular methods in short time spans, enabling students to perform high level molecular work while receiving immediate visual results.

Results and discussion

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We used a chimeric green PR protein (GPR) vector with designed restriction sites (Choi et al., 2013) and replaced the middle part of the PR with a "stuffer" DNA sequence, to avoid the high background (reddish colonies) observed with the original chimeric GPR vector (Supplementary Fig. S1). This introduced a stop codon after the downstream restriction site and hence an inactive PR.

The DNA samples from all depths (0 to 100 meter with 20 meter intervals), tested positive for PRs, resulted in two PCR amplicons of ~400 bp and ~330 bp (Supplementary Fig. S2). Two 96-well plates of clones were picked for each depth to detect coloured colonies for further study (Fig. 2). Out of a total of 1,152 colonies, 45 had visible colour and were chosen for further characterization. As expected, the depth from which the sample was collected from, showed the expected spectral tuning of the PRs; with yellow absorbing PR (YPR; purple colonies) and GPR (red colonies) dominating surface waters, while BPRs (orange colonies) were mostly found in deeper samples (Fig. 3B). Samples were collected in March, a time when the Red Sea is not stratified (Supplementary File S1), when light is the only variable changing rapidly with depth, while temperature, density, salinity, and oxygen remain constant. The different PRs amplified by our primers were diverse with some similar to PRs from different microbial groups ranging from Marine Group II (MGII) Euryarchaeota through the proteobacterial groups SAR11, SAR86, SAR92 and to Flavobacteria (Fig. 3A).

Position 105 in PR is believed to be the main determining position for PR wavelength absorption (Man et al., 2003); non polar Methionine (Gómez-

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Consarnau et al., 2007) or Leucine at position 105 result in GPR, while the polar Glutamine residue at position 105 results in a BPR. Several interesting exceptional cases were observed in our screen; clone 1 and 10 are YPR although they contain Leucine at position 105, while clones 36, 37 and 44 appear as GPR although having a glutamine at position 105 (Fig. 3). In the case of clones 36 and 37 we can presume the cause of the red-shift, as another similar clone (clone 34, with only one amino acid change compared to clone 36; see Fig. 4) absorbed in the blue. The change is a Cysteine to a Tyrosine residue in the loop region between transmembranal-domains E and F. This region was previously reported to have influence on spectral tuning (Yoshitsugu et al., 2008). Further work with clones 1, 10 and 44 should be performed in order to understand their spectral tuning mechanism.

In order to verify that the clones are functional proton pumps, a light-driven proton pumping assay was performed. Proton pumping activity was observed in all chimeric PRs with varying intensities (Supplementary File S2). This shows that although the clones obtained are chimeric, they are able to transport protons across the membrane and therefore could be used for teaching and discussing various aspects connected to rhodopsins; the use of uncouplers, change of membrane potential, proton pumping under various wavelengths, the conversion of light energy to potential or chemical energy, *etc*.

In this study we employed two strategies for obtaining absorption spectra; whole cell measurement, and purified protein measurement. A comparison of the two methods is presented in supplementary material

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Supplementary File S2). The advantage of whole cell measurement is rapid result acquisition with limited expenses - significantly relevant to basic laboratory courses. The advantage of purified protein measurement is much greater signal to noise ratio, due to absence of cell content interference. Both spectra obtaining methods are influenced by the pH of the medium, so the exact pH should be noted at each measurement. An abbreviated and detailed protocol is attached to supplementary material for the use in teaching and research (Supplementary File S3).

Concluding remarks

Expression of the environmental fragments depends on primer matching, correct length of PCR product, frame and compatibility to the chimeric construct, and therefore it would be interesting to compare the chimeric ligation results to standard TA cloning of the PCR fragments. This would allow the estimation of the rhodopsins “left in the dark” in such an experiment, and comparing the sequences to the ones that underwent expression in the chimeric vector. This could also be useful for altering the primers to better express rhodopsins from a specific environment. Even though conserved regions between helix C and F are important for colour tuning (Man et al., 2003; Choi et al., 2013), it would also be interesting to create chimeric constructs, where the constant part is of BPR or YPR origin for the expression of protein and testing the dominance of each part over the spectral tuning residues.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

With respect to environmental implications, the fact that we retrieved two YPRs implies that besides the known GPRs and BPRs (and now YPR) more is there to be found. Red light penetrates as deep as 25 meters in clear oceanic water and therefore the existence of more natural YPR, as well as Orange and Red absorbing PR variants are expected in future experiments with the new chimeric PR vector. Although we have tested using ocean genomic samples, it will also apply to freshwater, lake, glacier, etc.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Experimental procedures

Environmental sampling, DNA extraction and PCR

Sampling was performed in March 2014 in the Gulf of Aqaba, Station A (29°28′N, 34°55′E) (Supplementary File S1). Twenty liters of water from 0 m, 20, 40, 60, 80, and 100 meters. Each depth was filtered through GFD filter (Whatman) and collected on two 0.22 μm Durapore filters (Millipore). DNA was extracted from both filters using a phenol-chloroform protocol (Wright et al., 2009).

