J Mol Evol (2004) 58:348–358
DOI: 10.1007/s00239-003-2557-8

JOURNAL OF MOLECULAR EVOLUTION

New Insights into the Evolutionary History of Type 1 Rhodopsins

Mario X. Ruiz-González,[1,2] Ignacio Marín[1]

[1] Departamento de Genética, Universidad de Valencia, Calle Doctor Moliner 50, Burjassot 46100, Valencia, Spain
[2] Department of Zoology, Trinity College, Dublin, Ireland

Received: 15 May 2003 / Accepted: 9 October 2003

Abstract. Type 1 (archaeal) rhodopsins and related rhodopsin-like proteins had been described in a few halophile archaea, γ-proteobacteria, a single cyanobacteria, some fungi, and a green alga. In exhaustive database searches, we detected rhodopsin-related sequences derived not only from additional fungal species but also from organisms belonging to three groups in which opsins had hitherto not been described: the α-proteobacterium *Magnetospirillum magnetotacticum*, the cryptomonad alga *Guillardia theta*, and the dinoflagellate *Pyrocystis lunula*. Putative plant and human type 1 rhodopsin sequences found in the databases are demonstrated to be contaminants of fungal origin. However, a highly diverged sequence supposedly from the plant *Oryza sativa* was found that is, together with the *Pyrocystis* sequence, quite similar to γ-proteobacterial rhodopsins. These close relationships suggest that at least one horizontal gene transfer event involving rhodopsin genes occurred between prokaryotes and eukaryotes. Alternative hypotheses to explain the current phylogenetic range of type 1 rhodopsins are suggested. The broader phylogenetic range found is compatible with an ancient origin of type 1 rhodopsins, their patchy distribution being caused by losses in multiple lineages. However, the possibility of ancient horizontal transfer events between distant relatives cannot be dismissed.

Key words: Rhodopsin — Cryptomonad — Dinoflagellate — Horizontal transfer

Correspondence to: Ignacio Marín; *email:* ignacio.marin@uv.es

Introduction

Rhodopsins are members of the seven-transmembrane receptor family, able to react to light by using retinal as chromophore. There are two types of rhodopsins, type 1 and type 2 rhodopsins. The best-known type 2 rhodopsins are the animal visual photoreceptors (reviewed in Yokoyama 2000). Type 1 rhodopsins were discovered in halophile archaea, and therefore they are often called "archaeal opsins." Type 1 and type 2 rhodopsins are structurally similar and perform related functions, but it is unclear whether they are evolutionary related (reviewed by Spudich et al. 2000). Archaea contain four classes of type 1 rhodopsins. Two of them, sensory rhodopsin I (or simply "sensory rhodopsin"; sR) and sensory rhodopsin II, also known as phoborhodopsin (pR), function as sensors for phototaxis. The other two are transport rhodopsins called bacteriorhodopsin (bR), which functions as a proton-extruding pump, and halorhodopsin (hR), which is a chloride uptake pump (reviewed in Ihara et al. 1999; Spudich et al. 2000). After their discovery in archaea, genes with clear sequence similarities to Type 1 rhodopsins were characterized in other lineages. First, they were found in a few fungi (Graul and Sadee 1997; Bieszke et al. 1999a; Idnurm and Howlett 2001; Zhai et al. 2001). Some of these fungal genes encode proteins with the typical sequence signatures of light-modulated opsins (Bieszke et al. 1999a), and one has been demonstrated to bind retinal to form a photoactive pigment with a photochemical reaction cycle typical of type 1 rhodopsins (Bieszke et al. 1999b). However, despite being very similar in sequence, some other fungal proteins must be unable to act as photosensors, because they lack a critical lysine residue involved in retinal binding. This last type of molecules is called opsin-related proteins (Bieszke et al. 1999a; Spudich et al. 2000). Typical type 1 rhodopsin genes were later found in some uncultured γ-proteobacteria (Béjà et al. 2000, 2001; Man et al. 2003). Their product, called proteorhodopsin, is a transport rhodopsin, which works as a light-driven proton pump (Béjà et al. 2000). Finally, type 1 rhodopsins have recently been described in the green alga *Chlamydomonas reinhardtii*, where they function as receptors for phototaxis responses (Sineshchekov et al. 2002) and as light-activated proton channels when heterologously expressed (Nagel et al. 2002), and in the cyanobacteria *Nostoc* (*Anabaena*) sp., which shows features typical of sensory rhodopsins (Jung et al. 2003).

For type 1 rhodopsins and their relatives, it has been suggested that vertical transmission plus rare horizontal gene transfer episodes explains the generation of the patchy evolutionary distribution observed (Bieszke et al. 1999a; Béjà et al. 2000). Determination of the evolutionary history of these proteins critically depends on precisely establishing their evolutionary range and the phylogenetic relationships among rhodopsin genes. We have performed a comprehensive survey of the available type 1 rhodopsin sequences to determine whether horizontal transmission may have occurred. Although the information is insufficient to determine the evolutionary history of all these genes, at least one episode of horizontal transfer is suggested by the close similarity of some eukaryotic and γ-proteobacterial rhodopsins. Alternative hypotheses are proposed to explain the current phylogenetic range of these genes, so far found in a few, very distant organisms.

Materials and Methods

Exhaustive TBLASTN searches (Altschul et al. 1997) for type 1 rhodopsins were performed on all the databases available at the NCBI Web site (http://www.ncbi.nlm.nih.gov/). The database compiled in the GOLD homepage (found at http://igweb.integratedgenomics.com/GOLD/) were also screened for sequences not yet present in the NCBI databases. These searches were finished in March 2003.

Multiple sequence alignments of the protein sequences, conceptually translated from the corresponding nucleotide sequences, were performed using Clustal X, v. 1.83 (Thompson et al. 1987). The alignments were corrected manually, editing the sequences with GeneDoc 2.6 (Nicholas and Nicholas 1997). GeneDoc was also used to highlight the similarities among sequences shown in Fig. 1. Rhodopsins are membrane proteins with seven transmembrane helices. The most conserved region, included in our analyses, spans from the middle of the second helix (helix B) to a few amino acids beyond the C terminus of the seventh helix (helix G; for a total of 180 to 200 amino acids). Final phylogenetic trees were obtained from the protein multiple alignments, after eliminating the region between helix B and helix C, which cannot be reliable aligned. For phylogenetic reconstruction, three types of analyses were performed. First, we used the neighbor-joining (NJ) method (Saitou and Nei 1987), as implemented in Clustal X, v. 1.83, with correction for multiple substitutions. One thousand bootstrap replicates were performed to determine the reliability of tree topology. Second, we utilized maximum parsimony (MP), as implemented in MEGA 2.1 (Kumar et al. 2001). We performed two types of MP analyses. First, we generated a fast MP analysis from which a bootstrap consensus tree (Nei and Kumar 2000) was obtained. For this first analysis, parameters were as follow: (1) all sites included; (2) initial trees obtained by random addition, with 10 replicates; and (3) close-neighbor interchange with search level 3. The second analysis was much more exhaustive. Parameters were the same as in the first analysis, except that the number of trees obtained by random addition was 100,000. This large number of trees precludes the generation of a bootstrap consensus tree due to computer power limitations. However, it allows exploration of the consistency of the topology obtained with the previous, less exhaustive, method. Finally, we performed, as the third approach, a maximum likelihood (ML) analysis, using PROTML 2.2 (Adachi and Hasegawa 1992). The version of this program compiled for PC computers by R. L. Malmberg, which is publicly available at his Web page (http://dogwood.botany.uga.edu/malmberg/software.html), was used. We employed the random addition option (−q) and the JTT-f amino acid replacement model. The 50 best trees were evaluated using the resampling of estimated log likelihood method (RELL; Kishino et al. 1990), equivalent to bootstrapping, to obtain the values presented below. Trees presented below were drawn using TreeView 1.6.6 (Page 1996).

