

DISCOVERY NOTES

Open Access

# Proteorhodopsin genes in giant viruses

Natalya Yutin and Eugene V Koonin\*

## Abstract

Viruses with large genomes encode numerous proteins that do not directly participate in virus biogenesis but rather modify key functional systems of infected cells. We report that a distinct group of giant viruses infecting unicellular eukaryotes that includes Organic Lake Phycodnaviruses and *Phaeocystis globosa* virus encode predicted proteorhodopsins that have not been previously detected in viruses. Search of metagenomic sequence data shows that putative viral proteorhodopsins are extremely abundant in marine environments. Phylogenetic analysis suggests that giant viruses acquired proteorhodopsins via horizontal gene transfer from proteorhodopsin-encoding protists although the actual donor(s) could not be presently identified. The pattern of conservation of the predicted functionally important amino acid residues suggests that viral proteorhodopsin homologs function as sensory rhodopsins. We hypothesize that viral rhodopsins modulate light-dependent signaling, in particular phototaxis, in infected protists.

This article was reviewed by Igor B. Zhulin and Laksminarayan M. Iyer. For the full reviews, see the Reviewers' reports section.

## Findings

Many if not all viruses encode proteins that counter-act host defense or more generally affect the functions of cellular systems, presumably tweaking them in a manner that favors virus reproduction. Even viruses with small genomes, for example picornaviruses, typically encode a 'security protein' that modifies the host translation system in favor of viral RNA translation [1]. Viruses with larger genomes encode multiple proteins with dedicated functions in the modulation of virus-host interaction at different levels rather than direct roles in virus reproduction [2-5]. A striking example is the presence in numerous cyanophages of genes encoding multiple proteins involved in photosynthesis including complete photosystems I and II [6-8]. These phage-encoded proteins apparently support photosynthesis in infected cyanobacteria and hence promote phage reproduction [9,10]. Here we report the presence in genomes of giant viruses infecting marine unicellular eukaryotes of genes that encode another light-dependent energy-transduction system, proteorhodopsin. We investigate the origin of these genes and discuss their possible roles in the cellular functions of infected protists.

## Bacteriorhodopsins encoded in the genomes of Organic Lake Phycodnaviruses and *Phaeocystis globosa* virus and their abundant homologs in marine environments

In the course of comprehensive comparative genomic analysis of the giant viruses in the families Mimiviridae and Phycodnaviridae, our attention was caught by 5 viral proteins [one from the nearly complete genome of Organic Lake Phycodnavirus (OLPV) 2, three from fragments of other OLPV genomes [11], and one from the more distantly related *Phaeocystis globosa* virus (PGV)] that showed significant sequence similarity to proteorhodopsins from marine bacteria, in particular the most abundant bacterium in the ocean, *Pelagibacter ubique*. The sequences of these proteins were up to 28% identical to proteorhodopsins (expectation value  $<e-05$ ); all viral proteins with similarity to proteorhodopsins are currently annotated as 'hypothetical proteins' in GenBank although for two of the OLPV proteins the similarity to bacteriorhodopsins is pointed out in a note. Proteorhodopsins represent a distinct, comparatively simple phototrophic system that is of crucial importance in marine ecology [12-14]. These proteins belong to the broader family of bacteriorhodopsins (or Type I rhodopsins) that originally were discovered in Halobacteria (Euryarchaeota) and subsequently identified in diverse bacteria as well as protists and fungi [15]. To our knowledge, proteorhodopsins (or any other rhodopsin

\* Correspondence: koonin@ncbi.nlm.nih.gov  
National Center for Biotechnology Information, National Library of Medicine,  
National Institutes of Health, Bethesda, MD 20894, USA

superfamily members) have not been previously detected in viruses, so we were interested in a detailed analysis of these sequences.