Degenerate primers were designed based on multiple sequence alignment of known PRs for maximum diversity coverage of using the conserved regions in helix C and F. Each PCR reaction was performed using Takara Ex Taq^{TM} polymerase (Takara-bio, Korea). Polymerase chain reaction amplification was carried out in a total volume of 50 μl containing 1 μl DNA template, 0.2 mM dNTPs, 1X Ex Taq Buffer (Mg^{2+} plus), 0.6 μM primers (each), and 1.25 U Takara Ex Taq™ DNA polymerase. The amplification conditions included steps at 98°C for 5 sec, 40 cycles of 98°C for 10 sec, 50°C for 1.5 min and 72°C for 2 min.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Primer name	Nucleotide sequence
*Kpn*I-LRYLDWIL	AAGAATTCMGGTACCTNGAYTGG
GWAIYP-*ngoM*IV	GCGCGCAAGCTTGCCGGCNGGRTADATNVHCCANCC

Table 1. Degenerate primers designed to amplify diverse PRs using PCR reactions.

195

Cloning and expression

PCR products were double digested with KpnI and NgoMIV in 1.1 buffer (New England BioLabs) for 2 hours in 37°C, and cleaned using ½ volume of phenol and ½ chloroform, then 1 volume of chloroform, and heated to 70°C for 10
200 min to remove residual chloroform by evaporation.

Cloning vector was extracted from cells using only QIAprep Spin Miniprep Kit (Quiagen), since other methods didn't allow efficient restriction, and digested with KpnI and NgoMIV in 1.1 buffer (New England BioLabs) and separated from the 114 bp insert on 1% agarose gel. Cut vector was
205 extracted from gel using NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany). Cloning vector and insert were ligated using T4 DNA ligase (Thermo Scientific, Lithuania) at 4°C overnight. Next day ligation

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

was inactivated using 70°C for 10 min and dialysed 1 hour on VSWP-25 filters (Millipore) against DDW. Ligations were transformed into electro competent DH10B *E. coli* cells (4.5 µl into 30 µl cells) and shaken in 0.5 ml SOC medium at 37°C and 200 rpm. 0.25 ml was plated on LB-agar ampicillin (Amp) plates and incubated overnight. Colonies were transferred to 96 well plated with 230 µl LB-10% glycerol Amp for long term storage. For initial colour expression, plates were duplicated to a U shape bottom 96 well plates (Thermo Scientific, Denmark) with 150 µl LB, 50 µg/ml Amp, 1.2 mM IPTG and 15 *µ*M all-*trans* retinal, and shaken at 37°C and 250 rpm covered by AeraSeal™ air permeable sheet (EXCEL Scientific, Celifornia, USA) overnight. Plates were centrifuged at 3,000X RCF (Sigma 4-16KS centrifuge) at room temperature for the detection of cell colour.

Absorption assay of intact cells and purified protein

A fresh colony was used to inoculate 50 ml LB 50 µg/ml Amp, 1.2 mM IPTG and 15 µM all-trans retinal, and shaken at 37°C and 200 rpm in a 125 ml Erlenmeyer. Cells were collected by centrifugation, and washed twice with buffer A (50 mM Tris-HCl pH 8 and 5 mM $MgCl_2$), then resuspended in 1 ml same solution. The absorbance of the supernatant was recorded between 400 and 800 nm using spectrophotometer (Shimadzu UV-1800, Japan), against buffer A blank. In addition, absorbance spectra of purified protein was measured as described previously (Choi et al., 2013). The results are summarized in supplementary material (Supplementary File S2)- raw, with subtraction of negative control signal, and purified protein spectra.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sequencing and phylogenetic analysis

All clones were extracted using standard alkaline lysis miniprep protocol and sequenced using standard M13R primer (GCGGATAACAATTTCACACAGG, Macrogen, Korea). The middle unique part of each clone, excluding any primer sequences, were deposited in GeneBank under the accession numbers: KY963379-KY963416 eliminating non unique sequences in each depth. Full detailed chimeric sequences of all clones are available in supplementary material (Supplementary File S4).