To establish whether other genes have phylogenetic ranges similar to the one found for type 1 rhodopsin genes, we explored online the Cluster of Orthologous Groups (COGs) database (http://www.ncbi.nlm.nih.gov/COG/phylox.html [Tatusov et al. 1997; Tatusov et al. 2001]) and later we checked the results obtained by systematic TBLASTN searches against the NCBI databases, as indicated above.

Results

Table 1 summarizes the information for the 67 type 1 rhodopsins and rhodopsin-like sequences detected. The multiple alignment of those sequences is shown in Fig. 1, where it can be observed that 10 sequences are incomplete. Several organisms for which rhodopsins were not hitherto described appear in Table 1: the α-proteobacterium *Magnetospirillum magnetotacticum*, the cryptomonad alga *Guillardia theta*, the dinoflagellate *Pyrocystis lunula*, several plant species, and even a sequence annotated as belonging to our own species. In Fig. 2, we show a preliminary NJ tree. It is obvious from Fig. 3 results that all but one of the putative plant-derived sequences and also the supposedly human-derived sequence are actually most likely of fungal origin. The single exception is a rice sequence (Oryza1) that is quite different from any other rhodopsin sequence described so far and very different from the fungal proteins.

We then proceeded to eliminate all partial and obviously contaminant sequences, carefully reanalyzing the remaining 56 full-length sequences. Results are presented in Fig. 3, where we show the NJ tree, which is similar to the one shown in Fig. 2, together

Table 1. Type 1 rhodopsin sequences detected in our searches

Species and gene name	Sequence name (Figs. 1 to 3)	Accession no.
Eukaryotes (all fungal species, unless indicated)		
Aspergillus nidulans	Aspergillus	AACD0100055
Botrytis cinerea	Botrytis	AL115930
Candida albicans	Cand albicans HSP30	PEDANT ca1507*
Candida albicans	Cand albicans HSP31	PEDANT ca4034*
Candida glabrata	Cand glabrata	BZ293756 + BZ297564
Chaetomium globosum	Chaetomium	BP099207
Chlamydomonas reinhardtii (green alga)	Chlamydomonas CSOA	AV640817
Chlamydomonas reinhardtii (green alga)	Chlamydomonas CSOB	AB058891
Coccidioides posadasii	Coccidioides	AY059409
Coriolus versicolor	Coriolus	AB018405
Cryptococcus neoformas	Cryptococcus	SGTC cneo011005.C1391**
Gibberella zeae	Gibberella	BU059532 + BU067691
Guillardia theta (cryptomonad alga)	Guillardia	AW342219
Histoplasma capsulatum	Histoplasma	F_HGC186AR.contig-2909***
Homo sapiens (animal)	Homo	AC138524
Leptosphaeria maculans	Leptosphaeria	AF290180
Magnaporthe grisea	Magnaporthe	AC098842
Mycosphaerella graminicola	Mycosphaerella1	AW180117
Mycosphaerella graminicola	Mycosphaerella2	COGEME mg[0384]****
Neurospora crassa NOP1	Neurospora NOP1	AF135863
Neurospora crassa YRO2	Neurospora YRO2	AL356815
Oryza sativa (*indica*) (plant)	Oryza1	AAAA01084480
Oryza sativa (*indica*) (plant)	Oryza2	CA764330
Paracoccidiodes brasiliensis	Paracoccidioides	BQ497887
Prunus persica (plant)	Prunus	BU041889
Pyrocystis lunula (dinoflagellate)	Pyrocystis	AF508258
Saccharomyces bayanus	Sacch bayanus	AACA01000331
Saccharomyces castellii	Sacch castellii	AZ926138
Saccharomyces cerevisiae YDR033W	Sacch cerev YDR033W	Z68196
Saccharomyces cerevisiae YRO2	Sacch cerev YRO2	Z35923
Saccharomyces cerevisiae HSP30	Sacch cerev HSP30	M93123
Saccharomyces mikatae	Sacch mikatae	AABZ01000282
Saccharomyces paradoxus	Sacch paradoxus	AABY01000119
Schizosaccharomyces pombe	Schizosaccharomyces	CAA21219
Shorgum bicolor (plant)	Shorgum	AW564434
Triticum aestivum (plant)	Triticum	AL821328
Zygosaccharomyces rouxii	Zygosaccharomyces	AL395409
Archaea		
Haloarcula argentinensis bR	Ha argentinensis bR	D31880
Haloarcula japonica bR	Ha japonica bR	AB029320
Haloarcula sp. (Andes) bR	Ha sp bR	S76743
Haloarcula vallismortis hR	Ha vallismortis hR	D31881
Haloarcula vallismortis pR	Ha vallismortis pR	Z35308
Haloarcula vallismortis sR	Ha vallismortis sR	D83748
Halobacterium salinarum hRp	Hb salinarum hR	P16102
Halobacterium salinarum hR	Hb salinarum hR	D43765
Halobacterium salinarum bR	Hb salinarum bR	AF306937
Halobacterium salinarum pR	Hb salinarum pR	AE005080
Halobacterium salinarum sR	Hb salinarum sR	X51682
Halobacterium sp. AUS bRp	Hb sp AUS2 bR	S56354
Halobacterium sp. SG1 Hr	Hb sp SG1 hR	X70292
Halobacterium sp. SG1 bRp	Hb sp SG1 bR	X70291
Halobacterium sp. SG1 sR	Hb sp SG1 sR	X70290
Halobacterium sp. NRC-1 bR	Hb sp NRC1 bR	AE004437
Halobacterium sp. (BOP) bR	Hb sp bR	AB009620
Halorubrum sodomense hR	Hr sodomense hR	AB009622
Halorubrum sodomense bR	Hr sodomense bR	D50848
Halorubrum sodomense sR	Hr sodomense sR	AB009623
Haloterrigena sp. st. arg-4 hR	Hr sp arg4 hR	AB009621
Natromonas pharaonis hR	N pharaonis hR	J05199
Natromonas pharaonis pR	N pharaonis pR	Z35086