Comparison of the putative viral proteorhodopsins to databases of environmental sequences revealed numerous highly similar sequences, some more than 50% identical and with e-values as low as  $e^{-64}$ . The much greater similarity between the putative viral proteorhodopsins and the environmental sequences, all coming from the Global Ocean Survey (GOS) [16], than between any of these sequences and bacterial proteorhodopsins suggests that the detected environmental sequences are also of viral origin. We collected a representative set of putative viral, bacterial, archaeal and eukaryotic bacteriorhodopsins and constructed a multiple alignment of these sequences (Additional file 1). Inspection of this alignment indicates the conservation of all 7 transmembrane helices that were also independently predicted in the viral proteins (Figure 1 and Additional file 2). Furthermore, the viral sequences contained the invariant lysine residue that is involved in retinal binding as well as the aspartic acid that serves as the proton acceptor; the proton donor glutamate is not conserved; the position known to be important for spectral tuning [17] is occupied by a methionine as it is in some of the proteorhodopsins (Figure 1). The lack of conservation of the proton donor carboxylate indicate that viral proteorhodopsin homologs are sensory rhodopsins rather than light-dependent proton pumps [18,19]. The presence of a hydrophobic residue in the spectral tuning position suggests that viral proteorhodopsins absorb light in the green part of the spectrum [17].

#### Origin of the viral proteorhodopsins

We used the alignment shown in Figure 1 to construct a phylogeny of bacteriorhodopsins (see Additional file 1 for the full alignment). In the resulting phylogenetic tree, the proteorhodopsin homologs from giant viruses and their environmental homologs form a strongly supported clade with two distinct subclades, each with numerous environmental sequences (Figure 2; see Additional file 3 for details). Taking into account the much higher sequence similarity between the viral sequences and proteorhodopsins as opposed to all other groups of bacteriorhodopsins, the root position can be inferred between two major clades one of which consists of viral and bacterial proteorhodopsins and the other one encompasses the rest of prokaryotic (halorhodopsins, sensory rhodopsins, xenorhodopsins and others) and eukaryotic Group I rhodopsins (Figure 2). This tree structure implies that giant viruses acquired proteorhodopsin genes via horizontal transfer from bacteria or more likely from proteorhodopsin-encoding eukaryotes.

Notably, *Phaeocystis globosa*, the protist host of PGV, encodes two closely related rhodopsins. However, these rhodopsins confidently group within the eukaryotic branch of Proteorhodopsin Group II (Additional file 4) and accordingly are not the ancestors of the proteorhodopsins of PGV or other viruses. In the tree shown in Figure 2, the viral rhodopsins join the proteorhodopsin clade at its base which at face value seems to suggest ancient acquisition of the proteorhodopsin gene by ancestral giant viruses. However, we cannot rule out that some of the environmental sequences in the 'viral' clade actually come from planctonic protists and represent the (still uncharacterized) source(s) of the rhodopsin genes in giant viruses.

#### Implications for virus-host interaction in giant viruses

It appears likely that proteorhodopsins of giant viruses modulate phototrophic process in the infected protists. Although proteorhodopsins originally were discovered and characterized in bacteria [13] and subsequently in mesophilic Archaea [20], more recently, members of this family have been identified in several dinoflagellates [21-23]. Notably, in the marine dinoflagellate *Oxyrrhis marina*, proteorhodopsin is the most highly expressed nuclear protein, suggesting a major physiological role(s) [21]. Database searches also indicate the presence of two closely related proteorhodopsins in the prasinophyte *P. globosa* (Figure 1 and Additional file 4). There are no experimental data on the functions of proteorhodopsins in these unicellular eukaryotes. However, by analogy with the well characterized bacterial proteorhodopsins [24], it appears likely that those of the eukaryotic proteorhodopsins that possess the proton donor carboxylate function as light-driven proton pumps involved in ATP synthesis, particularly under oligotrophic conditions, whereas those that lack the proton donor perform sensory functions, in particular in phototaxis [21,25]. By this token, the proteorhodopsins of *P. globosa* are predicted to possess proton-pumping activity (see Figure 1; the second paralogous sequence from *P. globosa* is nearly identical and is not shown).

