Genomic Blueprints of Vision: A Comparative History of Rhodopsin Evolution and Phototransduction

[1]*Aastha Maheshwari and* [1, 1*]*Ritu Rai*[*]

[1, 1*]*Department of Zoology, Dyal Singh College, University of Delhi, Delhi, India*

maheshwariaastha1@gmail.com, riturai@dsc.du.ac.in

Abstract:

Light is the fundamental signal for living entities as life on Earth is ultimately sustained by light energy. Many animals have evolved complex sensory mechanisms to utilize light cues for regulating homeostatic systems such as eyesight and the circadian clock. Organisms adapt to diverse lighting by modifying visual-pigment subtypes and their spectral tuning. These changes affect both color perception and dim-light vision. Through the lens of Comparative Genomics, evolution of opsins acts as a bridge between molecular genetics and ecological adaptation. Opsins are a diverse family of G-protein coupled receptors (GPCRs) that serve as the fundamental molecular interface between light and biological signaling. Rhodopsin is the 7-helical G-coupled receptor transmembrane protein. It also contains 11-cis-retinal, bound to the opsin protein, primarily absorbs light and responsible for scotopic vision in dim light conditions. This study compared Rhodopsin gene at gene sequence level, coding region and protein level among the twenty -six selected organisms from diverse groups including a set of model organisms such as *Drosophila melanogaster*, *Danio rerio*, *Xenopus laevis*, *Rattus norvegicus*, *Gallus gallus*, *Macaca mulatta*, *Pan troglodyte* and *Homo sapiens*. The sequence alignment studies show percentage similarities across the genomes and derive different functional annotations that exhibit a diverse role for Rhodopsin. Further the phylogenetic analysis shows how they evolved over the time. The comparative genomic analysis of rhodopsin across diverse organisms effectively demonstrates how evolutionary pressures shape genetic architecture to meet specific environmental demands. Rhodopsin plays a vital role in studying GPCRs vividly, mutations in rhodopsin lead to diseases like Retinitis Pigmentosa and Congenital Stationary Night Blindness which disrupts normal vision and also affect the circadian rhythm, dysregulation causing sleeping disorders. They serve as valuable biomarkers for the early retinal degeneration and detection of neurodegenerative diseases such as Alzheimer's and Parkinson's.

Key words: Rhodopsin, G-coupled photoreceptor, vision, comparative genomics, evolution

Introduction:

Rhodopsin are the seven transmembrane, retinal binding proteins that mediate light detection in virtually all domains of life, serving as visual and non -visual phototransduction, maintaining circadian clock in vertebrates and as ion pumps or sensory receptors in microbes. (Kojima, K., & Sudo, Y. 2023). It was discovered in the 1930s as a visual pigment in rod receptors present in animal retina. (Tansley, 1931; Wald, 1935). It contains chromophore 11-cis retinal which is bound to rhodopsin protein. During photo-activation, the absorption of light causes 11-cis retinal to converts to all trans configuration leading to conformational changes in protein and results in a series of chemical reactions (Wald, G. 1953; Morton, R. A., and Pitt, G. A. J. 1957; Dartnall, H. J. A. 1962). This activated rhodopsin stimulates a

GPCR protein vividly known as transducin, which sends signals to the brain and is responsible for scotopic vision or dim light vision.

Rhodopsin is reported to be present in all domains of life including Bacteria, Archaea and Eukarya. Rhodopsin is categorized as microbial rhodopsin also termed as Type1 rhodopsin, majorly functions as light driven pumps and ion channels. Upon phototransduction all-trans configuration of the proteins gets converted to 13-cis configuration (Smith et al., 1985; Oesterhelt, 1998). This Type 1 rhodopsin are identified in different organisms includes archaea like (*Halobacterium salinarum)*, including cyanobacteria, unicellular eukaryotes (algae, fungi, yeast) and in choanoflagellates and viruses as well. (Lamarche et al., 2017; Bratanov et al., 2019; Zabelskii et al., 2020; Rozenberg et al., 2021; Govorunova et al., 2022b; Nagata and Inoue, 2022). *Halobacterium salinarum* rhodopsin was discovered in the 1970s, coined as bacteriorhodopsin. (Oesterhelt and Stoeckenius, 1971; Oesterhelt and Hess, 1973; de Grip WJ and Ganapathy S, 2022).

Animal rhodopsin also termed as Type 2 rhodopsin are GPCR proteins that mediate photon-induced signaling in vertebrate and invertebrate eyes. Animal rhodopsin also perform non -visual functions like Melanopsin, found in the brain and eyes, and may play a role in circadian cycles and papillary reflexes. (Provencio et al., 1998). Neuropsin (Opn5) is primarily expressed in neural tissues. (Tarttelin et al., 2003). The brain and visceral organs express Encephalopsin (Blackshaw et al.,1999). RGR opsin, found in the Retinal Pigment Epithelium (RPE) and Muller cells, serves as a photoisomerase. (M. Jhang et al.,1993; D Shen et al.,1994). Peropsin is expressed in RPE cells (Sun et al.,1997). The Bovine rhodopsin, the first 3D high resolution structure of rhodopsin was solved via X-ray crystallography (Palczewski et.al., 2000).