(Continued)

Table 1. Continued

Species and gene name	Sequence name (Figs. 1 to 3)	Accession no.
Eubacteria		
Uncultured γ-proteobacterium (MEDA17)	Proteobacterium1	AY250738
Uncultured γ-proteobacterium (MEDA15)	Proteobacterium2	AY250740
Uncultured γ-proteobacterium (PAL B6)	Proteobacterium3	AF349998
Uncultured γ-proteobacterium (BAC 31A8)	Proteobacterium4	AF279106
Uncultured γ-proteobacterium (BAC 40E8)	Proteobacterium5	AF349976
Magnetospirillum magnetotacticum (α-proteobacterium)	Magnetospirillum	AAAP01002252
Nostoc sp. (cyanobacterium)	Nostoc	AP003592

Note. Archaeal gene nomenclature follows Ihara et al. (1999) (bR, bacteriorhodopsin; hR, halorhodopsin; sR, sensory rhodopsin; pR, phoborhodopsin). All numbers refer to the NCBI database with the exception of the following: (*) obtained from the PEDANT database (http://pedant.gsf.de/); (**) from the Stanford Genome Technology Center database (http://www-sequence.stanford.edu); (* * *) from the Genome Sequencing Center, Washington University School of Medicine database (http://genome.wustl.edu); and (* * **) from the COGEME database (http://cogeme.ex.ac.uk).

with the information for bootstrap/RELL values of those branches reasonably supported by the three methods of phylogenetic reconstruction used (see Materials and Methods). These results fully confirm previous phylogenetic studies for archaeal proteins (e.g., Ihara et al. 1999), with the exception of the lack of support for a monophyletic clade containing all phoborhodopsins. This may be related to phoborhodopsins evolving at the highest rate among all archaeal proteins (Ihara et al. 1999). It is also congruent with trees obtained with fewer sequences by Bieszke et al. (1999a) and Zhai et al. (2001). The position of proteorhodopsins in our Fig. 3 is different from the one shown by Béjà et al. (2000), where they appeared closer to archaeal sensory rhodopsins than to archaeal halorhodopsins or bacteriorhodopsins. However, the topology for the most internal branches in our tree cannot be ascertained with the available data (e.g., compare Figs. 2 and 3), and therefore, we conclude that the precise position of proteorhodopsins is ambiguous. Support for all fungal sequences forming a single monophyletic group increases slightly when partial sequences are eliminated (see NJ values in Figs. 2 and 3). Moreover, the low level of support for this branch is mainly caused by a single highly divergent sequence, from *Cryptococcus neoformans*, whose phylogenetic position is ambiguous. It appears in some analyses together with nonfungal sequences (see Fig. 3 legend). Without this sequence, support for a fungal monophyletic group is much higher in the three methods of phylogenetic reconstruction (not shown). Two fungal groups are apparent in Figs. 2 and 3 (and, again, they obtain stronger support when the *C. neoformans* sequence is eliminated). The existence of these two classes of type 1 rhodopsin-like proteins in fungi was already suggested by Bieszke et al., (1999a). This dichotomy is congruent with the known features of these proteins, because one of the groups contains the opsin-related sequences in multiple ascomycete species and the basidiomycete *Coriolus versicolor*, while the second branch includes the *bona fide Neurospora* rhodopsin gene *Nop1* (Bieszke et al. 1999a) and very similar genes of other ascomycete species. Considering the sequences for which we have information (see Fig. 1), we found that only 1 of the 15 genes in the opsin-related branch contains the critical lysine residue in the seventh helix, known to be involved in retinal binding, while all the genes in the *Nop1* branch have it. The already-mentioned problematic sequence of the basidiomycete *C. neoformans* also contains that residue and may thus belong to the *Nop1* class. The latter result suggests that the duplication that gave rise to the two main groups of fungal genes is older than the ascomycetes–basidiomycetes split.

Five eukaryotic sequences are excluded from the fungal group (Figs. 2 and 3). Among them, we found the two rhodopsin genes recently described in *Chlamydomonas reinhardtii* (Nagel et al. 2002; Sineshchekov et al. 2002). Their high similarity suggests that they derive from a recent duplication. In addition, we detected sequences putatively derived from the cryptomonad alga *Guillardia*, the dinoflagellate *Pyrocystis*, and the plant *Oryza*. The phylogenetic relationships of the algal sequences with other prokaryotic or eukaryotic rhodopsins are unclear (see Figs. 2 and 3). However, the *Pyrocystis* and *Oryza* sequences are quite similar and appear together with proteorhodopsins. The sequences of the α-proteobacterium *Magnetospirillum magnetotacticum* and the cyanobacterium *Nostoc* sp. (Jung et al. 2003) are very different from proteorhodopsins and appear in the phylogenetic trees as independent branches.

To determine how frequently genes are limited to four groups as distant as archaea, eukaryotes, proteobacteria, and cyanobacteria, we searched the COGs database, using the phylogenetic pattern search option, for genes present in those groups but absent in other eubacteria. All genes with those characteristics were then checked individually in searches against the NCBI sequence databases to confirm the COGs results. We finally found three