Maximum likelihood phylogenetic tree was constructed using the phylogeny.fr pipeline (Dereeper et al., 2008), which included PhyML v3.0 (Guindon et al., 2010) and the WAG substitution model for amino acids (Whelan and Goldman, 2001), and 100 bootstrap replicates.

Proton pumping activity assay

Cells were inoculated from a fully thawed plate into two 96-well 2.2 ml plates (ABgene, UK, Cat. No. AB-0932) filled with 1ml LB (50 µg/ml ampicillin, 1.2 mM IPTG and 15 *µ*M all-*trans* retinal) in each well and grown at 30°C, shaking at 700 rpm for 17 hours. Cells were collected by centrifugation at 3,000X RCF (Sigma 4-16KS centrifuge) and washed twice with 0.5 ml minimal salt solution (10 mM NaCl, 10 mM $MgSO_4$ and 10 mM $CaCl_2$). Finally, cells from both plates were re-suspended in a 150 µl (final volume) minimal salt solution in

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dark well plates with a transparent bottom (Greiner Bio-One, Cat. No. 655096). Cells were allowed to settle for 10 min in the dark at RT, after which a functional screening was performed using a customized robotic system (TECAN, Männedorf, Switzerland) as follows: eight pH electrodes (Sentek UK, Cat. No. P13/2.5M/BNC) were measured by a multi-parameter analyzer (Consort, Belgium, model 3060). Changes in pH were logged every 1 sec for 3 min of dark, followed by 2 min of illumination by 2 LED lights: warm white light (2,600-3,700 CCT spanning 420-700 nm, intensity 12 lm, Cree Inc.) and a blue light (485 nm peak, intensity 10.7 lm, Cree Inc.) constituting all visible spectrum under each well. The dark/light cycles were measured twice to confirm consistency. Eight wells were measured simultaneously, each well illuminated by 2 LED lights with a distance of 17 mm between the LEDs and the cells, a setup constructed by Neotec Scientific Instrumentation Ltd. (Kefar Sava, Israel). Each well received a light intensity of 450 µE.

Acknowledgments

The authors would like to thank the captain and crew of the R/V Sam Rothberg of the Inter University Institute in Eilat, Israel, for their expert assistance at sea and for providing their facilities for primary processing of the samples.

This work was supported by Israel Science Foundation grants 1769/12, the I-CORE Program of the Planning and Budgeting Committee, and the Grand Technion Energy Program (GTEP), and is part of The Leona M. and

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Harry B. Helmsley Charitable Trust reports on Alternative Energy series of the Technion – Israel Institute of Technology and the Weizmann Institute of Science, and by grants from the NRF of Korea-2016R1A6A3A11934084 and 2015R1D1A1A01058917.

Conflict of Interest

The authors declare no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures

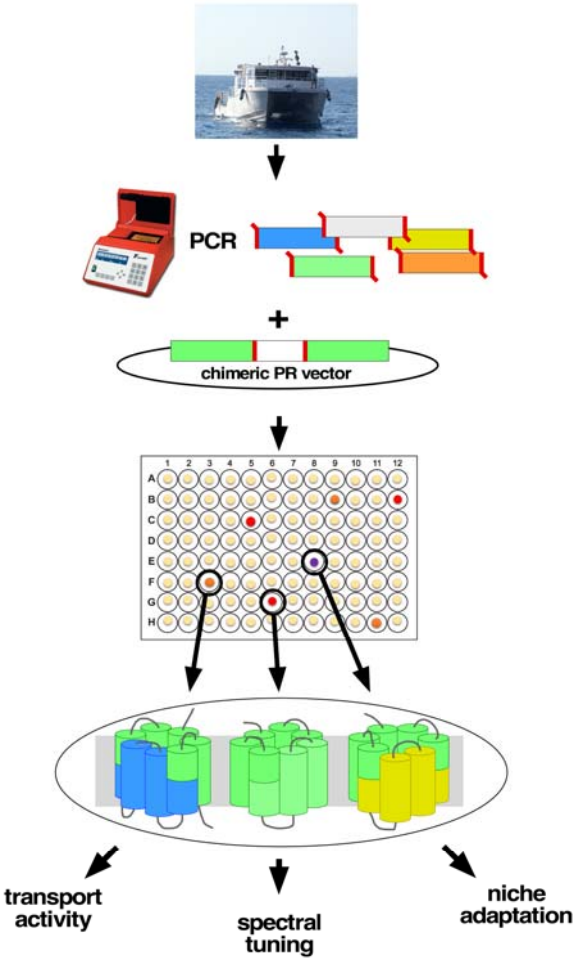

Figure 1. A schematic representation of suggested protocol for the use of chimeric rhodopsin. DNA samples are obtained by various methods from desired niches, then a PCR reaction is done using degenerate primers listed in table 1. The PCR product is then digested and cloned directly into the expression vector. The resulting colonies are harvested into individual cells in a 96 well format for storage, expression and further study.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