Comparative phylogenomic studies reveals that microbial and animal rhodopsin have seven transmembrane structures, despite having less or negligible sequence similarity suggests that deep structural convergence and distant evolutionary links (Mackin et al.2014). This work aims to construct the ancestral rhodopsin repertory, map events across major clades, trace the origins of ion-pumping, sensory, kinase functions and assess how structural and functional diversification correlates with ecological niches such as dim-light habitats and shows their species - species spectral tuning.

2. Materials and methods:

2.1. Sequence selection: Reference sequences were obtained by searching public genomic databases such as NCBI databases with key word "Rhodopsin". Sequence search was followed by different runs of "BLAST" searches. Finally twenty-six rhodopsin sequences were selected for the study. Multiple rhodopsin sequences were obtained for many organisms. The selected rhodopsin sequences belong to organisms from diversified groups. The rhodopsin gene and protein sequences used in the present study are summarized in Table 2.1. The table also shows the chromosomal location of the gene in respective organisms.

Table 2.1: NCBI accession numbers, location with their Gene ID of gene sequence and protein sequences of twenty six diverse organisms.

Sl. No.	Organism (common name)	Gene ID	Accession no. (NCBI) gene sequences	Accession no. (NCBI) protein sequences	Chromosomal Location
1.	*Homo sapiens* (Human)	6010	NC_000003.12	NP_000530.1	Chromosome 3
2.	*Pan troglodytes* (Chimpanzee)	460685	NC_086015.1	XP_063663190.1	Chromosome 2
3.	*Macaca mulatta* (Rhesus monkey)	702931	NC_133407.1	XP_001094250.1	Chromosome 2
4.	*Rattus norvegicus* (Rats)	24717	NC_086022.1	NP_254276.1	Chromosome 4
5.	*Bos taurus* (Cattle)	509933	NC_037349.1	NP_001014890.1	Chromosome 22
6.	*Tursiops truncatus* (Common bottlenose Dolphin)	101320374	NC_047043.1	NP_001267588.1	Chromosome 10
7.	*Ornithorhynchus anatinus* (Platypus)	100048940	NC_041749.1	NP_001121099.1	Chromosome X1
8.	*Gallus gallus* (Chicken)	751791	NC_052543.1	NP_001384426.1	Chromosome 12
9.	*Alligator mississippiensis* (Alligator)	102559120	NC_081835.1	NP_001274211.1	Chromosome 12
10.	*Xenopus laevis* (African clawed frog)	380209	NC_054377.1	NP_001080517.1	Chromosome 4L
11.	*Latimeria chalumnae* (Coelacanth fish)	102353290	NC_088152.1	XP_005997879.1	Chromosome 14
12.	*Danio rerio* (Zebra fish)	30295	NC_133183.1	NP_571159.2	Chromosome 8
13.	*Petromyzon marinus* (Sea lamprey)	116952281	NC_133712.1	XP_032827379.1	Chromosome 46

14.	*Branchiostoma floridae* (Florida lancelet)	118415307	NC_049979.1	XP_035675706.1	Chromosome 1
15.	*Strongylocentrotus purpuratus* (Sea urchin)	100888012	NW_022145538.1	XP_030839863.1	Chromosome Unknown
16.	*Octopus bimaculoides*(California two spot octopus)	106878312	NC_069003.1	XP_014782988.1	Chromosome 23
17.	*Magallana gigas* (Oyster)	105333166	NC_088854.1	XP_011434309.2	Chromosome 2
18.	*Drosophila melanogaster* (Fruit fly)	34615	NT_033779.5	NP_477096.1	Chromosome 2L
19.	*Anopheles gambiae* (Mosquito)	1279519	NC_064602.1	XP_319247.2	Chromosome 3
20.	*Apis mellifera* (Honey bee)	406128	NC_037651.1	NP_001011606.1	Chromosome LG14
21.	*Danaus plexippus* (Monarch butterfly)	116772435	NC_083545.1	XP_032520522.2	Chromosome 12
22.	*Schistosoma mansoni* (Blood fluke)	8340877	NC_031502.1	XP_018655262.1	Chromosome W
23.	*Acropora millepora* (staghorn Coral)	114955892	NC_058068.1	XP_029188659.1	Chromosome 3
24.	*Methanothermobacter marburgensis* (Archaea)	77400104	NZ_CP069376.1	WP_013296133.1	NA
25.	*Natronomonas pharaonis DSM 2160* (Archaea)	3703211	NC_007426.1	UEB91126.1	NA

26.	*Halobacterium salinarum* (Archaea)	68694180	NZ_CP085882.1	WP_136361479.1	NA

2.2. Sequence Alignment: Multiple sequence alignment using ClustalW with default parameters was performed. Comparison of *Homo sapiens* Rhodopsin gene sequences (6706 bp) and protein sequence (348 aa) has been done with twenty- five different organisms showing its percentage similarity.