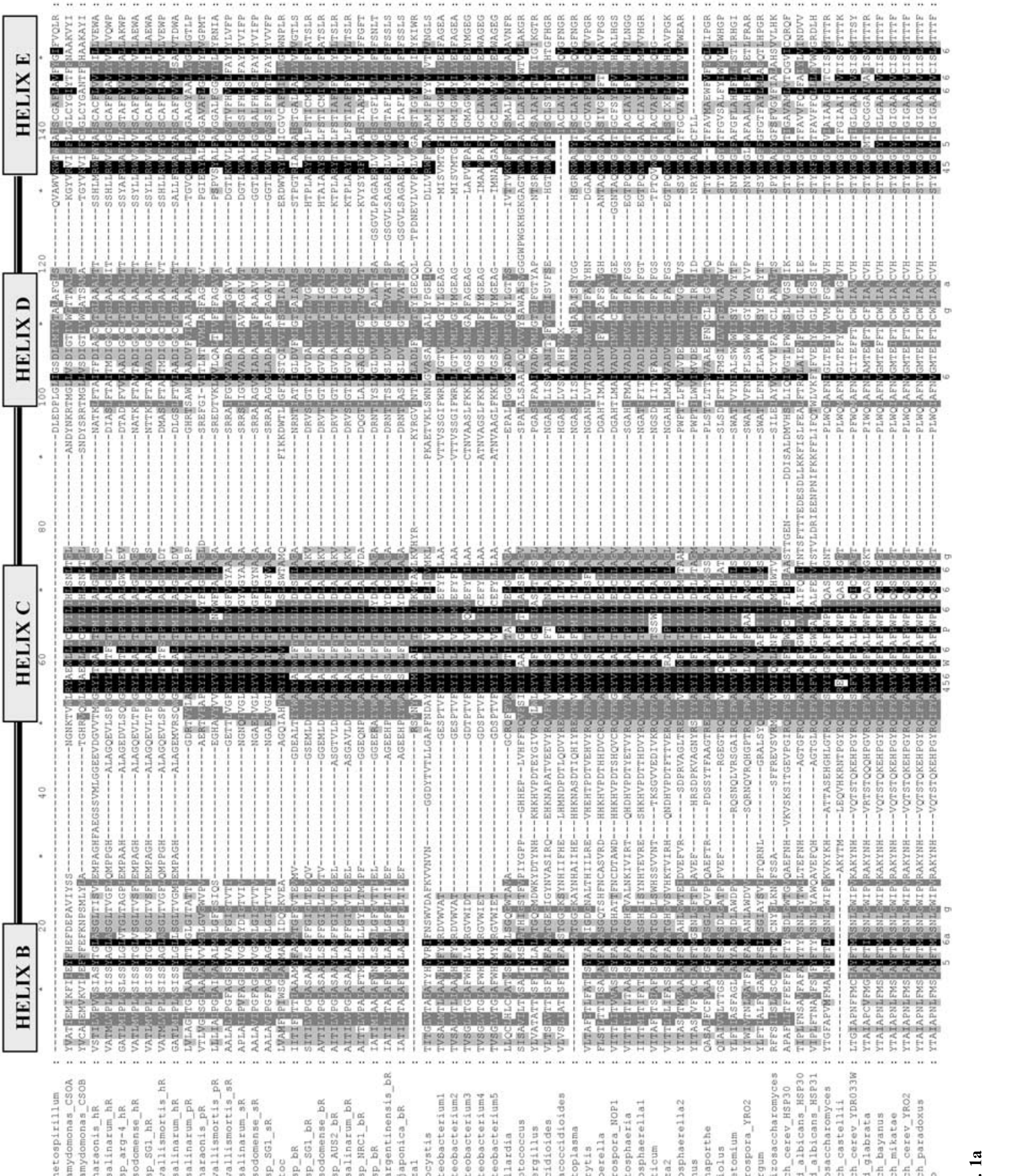
HELIX B
HELIX C
HELIX D
HELIX E
Magnetospirillum
Chlamydomonas_CSOA
Chlamydomonas_CSOB
N_pharaonis_hR
Hb_salinarum_hR
Ht_sp_arg-4_hR
Hr_sodomense_hR
Hb_sp_SG1_hR
Ha_vallismortis_hR
Hb_salinarum_hR
Hb_salinarum_pR
N_pharaonis_pR
Ha_vallismortis_pR
Ha_vallismortis_sR
Hb_salinarum_sR
Hr_sodomense_sR
Hb_sp_SG1_sR
Nostoc
Hb_sp_bR
Hb_sp_SG1_bR
Hr_sodomense_bR
Ha_sp_AUS2_bR
Hb_salinarum_bR
Hb_sp_NRC1_bR
Ha_sp_bR
Ha_argentinensis_bR
Ha_japonica_bR
Oryza1
Pyrocystis
Proteobacterium1
Proteobacterium2
Proteobacterium3
Proteobacterium4
Proteobacterium5
Guillardia
Cryptococcus
Aspergillus
Coccidioides
Paracoccidioides
Histoplasma
Botrytis
Gibberella
Neurospora_NOP1
Leptosphaeria
Mycosphaerella1
Triticum
Oryza2
Mycosphaerella2
Prunus
Magnaporthe
Coriolus
Chaetomium
Neurospora_YRO2
Shorgum
Schizosaccharomyces
Sacch_cerev_HSP30
Cand_albicans_HSP30
Cand_albicans_HSP31
Zygosaccharomyces
Sacch_castellii
Sacch_cerev_YDR033W
Cand_glabrata
Sacch_bayanus
Sacch_mikatae
Sacch_cerev_YRO2
Sacch_paradoxus
Homo