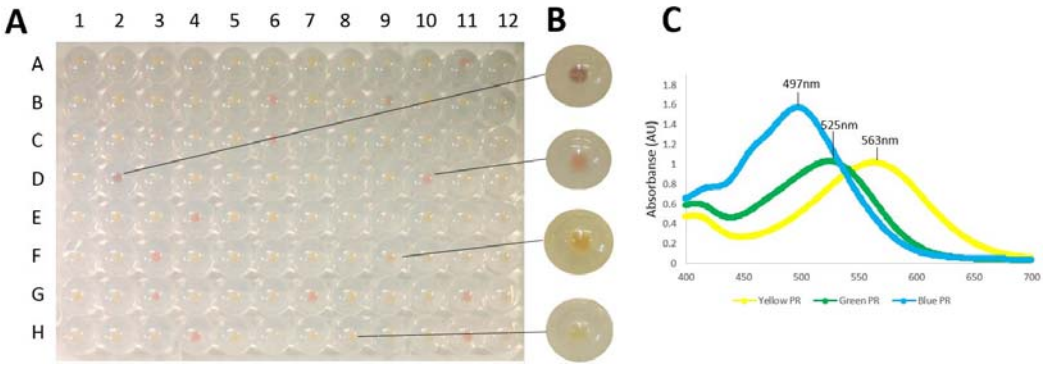

Figure 2. A photograph demonstrating results from one of the expression plates in this study. (A) Colonies were grown in 96 well plates as described in materials and methods, induced by IPTG in the presence of *all-trans* retinal. (B) A close-up of representative clones (from top to bottom); yellow, green, and blue absorbing chimeric rhodopsin and the empty vector pKa00x. (C) Absorption spectra of purified protein representing yellow, green and blue absorbing chimeric rhodopsins (BPR; #42, YPR; #1, GPR; #11).

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

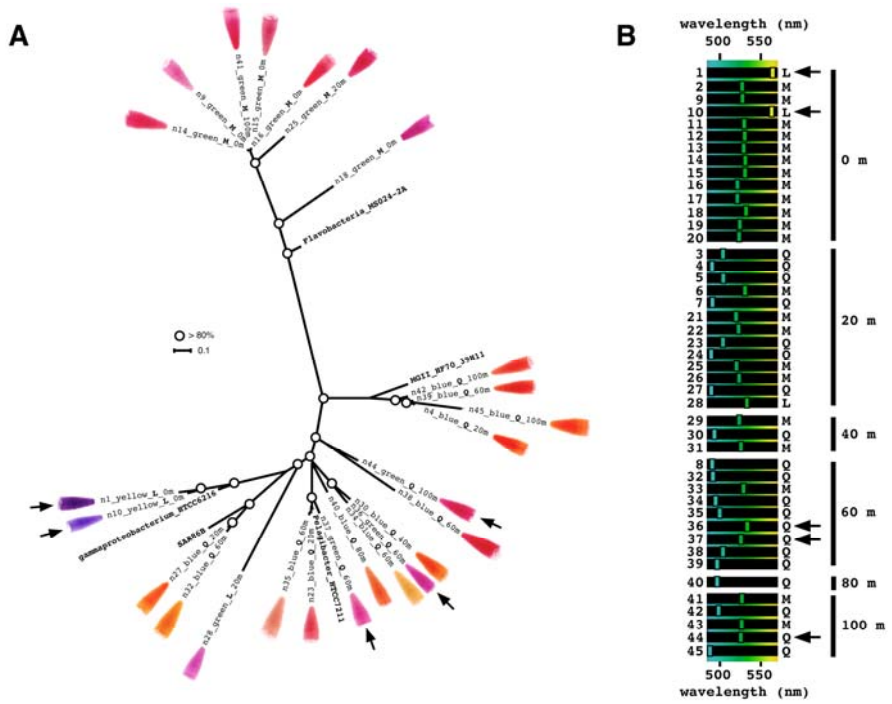

Figure 3. Diverse chimeric clones obtained from different depths in a single cast at station A, the Red Sea. Black arrows mark the clones which show discrepancies between genotype and phenotype according to homologous position 105. (A) A phylogenetic tree was constructed using unique representatives from each depth with rhodopsins from representative cultured microbes. Circles represent bootstrap values higher than 80%. Next to each clone, appears a picture of purified protein after induction in the presence of *all-trans* retinal. (B) The absorption spectra of each clone was measured after protein purification, and the peak is represented on the spectrum between 480 and 570nm. Clones are arranged according to depth.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