Viral proteorhodopsins that are predicted to function as sensory rhodopsins could affect signaling and in particular phototaxis in the infected protists, perhaps stimulating relocation of the infected protists to areas that are rich in nutrients required for virus reproduction. Complete sequencing of the genome of *P. globosa* and the still unidentified hosts of OLPV (most likely, also prasinophytes [11]) will show whether the putative viral sensory rhodopsins complement a pre-existing host function or confer a functionality that is new to the host. Given that *P. globosa* is a dominant component of marine phytoplankton and that its population dynamics is substantially affected by viruses [26], viral proteorhodopsin homologs described here,



**Figure 1 Conserved sequence blocks in the rhodopsin superfamily.** The conserved blocks are separated by numbers that indicate the length of less well conserved sequence segments that are not shown (see Additional File 1). The alignment columns are colored on the basis of the respective position conservation throughout the superfamily: yellow background indicates hydrophobic residues (ACFILMWWY), red letters indicate polar residues (DEHKNQR), and green background indicates small residues (ACGNPSTV). The transmembrane helices are indicated following transmembrane helix prediction for PGV sequence (helix A is not shown; see Additional File 2 for all 7 predicted transmembrane helices). The functionally important residues are numbered: 1, proton acceptor; 2, position important for spectral tuning; 3, proton donor; 4, retinal-binding amino acid residue. Each sequence is denoted by the corresponding taxon abbreviation followed by the species abbreviation and GenBank Identification (GSI) number. Taxa abbreviations: A, Archaea; B, Bacteria; E, Eukaryota; Ae, *Euryarchaeota*; Ba, *Actinobacteria*; Bb, *Bacteroidetes/Chlorobi* group; Bc, *Cyanobacteria*; Bd, *Deinococcus-Thermus*; Bf, *Firmicutes*; Bh, *Chloroflexi*; Bo, *Planctomycetes*; Bp, *Proteobacteria*; E9, *Viridiplantae*; Ec, *Alveolata*; Eh, *Cryptophyta*; El, *Opisthokonta*; Em, *Glaucocystophyceae*; Ep, *Haptophyceae*. Species abbreviations: are listed in Additional File 3.

regardless of their exact role(s) that remains to be elucidated experimentally, could be major players in the ocean ecology.