Fig. 1a

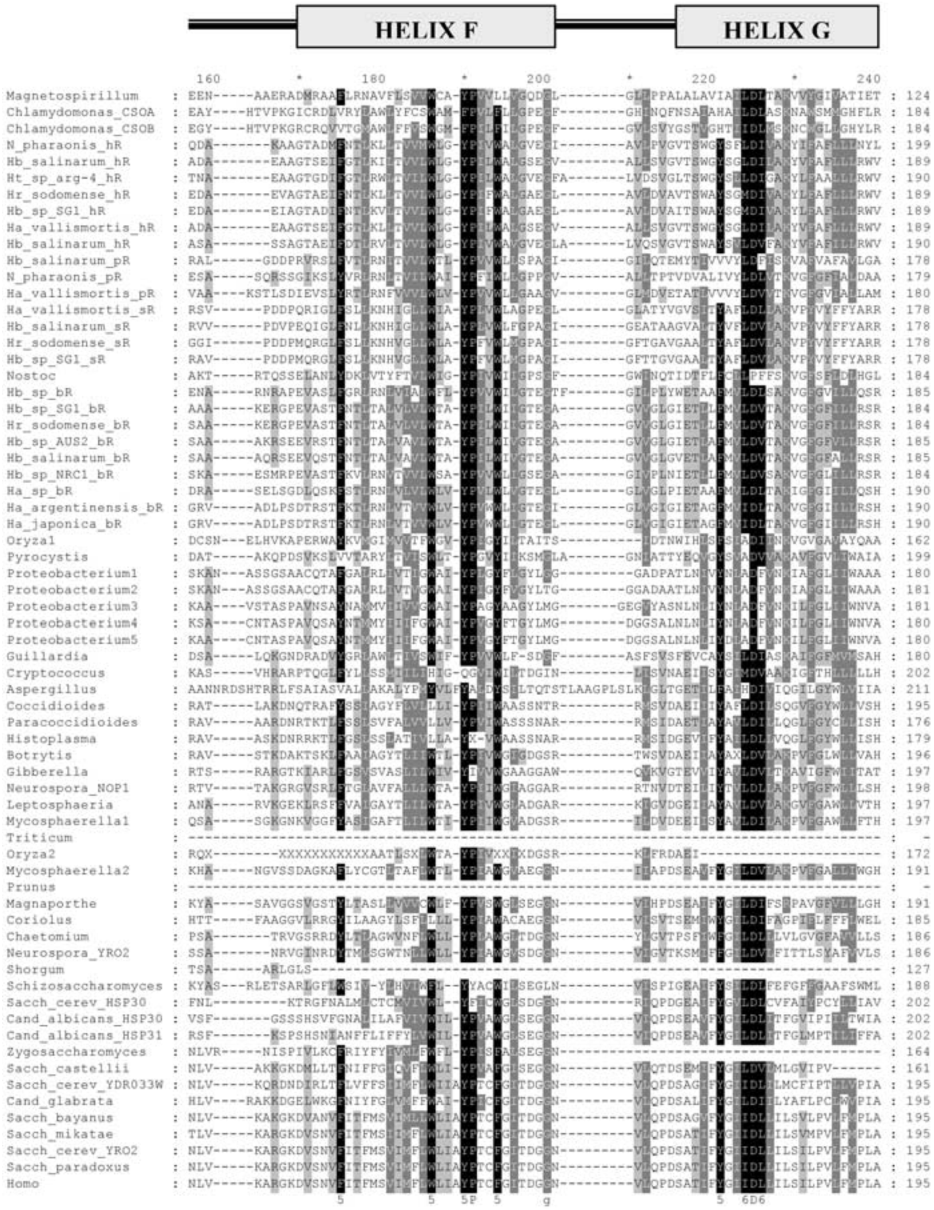

Fig. 1. Multiple-sequence alignments of type 1 rhodopsins and rhodopsin-like proteins. Positions of helices B to G (according to Spudich et al. 2000) are indicated at the top. Biochemically conserved residues are highlighted with gray (conserved) to black (highly conserved) shading. Sequences are ordered according to the phylogenetic tree shown in Fig. 2.

clear-cut cases with a similar phylogenetic pattern. They correspond to (1) the orthologues of the gene encoding for yeast translation initiation factor SUI1 (COG0023), already described by Kyrpides and Woese (1998); (2) COG0278, which includes glutaredoxin-like genes of *E. coli* (*ydhD*) and *S. cerevisiae* (*GRX3*, *GRX4*, and *GRX5*) among other organisms; and (3) orthologues of the *E. coli* stress response gene *bolA* (COG0271). These results show that the patchy phylogenetic distribution observed for type 1 rhodopsins and rhodopsin-related sequences is not particularly unusual. It must be considered that these automatic searches are expected to produce a significant underestimation of the number of genes with the phylogenetic range that we explored, due to the fact that only a single eukaryote (*Saccharomyces cerevisiae*) and a single cyanobacterium (*Synechocistis* sp.) are included in the COGs database. In fact, rhodopsins themselves cannot be detected by such methods, because they are not present in the *Synechocistis* genome.

Discussion

Type 1 rhodopsins are present in archaea, eubacteria, and eukaryotes. However, in each of these three domains, they have been found in just a few species. In archaea, they seem to be restricted to some related halophilic species. In eubacteria, they have appeared in γ-proteobacteria, a single cyanobacterium, and, as described above, an α-proteobacterium. In eukaryotes, in addition to the already known rhodopsins in fungi and green algae, we found that sequences annotated as belonging to a cryptomonad alga (*Guillardia theta*), a dinoflagellate (*Pyrocystis lunula*), several plants, and even humans encoded rhodopsin-like proteins. However, phylogenetic analyses strongly suggested that all but one of the putative plant sequences and the human sequence were false positives of fungal origin. Only one plant sequence, supposedly from *Oryza sativa*, had a level of divergence that suggested it was not derived from fungi. The discovery of these multiple contaminant sequences raises the critical question of whether the other new eukaryotic sequences found may actually not belong to the species to which they have been attributed. This problem must be tackled individually. However, an important preliminary consideration is that it is obvious from our results (Fig. 2) that rhodopsin sequences provide reasonable phylogenetic information only for relatively closely related species. Several data can be used to determine the age of the splits that separated organisms that appear in supported groups in our trees. It turns out that they are separated by a few hundreds of millions of years of independent evolution. Thus, estimates of divergence time among archaea allowed the origin of all species containing archaeal type 1 rhodopsins to be traced back in time to about 700 million years ago (Ihara et al. 1999). Similarly, rDNA data presented by Suzuki et al. (2001) and our own data on sequence distances suggest that all characterized proteorhodopsins are present in organisms that diverged at most a few hundreds of millions of years ago (e.g., compare in Figs. 2 and 3 the branch lenghts for proteorhodopsins with even the slowest-evolving archaeal opsins). Finally, the finding of rhodopsins in basidiomycetes and ascomycetes determines the origin of these proteins in fungi to be older than 600 million years, when those two groups diverged (Redecker et al. 2000; Redecker 2002). Heckman et al. (2001) have suggested that the basidiomycetes/ascomycetes split would have occurred much earlier (perhaps 1.2 billion years ago), but this conclusion has been critizised by other authors (e.g., Rodriguez-Trelles et al. 2002). In opposition to those data, organisms that appear in our trees in independent branches, and whose divergences can be approximately situated in time, split much longer ago. For example, cyanobacterial and proteobacterial rhodopsins appear in our trees as independent branches. Feng et al. (1997) estimated that the split of the lineages that generated proteobacteria and cyanobacteria occurred about 2.2 billion years ago. In a recent combined analysis of paleontological, phenotypic, and molecular data, Cavalier-Smith (2002) has suggested that the split of the ancestors of cyanobacteria and proteobacteria may have occurred even earlier, more than 2.5 billion years ago.

These considerations suggest that the strong similarity of the putative *Oryza* and *Pyrocystis* sequences to γ-proteobacterial rhodopsins cannot be explained by conventional vertical transmission. We should then postulate a totally unrealistic deacceleration of the rate of change of those proteins. It is also illogical to argue that they are of mitochondrial origin. In that case, the eukaryotic proteins should be more similar to the α-proteobacterial *Magnetospirillum* rhodopsin than to the γ-proteobacterial rhodopsins. We are therefore left with two reasonable alternatives. First, we can interpret that high similarity as supporting a relatively recent horizontal transmission. Second, those sequences may be rhodopsins of proteobacterial origin annotated as belonging to eukaryotic species. We favor the first option because there is good preliminary evidence that the *Pyrocystis* sequence indeed originated from the genome of the dinoflagellate. The clone that contains this sequence derives from a cDNA library obtained by Okamoto et al. (2001). Although the starting RNA derived from a nonaxenic culture (CCMP731; see Provasoli-Guillard National Center for Culture of Marine Phytoplankton, http://ccmp.bigelow.org/SD/dculture.php?strain = CCMP731), the cDNAs were obtained from poly(A) RNA (Okamoto et al. 2001), thus making the probability of contaminants of bacterial origin much less likely. Moreover, Okamoto et al. (2003) have obtained evidence that the expression of this gene follows a circadian rhythm. So far, this type of rhytmicity has not been described in proteobacteria (but see Dvornyk et al. 2003). Thus, it is unlikely that this clone has a proteobacterial contaminant origin, although it is still formally possible and independent validation is advisable. In any case, the fact that rhodopsins exist in green algae and probably also in dinoflagellates suggests that they may also exist in groups evolutionarily close to those two. In this context, we think that the *Guillardia* sequence may be a bona fide cryptomonad rhodopsin. There is growing evidence that cryptomonads and dinoflagellates derive from a common ancestor, originated after an endosymbiotic event involving a red alga and a flagellate (reviewed in Archibald and Keeling 2002). It is attractive then to hypothesize that rhodopsins were present in the lineage that generated the closely related green and red algae (e.g., Baldauf et al. 2000; Moreira et al. 2000; Van de Peer et al. 2000) and that cryptomonads and dinoflagellates obtained those genes from their red algal ancestor. In the case of *Guillardia*, the genome of the red algae has mostly dissapeared, with fewer than 500 genes in three small chromosomes found in a structure derived