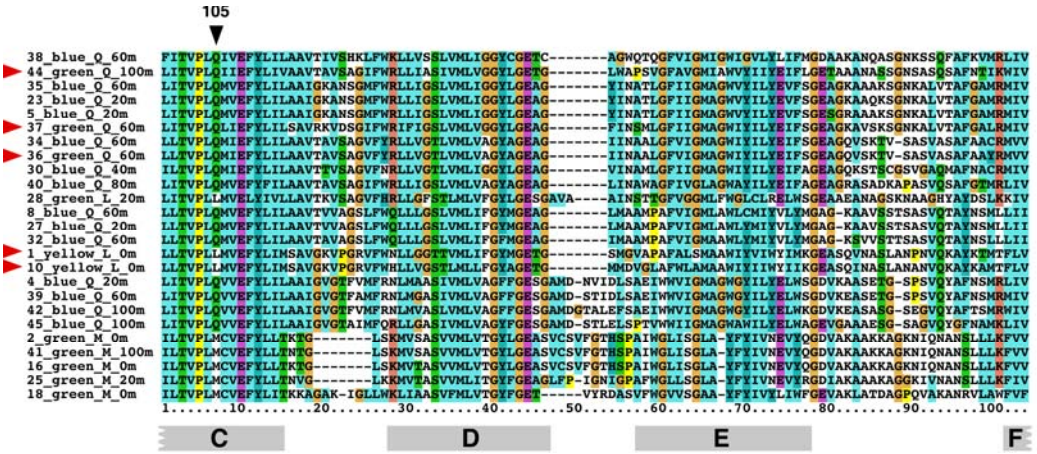

Figure 4. Protein sequence alignment of representative clones found in this study. Position 105 is marked (black arrow). Red arrows mark the clones which show discrepancy between phenotype and genotype according to position 105. At the bottom of the alignment, represented are corresponding transmembrane helixes of PR.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Béjà, O., Pinhassi, J., and Spudich, J.L. (2013) Proteorhodopsins: Widespread Microbial Light-Driven Proton Pumps. In *Encyclopedia of Biodiversity*. Levin, S.A. (ed). New York: Elsevier, pp. 280-285.

Béjà, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411**: 786-789.

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P. et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902-1906.

Choi, A.R., Kim, S.J., Jung, B.H., and Jung, K.H. (2013) Characterization of the chimeric seven-transmembrane protein containing conserved region of helix C-F of microbial rhodopsin from Ganges River. *Appl Microbiol Biotechnol* **97**: 819-828.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F. et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465-469.

Finkel, O.M., Béjà, O., and Belkin, S. (2012) Global abundance of microbial rhodopsins. *ISME J* **7**: 448-451.

Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R. et al. (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**: 210-213.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321.

Man, D., Wang, W., Sabehi, G., Aravind, L., Post, A.F., Massana, R. et al. (2003) Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**: 1725-1731.

Martínez, A., Bradley, A.S., Waldbauer, J., Summons, R.E., and DeLong, E.F. (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in heterologous host. *Proc Natl Acad Sci U S A* **104**: 5590-5595.

Oesterhelt, D., and Stoeckenius, W. (1971) Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat New Biol* **233**: 149-152.

Philosof, A., and Béjà, O. (2013) Bacterial, archaeal and viral-like rhodopsins from the Red Sea. *Environ Microbiol Rep* **5**: 475-482.

Pinhassi, J., DeLong, E.F., Béjà, O., Gonzalez, J.M., and Pedros-Alio, C. (2016) Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology and ecology. *Microbiol Mol Biol Rev* **80**: 929-954.

Pushkarev, A., and Béjà, O. (2016) Functional metagenomic screen reveals new and diverse microbial rhodopsins. *ISME J* **10**: 2331-2335.

Whelan, S., and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**: 691-699.

Wright, J.J., Lee, S., Zaikova, E., Walsh, D.A., and Hallam, S.J. (2009) DNA extraction from 0.22 μM sterivex filters and cesium chloride density gradient centrifugation. *J Vis Exp* **31**: 1352.

bioRxiv preprint doi: https://doi.org/10.1101/162016; this version posted July 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Yoshitsugu, M., Shibata, M., Ikeda, D., Furutani, Y., and Kandori, H. (2008) Color change of proteorhodopsin by a single amino acid replacement at a distant cytoplasmic loop. *Angew Chem Int Ed Engl* **47**: 3923-3926.

380 Yutin, N., and Koonin, E.V. (2012) Proteorhodopsin genes in giant viruses. *Biol Direct* **7**: 34.