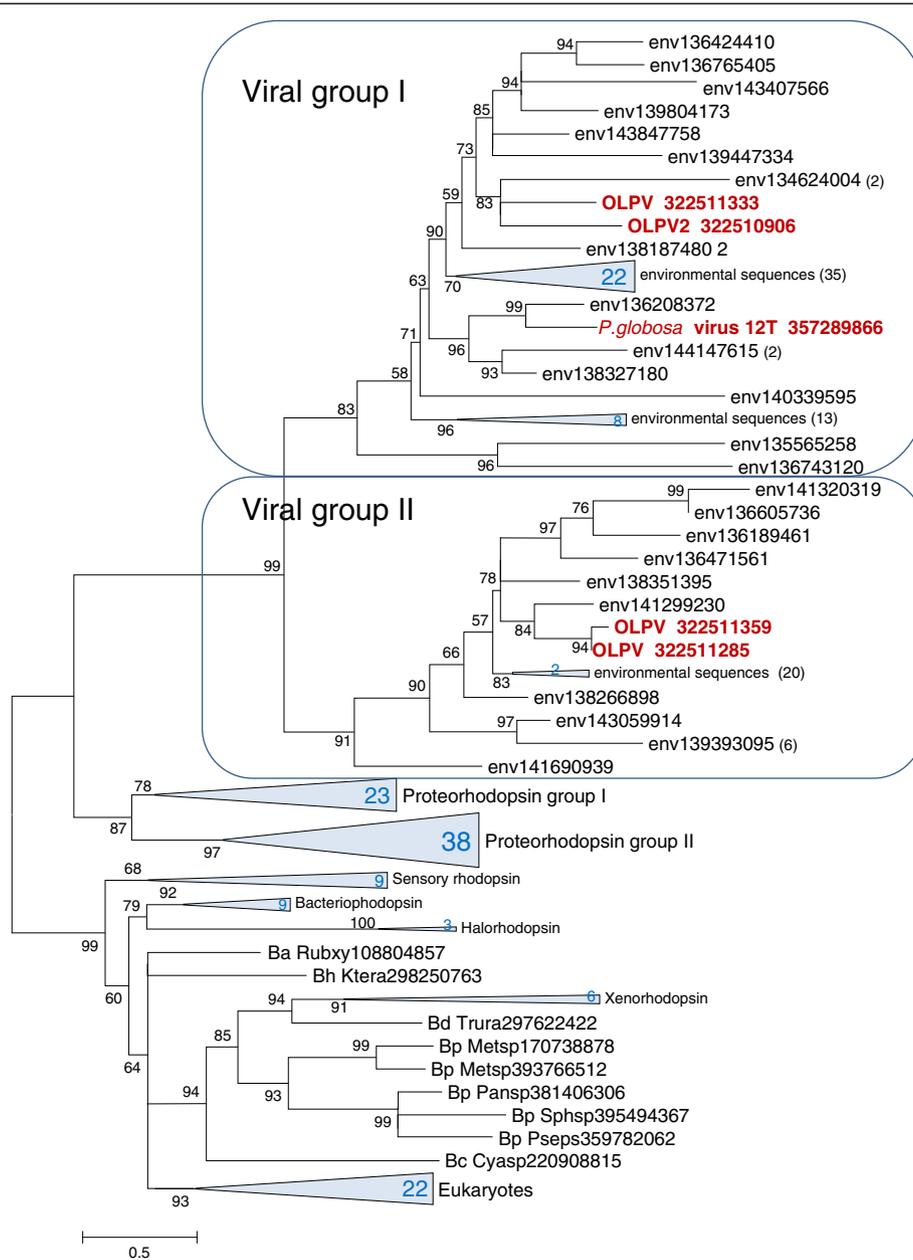
**Conclusions**

Proteorhodopsin homologs encoded by giant viruses belong to a distinct proteorhodopsin subfamily that additionally includes numerous uncharacterized sequences from marine environments that are likely to be of virus and/or eukaryotic origin. The viruses probably acquired proteorhodopsin genes from unicellular eukaryotic hosts

although the identity of the donors remains unknown. These proteins are predicted to perform light-dependent sensory functions, potentially altering the behavior of the resulting protist host, e.g. by inducing phototaxis and perhaps stimulating the host relocation to nutrient-rich areas.

**Methods**

Protein sequences were retrieved from the non-redundant database at the National Center for Biotechnology Information (NIH, Bethesda). Reference sequences for halo-



**Figure 2 Phylogenetic tree of the rhodopsin superfamily.** Branches with bootstrap support less than 50 were collapsed. Several large clades are shown by triangles with the number of the collapsed branches shown within the triangle. Numbers in parentheses represent number of environmental sequences clustered into the branch. Each sequence is denoted by the corresponding species abbreviation and GenBank Identification (GI) number. Abbreviations: OLPV, Organic Lake Phycodnavirus; OLPV2, Organic Lake Phycodnavirus 2; env, environmental sequence (marine metagenome); Ba, *Actinobacteria*; Bc, *Cyanobacteria*; Bd, *Deinococcus-Thermus*; Bh, *Chloroflexi*; Bp, *Proteobacteria*; Cyasp, *Cyanotheca* sp. PCC 7424; Ktera, *Ktedonobacter racemifer* DSM 44963; Metsp, *Methylobacterium* sp. 4–46; Pansp, *Pantoea* sp. Sc1; Pseps, *Pseudomonas psychrotolerans* L19; Rubxy, *Rubrobacter xylanophilus* DSM 9941; Sphsp, *Sphingomonas* sp. PAMC 26617; Trura, *Truepera radiovictrix* DSM 17093. Expanded subtrees for Proteorhodopsin group I and II are shown in Additional file 4.

bacterio-, xeno, and sensory rhodopsins were taken from [27]. The non-redundant protein sequence database was searched using the PSI-BLAST program [28], with default parameters and the predicted viral proteorhodopsin sequences used as queries. The reported results reflect searches performed on 13-15/08/2012. Marine

metagenomics blast hits were clustered before the alignment by blastclust (<http://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/blastlab.html>); a representative (the longest) sequence from each cluster was taken. Protein sequences were aligned using the MUSCLE program with default parameters [29]; columns containing a large

fraction of gaps (greater than 30%) and non-homogenous columns defined as described previously [30] were removed from the alignment. The resulting 160-column alignment was used to construct a maximum likelihood phylogenetic tree using the FastTree program with default parameters (JTT evolutionary model, discrete gamma model with 20 rate categories) [31]. Transmembrane helices in proteins were predicted using the TMHMM program [32].