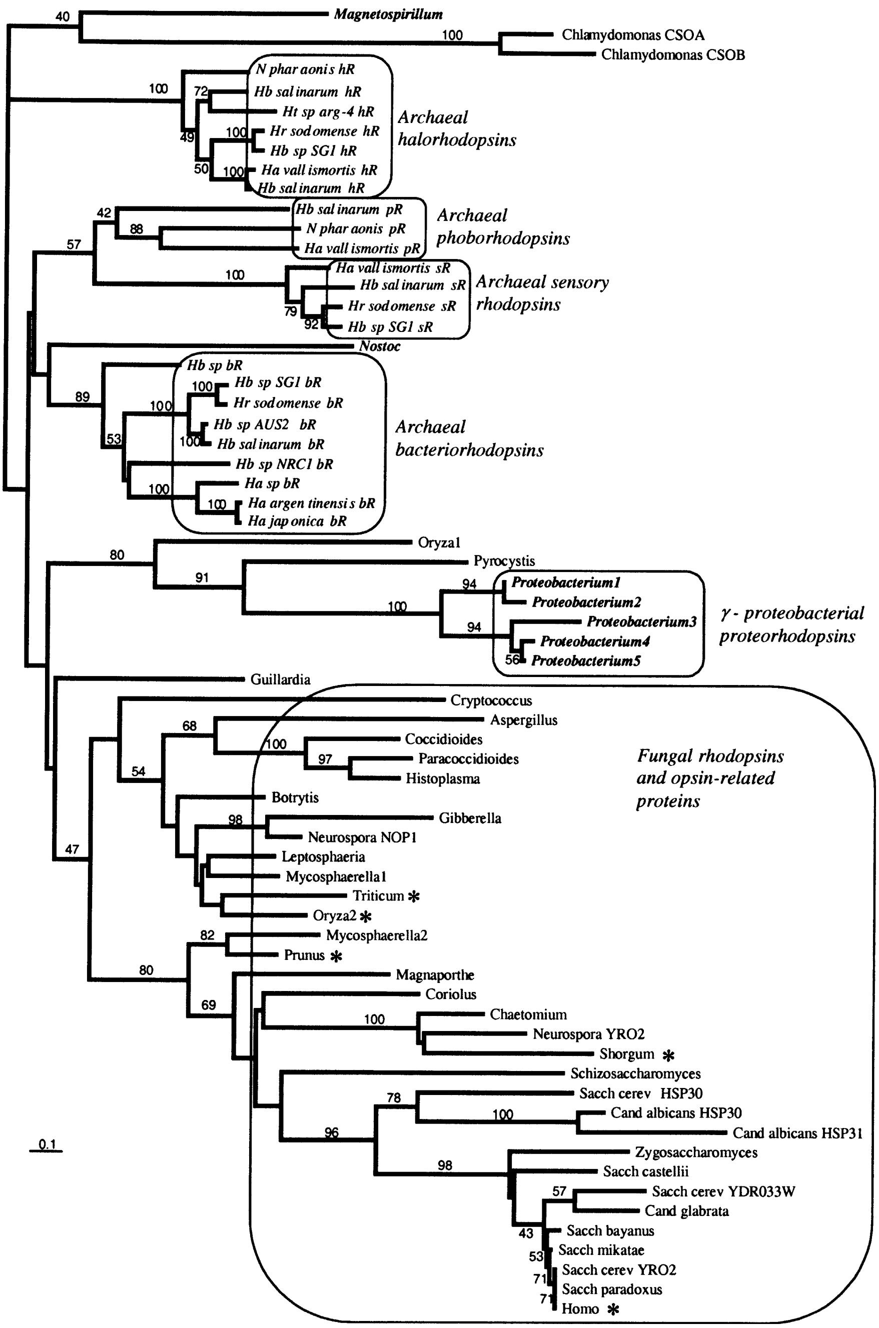

Fig. 2. Preliminary unrooted phylogenetic tree for type 1 rhodopsins and their relatives, obtained using the neighbor-joining method. Archaeal species are shown in italics, and eubacterial species in boldface italics. Numbers refer to bootstrap support (as a percentage). Only those branches with support of at least 40% are shown. Asterisks refer to rhodopsins assigned to nonfungal species that are most likely of fungal origin (see text).

from the algal nucleus, called the nucleomorph, remaining (Douglas et al. 2001). We found that the *Guillardia* rhodopsin gene does not come from the fully sequenced chloroplast or nucleomorph genomes. Therefore, it may derive from the mitochondria or the main nucleus. The first option is unlikely, considering that this gene is also absent from the mitochondrial genome of another cryptomonad, *Rhodomonas salina* (accession number NC_002572). Thus, it most likely comes from the main (flagellate) nucleus. Despite this fact, it still may be of red algal origin, because it may have been transferred, as has been shown to occur for other genes (e.g., Deane et al. 2000) from the algal nucleus (now nucleomorph) to the main nucleus. If we thus assume that the algal and dinoflagellate rhodopsins have an ancient common origin and also that a relatively recent horizontal transfer event occurred involving a dinoflagellate and a proteobacterium, the transfer must have been from the eukaryote to the prokaryote. This

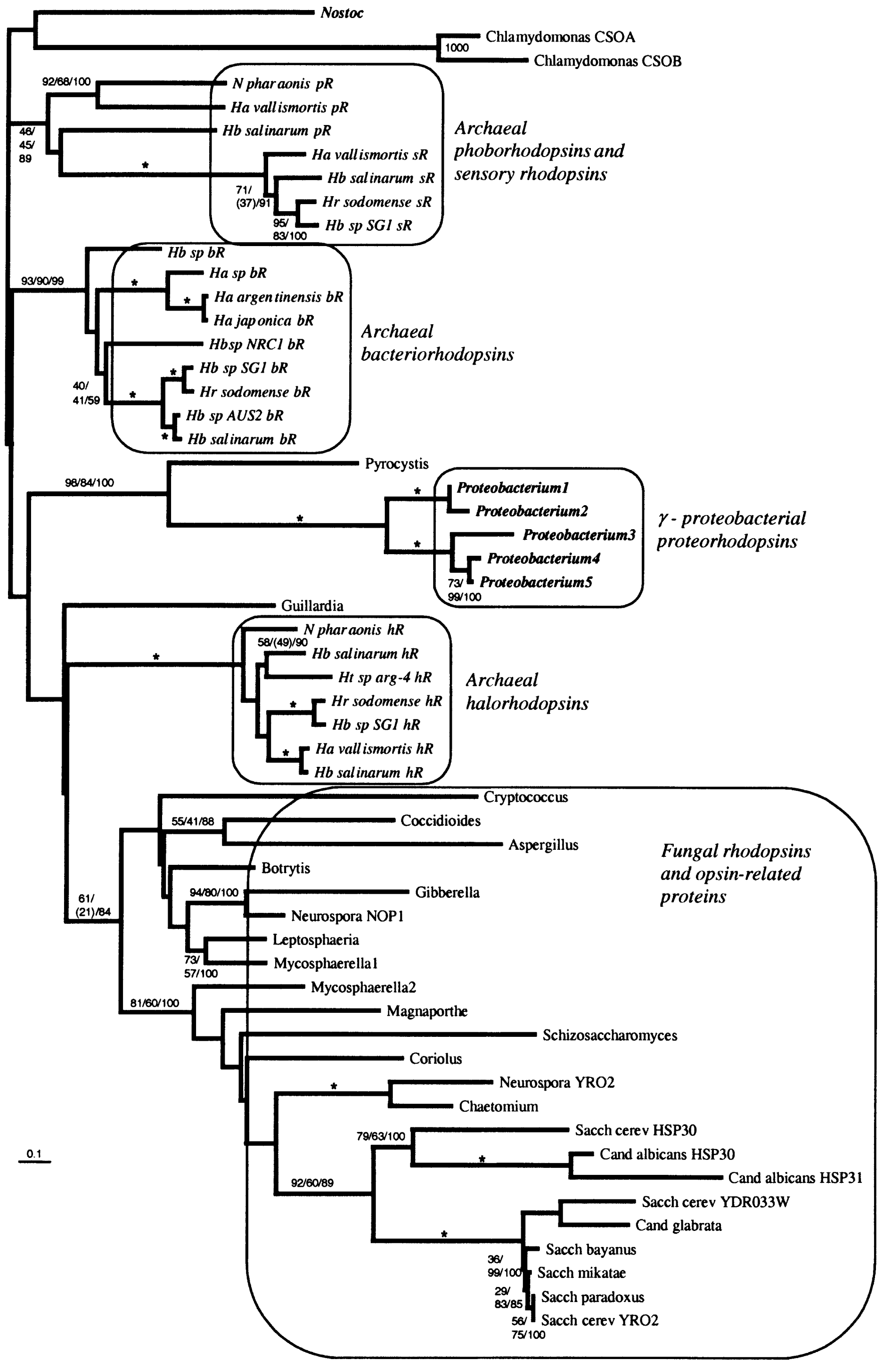

Fig. 3. Summary of the trees obtained with full-length rhodopsin sequences. Numbers refer to the bootstrap or RELL values supporting each branch (as percentages), shown as NJ/MP/ML (see Materials and Methods for details). Only values for those branches supported by the three methods of phylogenetic reconstruction and in which at least two of them gave values of support above 40% are shown. Asterisks refer to branches with support percentages above 95% for each of the three methods of phylogenetic reconstruction. When the MP value is in parentheses, it means that those branches were supported by the primary MP analysis but were not present in the most parsimonious trees obtained in the more exhaustive MP analysis made by generating 100,000 random trees (see Materials and Methods). The lack of support for a monophyletic fungal branch is caused by the *Cryptococcus* sequence appearing separated from the other fungi and together with the putative *Guillardia* rhodopsin.

is also the expected direction of transfer considering the relative likelihood of accepting foreign genes in eukaryotic versus prokaryotic species.

However, we are still left with a second problem: how species that belong to two evolutionarily very distant eukaryotic groups, dinoflagellates and plants, can have proteins that are so similar (Fig. 2). In this case, we favor the idea that the putative *Oryza* sequence indeed may have not originated from any plant genome. This truncated sequence is found in an isolated 676-bp-long fragment annotated as the "whole genome shotgun sequence," but it has not been detected in the assembled genomic DNA of either of the two *Oryza* varieties that have been extensively sequenced. The fully sequenced *Arabidopsis* and the other partially sequenced plant genomes also seem to lack rhodopsins. Moreover, all the other supposedly plant rhodopsin sequences have been demonstrated to be of contaminant origin (see above). The likelihood of this putative *Oryza* rhodopsin sequence being a false positive (i.e., not a plant sequence) is thus high. Its origin is mysterious, however, because it is clearly substantially different from all the other rhodopsins found so far.

Accepting that one relatively recent horizontal transfer may have occurred, the question remains whether the patchy distribution observed is the consequence of several additional ancient horizontal transfer events. Until very recently, support for an ancient origin for type 1 rhodopsins was very weak. They presented an extremely unusual phylogenetic distribution, appearing in just a few closely related species of archaea, proteobacteria, or fungi. These results were hardly compatible with an ancient origin, instead favoring explanations based on horizontal transfer events. However, the demonstration that rhodopsin genes exist in several other organisms such as green algae (Nagel et al. 2002; Sineshchekov et al. 2002) and, especially, cyanobacteria (Jung et al. 2003), which are among the most distant relatives of proteobacteria within the eubacterial domain, weakens the support for horizontal transmission-based hypotheses. In fact, this broader phylogenetic range is not unique anymore: We have found, using COGs-based analyses, three other genes with similar phylogenetic distributions. Considering the limitations of those analyses mentioned in the previous section, it is likely that more exist. It is therefore more reasonable now to suggest that type 1 rhodopsins may have emerged before the splits that gave rise to eubacteria, eukaryotes, and archaea. Actually, Kyrpides and Woese (1998) arrived at that conclusion for a case with a similar phylogenetic distribution. Our description of rhodopsins in additional groups leads to an even broader phylogenetic range and, thus, increases the likelihood of an extremely ancient origin for these genes. However, our results do not fully dismiss the possibility of several ancient horizontal transfers involving totally unrelated organisms. In summary, we envisage three main alternatives.

(1) Extremely ancient origin plus a single recent horizontal transfer: Type 1 rhodopsin origin predates the splits separating eubacteria, archaea, and eukaryotes, and a single horizontal transfer event occurred between eukaryotes and proteobacteria. In this case, we must postulate multiple losses in many independent lineages.

(2) Ancient horizontal transfer to archaea + symbiosis + recent horizontal transfer: Type 1 rhodopsins originated in eubacteria and were later horizontally transferred to archaea. They were transferred to eukaryotes by symbiosis, because they were present in the bacterial ancestors of mitochondria (α-proteobacteria) and chloroplasts (cyanobacteria). However, only a few lineages still conserve these genes: fungi (derived from mitochondria) and organisms (green algae, cryptomonads, dinoflagellates) derived from the ancestral cyanobacteria–eukayote symbiosis that generated chloroplast-containing eukaryotes or that obtained cyanobacterial genes through secondary symbioses (see Stechmann and Cavalier-Smith 2002; Yoon et al. 2002).