## Additional files

**Additional file 1: Secondary structure prediction for viral proteorhodopsin homologs.**

**Additional file 2: Multiple alignment of bacteriorhodopsins used for the construction of the phylogenetic tree in Figure 2.**

**Additional file 3: Supplementary data for Figure 1 and Figure 2.**

**Additional file 4: Expanded phylogenetic trees for Proteorhodopsin groups I and II.**

## Competing interests

The authors declare no conflict of interests.

## Authors' contributions

NY collected the data; NY and EVK analyzed the data; EVK wrote the manuscript which was read and approved by both authors.

## Reviewers' reports

Reviewer 1: Dr. Igor B. Zhulin, Oakridge National Laboratory and the University of Tennessee

The paper by Yutin and Koonin reports a discovery of proteorhodopsin genes in marine viruses. This is a very interesting finding expanding the repertoire of genes that viruses might carry in order to modify host's metabolism. As different organisms developed various forms of light-harvesting devices, and at least some of them have been already found in viruses (photosystems I and II), it does not come as a total surprise, and supports the notion that improving host's conditions promotes phage reproduction. When conditions are right, proteorhodopsin can be a very useful plug-and-play device for energy generation.

The paper is brief, clearly written and goes straight to the point. I do not have any particular comments or concerns rather than it would be useful to indicate the timing of database searches since NR changes so rapidly and parameters for PSI-BLAST and MUSCLE (presumably, default, but still...) – all in the Methods section

**Authors' response:** *The details proposed to be included were included.*

Reviewer 2: Dr. Laksminarayan M. Iyer, National Center for Biotechnology Information, NIH

The study details the presence, and analyzes the origins, of proteorhodopsins in certain marine NCLDV viruses. These proteins add to the small, yet interesting, list of laterally acquired genes in large dsDNA viruses that are predicted to alter the response of the infected host to various environmental inputs. The precise biology of how the viral proteorhodopsins contribute to the fitness of the virus and the host should elicit interest among experimental biologists. The analysis can be easily reproduced and the writing is lucid. I only have a minor comment. On the issue of the provenance of environmental sequences most closely related to the viral ones, the authors could consider performing a gene neighborhood analysis, where possible, to see if any predominant associations emerge that might provide clues to the origins of these sequences.

**Authors' response:** *Regrettably, the majority of the environmental sequences are too short for this type of analysis. Those few sequences that were long enough failed to yield useful clues.*

## Acknowledgements

The authors thank Oded Beja and Valerian Dolja for critical reading of the manuscript and useful suggestions. The authors' research is supported by

the US Department of Health and Human Services intramural funds (to National Library of Medicine).