(3) Two or more ancient horizontal transfers + a recent horizontal transfer: Rhodopsins originated in a domain of life (perhaps eubacteria, considering the very old dichotomy cyanobacteria/proteobacteria) and they were long ago horizontally transferred once to archaea and one or more times to eukaryotes. More recently, an additional horizontal transfer occurred from eukaryotes to eubacteria.

Only additional information, especially characterization of type 1 rhodopsins in more types of organims, may discriminate among these alternatives.

We close with two final notes. First, our phylogenetic reconstructions did not consistently group together sensory or transport rhodopsins. It is potentially possible to establish whether sensory and transport rhodopsins are two evolutionarily independent branches, which diverged long ago, or whether functional shifts between those two types of proteins have occurred in particular lineages. However, that demonstration will clearly depend on our ability to determine all the horizontal transfer events involving these genes that occurred in the past. Second, the available data suggest that type 2 rhodopsins may be of much more recent origin than type 1 rhodopsins. However, the question of whether they evolved from type 1 rhodopsins in eukaryotes or arose independently is still open.

Acknowledgments. We thank J. Woodland Hastings for sharing unpublished data on the Pyrocystis opsin. This work was supported by a Generalitat Valenciana grant (BM-011/2002) and a MCYT grant (GEN2001-4851-C06-02), which provided the funds required for computer equipment.

References

Adachi J, Hasegawa M (1992) Computer science monographs. Vol. 27: MOLPHY: Programs for molecular phylogeny I-PROT-ML: Maximum likelihood inference of protein phylogeny. Institute of Statistical Mathematics, Tokyo

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Archibald JM, Keeling PJ (2002) Recycled plastids: A “green movement” in eukaryotic evolution. Trends Genet 18:577–584

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–976

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN, DeLong EF (2000) Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. Science 289:1902–1906

Béjà O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001) Proteorhodopsin phototrophy in the ocean. Nature 411:786–789

Bieszke JA, Braun EL, Bean LE, Kang S, Natvig DO, Borkovich KO (1999a) The nop-1 gene of Neurospora crassa encodes a seven transmembrane helix retinal-binding protein homologous to archaeal rhodopsins. Proc Natl Acad Sci USA 96:8034–8039

Bieszke JA, Spudich EN, Scott KL, Borkovich KA, Spudich JL (1999b) A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment. Biochemistry 38:14138–14145

Cavalier-Smith T (2002) The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. Int J Syst Evol Micr 52:7–76

Deane JA, Farunholz M, Su V, Maier UG, Martin W, Durnford DG, McFadden GI (2000) Evidence for nucleomorph to host nucleus gene transfer: Light-harvesting complex proteins from cryptomonads and chlorarachniophytes. Protist 151:239–252

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG (2001) The highly reduced genome of an enslaved algal nucleus. Nature 410:1091–1096

Dvornyk V, Vinogradova O, Nevo E (2003) Origin and evolution of circadian clock genes in prokaryotes. Proc Natl Acad Sci USA 100:2495–2500

Feng DF, Cho G, Doolittle RF (1997) Determining divergence times with a protein clock: update and reevaluation. Proc Natl Acad Sci USA 94:13028–13033

Graul RC, Sadee W (1997) Evolutionary relationships among proteins probed by an iterative neighborhood cluster analysis (INCA). Alignment of bacteriorhodopsins with the yeast sequence YRO2. Pharm Res 14:1533–1541

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001) Molecular evidence for the early colonization of land by fungi and plants. Science 293:1129–1133

Idnurm A, Hewlett BJ (2001) Characterization of an opsin gene from the ascomycete Leptosphaeria maculans. Genome. 44:167–171

Ihara K, Umemura T, Kategiri I, Kitajima-Ihara T, Sugiyama Y, Kimura Y, Mukohata Y (1999) Evolution of the archaeal rhodopsins: Evolution rate changes by gene duplication and functional differentiation. J Mol Biol 285:163–174

Jung KH, Trivedi VD, Spudich JL (2003) Demonstration of a sensory rhodopsin in eubacteria. Mol Microb 47:1513–1522

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 31:151–160

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe

Kyrpides NC, Woese CR (1998) Universally conserved translation initiation factors. Proc Natl Acad Sci USA 95:224–228

Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, Spudich EN, Spudich JL, Béjà O (2003) Diversification and spectral tuning in marine proteorhodopsins. EMBO J 22:1725–1731

Moreira D, Guyader HL, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. Nature 405:69–72

Nagel G, Ollig D, Furhmann M, Kateriya S, Musti AM, Bamberg E, Hegemann P (2002) Channelrhodopsin-1: A light-gated proton channel in green algae. Science 296:2395–2398

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Nicholas KB, Nicholas Jr HB (1997) GeneDoc: A tool for editing and annotating multiple secuence alignments. Distributed by the authors (http://www.psc.edu/biomed/genedoc/)

Okamoto OK, Hastings JW (2003) Novel dinoflagellate clock-related genes identified through microarray analysis. J Phycol (in press)

Okamoto OK, Liu L, Robertson DL, Hastings JW (2001) Members of a dinoflagellate luciferase gene family differ in synonymous substitution rates. Biochemistry 40:15862–15868

Page RD (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Redecker D (2002) New views on fungal evolution based on DNA markers and the fossil record. Res Microbiol 153:125–130

Redecker D, Kodner R, Graham LE (2000) Glomalean fungi from the Ordovician. Science 289:1920–1921

Rodriguez-Trelles F, Tarrio R, Ayala FJ (2002) A methodological bias toward overestimation of molecular evolutionary time scales. Proc Natl Acad Sci USA 12:8112–8115

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 34:544–545

Sineshchekov OA, Jung KH, Spudich JL (2002) Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii. Proc Natl Acad Sci USA 99:8689–8694

Spudich JL, Yang CS, Jung KH, Spudich EN (2000) Retinylidene proteins: Structures and functions from archaea to humans. Annu Rev Cell Dev Biol 16:365–392

Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. Science 297:89–91

Suzuki MT, Béjà O, Taylor LT, DeLong EF (2001) Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. Environ Microbiol 3:323–331

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278:631–637

Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV (2001) The COG database: New developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29:22–28

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Van de Peer Y, Baldauf SL, Doolittle WF, Meyer A (2000) An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. J Mol Evol 51:565–576

Yokoyama S (2000) Molecular evolution of vertebrate visual pigments. Prog Ret Eye Res 19:385–419

Yoon HS, Hackett JD, Bhattacharya D (2002) A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. Proc Natl Acad Sci USA 99:11724–11729

Zhai Y, Heijne WHM, Smith DW, Saier Jr MH (2001) Homologues of archaeal rhodopsins in plants, animals and fungi: Structural and functional predications for a putative fungal chaperone protein. Biochim Biophys Acta 1511:206–223