Received: 30 August 2012 Accepted: 1 October 2012

Published: 4 October 2012

## References

1. Agol VI, Gmyl AP: **Viral security proteins: counteracting host defences.** *Nat Rev Microbiol* 2010, **8**(12):867–878.
2. Wei H, Zhou MM: **Viral-encoded enzymes that target host chromatin functions.** *Biochim Biophys Acta* 2010, **1799**(3–4):296–301.
3. de Souza RF, Iyer LM, Aravind L: **Diversity and evolution of chromatin proteins encoded by DNA viruses.** *Biochim Biophys Acta* 2010, **1799**(3–4):302–318.
4. Werden SJ, Rahman MM, McFadden G: **Poxvirus host range genes.** *Adv Virus Res* 2008, **71**:135–171.
5. Bugert JJ, Darai G: **Poxvirus homologues of cellular genes.** *Virus Genes* 2000, **21**(1–2):111–133.
6. Alperovitch-Lavy A, Sharon I, Rohwer F, Aro EM, Glaser F, Milo R, Nelson N, Beja O: **Reconstructing a puzzle: existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets.** *Environ Microbiol* 2011, **13**(1):24–32.
7. Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, Pinter RY, Partensky F, Koonin EV, Wolf YI, et al: **Photosystem I gene cassettes are present in marine virus genomes.** *Nature* 2009, **461**(7261):258–262.
8. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW: **Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts.** *PLoS Biol* 2006, **4**(8):e234.
9. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW: **Photosynthesis genes in marine viruses yield proteins during host infection.** *Nature* 2005, **438**(7064):86–89.
10. Bragg JG, Chisholm SW: **Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene.** *PLoS One* 2008, **3**(10):e3550.
11. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA, et al: **Virophage control of antarctic algal host-virus dynamics.** *Proc Natl Acad Sci U S A* 2011, **108**(15):6163–6168.
12. Beja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF: **Proteorhodopsin phototrophy in the ocean.** *Nature* 2001, **411**(6839):786–789.
13. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, et al: **Bacterial rhodopsin: evidence for a new type of phototrophy in the sea.** *Science* 2000, **289**(5486):1902–1906.
14. Moran MA, Miller WL: **Resourceful heterotrophs make the most of light in the coastal ocean.** *Nat Rev Microbiol* 2007, **5**(10):792–800.
15. Spudich JL, Yang CS, Jung KH, Spudich EN: **Retinylidene proteins: structures and functions from archaea to humans.** *Annu Rev Cell Dev Biol* 2000, **16**:365–392.
16. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooshep S, Wu D, Eisen JA, Hoffman JM, Remington K, et al: **The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific.** *PLoS Biol* 2007, **5**(3):e77.
17. Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, Spudich EN, Spudich JL, Beja O: **Diversification and spectral tuning in marine proteorhodopsins.** *EMBO J* 2003, **22**(8):1725–1731.
18. Atamna-Ismaeel N, Finkel OM, Glaser F, Sharon I, Schneider R, Post AF, Spudich JL, von Mering C, Vorholt JA, Iluz D, et al: **Microbial rhodopsins on leaf surfaces of terrestrial plants.** *Environ Microbiol* 2012, **14**(1):140–146.
19. Jung KH: **The distinct signaling mechanisms of microbial sensory rhodopsins in Archaea, Eubacteria and Eukarya.** *Photochem Photobiol* 2007, **83**(1):63–69.
20. Frigaard NU, Martinez A, Mincer TJ, DeLong EF: **Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea.** *Nature* 2006, **439**(7078):847–850.
21. Slamovits CH, Okamoto N, Burri L, James ER, Keeling PJ: **A bacterial proteorhodopsin proton pump in marine eukaryotes.** *Nat Commun* 2011, **2**:183.
22. Lin S, Zhang H, Zhuang Y, Tran B, Gill J: **Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates.** *Proc Natl Acad Sci U S A* 2010, **107**(46):20033–20038.

23. Okamoto OK, Hastings JW: **Novel dinoflagellate clock-related genes identified through microarray analysis.** *J Phycol* 2003, **39**:519–526.
24. Fuhrman JA, Schwalbach MS, Stingl U: **Proteorhodopsins: an array of physiological roles?** *Nat Rev Microbiol* 2008, **6**(6):488–494.
25. DeLong EF, Beja O: **The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times.** *PLoS Biol* 2010, **8**(4):e1000359.
26. Brussard CPG, Bratbak G, Baudoux AC, Ruardij P: **Phaeocystis and its interaction with viruses.** *Biogeochemistry* 2007, **83**:201–215.
27. Ugalde JA, Podell S, Narasingarao P, Allen EE: **Xenorhodopsins, an enigmatic new class of microbial rhodopsins horizontally transferred between archaea and bacteria.** *Biol Direct* 2011, **6**:52.
28. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389–3402.
29. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**(5):1792–1797.
30. Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin EV: **The deep archaeal roots of eukaryotes.** *Mol Biol Evol* 2008, **25**(8):1619–1630.
31. Price MN, Dehal PS, Arkin AP: **FastTree 2—approximately maximum-likelihood trees for large alignments.** *PLoS One* 2010, **5**(3):e9490.
32. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**(3):567–580.

doi:10.1186/1745-6150-7-34

**Cite this article as:** Yutin and Koonin: Proteorhodopsin genes in giant viruses. *Biology Direct* 2012 **7**:34.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