Table 2.2.: Multiple sequence alignment with similarity scores with respect to *Homo sapiens* gene and protein sequence.

Sl. No.	**Organism (Common name)**	**Gene Sequences (bp)**	**Similarity score w.r.t *Homo sapiens* gene sequences (%)**	**Protein Sequences (aa)**	**Similarity score w.r.t *Homo sapiens* protein sequences (%)**
1.	*Pan troglodytes* (Chimpanzee)	6667 bp	98.6201	348 aa	100
2.	*Macaca mulatta* (Rhesus monkey)	6629 bp	59.9638	348 aa	98.2759
3.	*Rattus norvegicus* (Rats)	5163 bp	34.631	348 aa	95.1149
4..	*Bovine taurus* (Cattle)	6329 bp	38.5369	348 aa	93.3908
5.	*Tursiops truncatus* (Dolphin)	4966 bp	44.4221	348 aa	91.954
6.	*Ornithorhynchus anatinus* (Platypus)	3916 bp	35.24	353 aa	91.3793
7.	*Gallus gallus* (Chicken)	4525 bp	32.0221	351 aa	85.9195
8.	*Alligator mississippiensis* (Alligator)	4333 bp	31.2024	352 aa	85.6322
9.	*Xenopus laevis* (African clawed frog)	3883 bp	30.4146	354 aa	81.8966
10.	*Latimeria chalumnae* (Coelacanth fish)	6525 bp	22.8966	355 aa	80.1724
11.	*Danio rerio* (Zebra fish)	1549 bp	53.1956	354 aa	79.3103
12.	*Petromyzon marinus* (Sea lamprey)	28665 bp	23.3671	353 aa	79.023

13.	*Branchiostoma floridae* (Florida lancelet)	13478 bp	17.7602	390 aa	16.092
14.	*Strongylocentrotus purpuratus* (Sea urchin)	21190 bp	17.8199	519 aa	17.5287
15.	*Octopus bimaculoides* (California two spot octopus	4328 bp	18.415	455 aa	20.6897
16.	*Magallana gigas* (Oyster)	7775 bp	16.9997	354 aa	19.5402
17.	*Drosophila melanogaster* (Fruit fly)	1680 bp	21.6071	382 aa	21.2644
18.	*Anopheles gambiae* (Mosquito)	1821 bp	20.4833	386 aa	21.2644
19.	*Apis mellifera* (Honey bee)	2591 bp	18.2555	377 aa	18.1034
20.	*Danaus plexippus* (Monarch butterfly)	7703 bp	16.8208	379 aa	19.2529
21.	*Schistosoma mansoni* (Blood fluke)	7575 bp	16.6418	190 aa	19.4737
22.	*Acropora millepora* (staghorn Coral)	8941bp	17.8646	407 aa	10.6322
23.	*Methanothermobacter marburgensis* (Archaea)	825 bp	25.4545	274 aa	12.7737
24.	*Natronomonas pharaonis DSM 2160* (Archaea)	720 bp	21.9444	239 aa	14.6444
25.	*Halobacterium salinarum* (Archaea)	789 bp	21.2928	262 aa	12.2137

2.3 Exon sequence alignment: In order to study the sequence similarity for coding region of the Rhodopsin gene, Multiple sequence alignment of five exon sequences of *Homo sapiens* was done using CLUSTAL W.

Table 2.3: Sequence alignment of five exon sequences of Rhodopsin gene of *Homo sapiens*. The sequence showing highest similarity with the human exon are highlighted.

Sl. No.	**Organism (Common name)**	**Exon1 (96…456)**	**Exon 2 (2238…2406)**	**Exon 3 (3613…3778)**	**Exon 4 (3895...4137)**	**Exon 5 (4970...5080)**
1.	*Pan troglodytes* (Chimpanzee)	99.72	99.40	98.79	99.16	**100**
2.	*Macaca mulatta* (Rhesus monkey)	95.30	97.04	93.97	**98.33**	96.39

3.	*Rattus norvegicus* (Rats)	87.56	**97.04**	90.36	86.66	88.28
4..	*Bos taurus* (Cattle)	90.33	**94.67**	84.93	88.75	86.48
5.	*Tursiops truncatus* (Dolphin)	90.60	**92.89**	87.34	90	84.68
6.	*Ornithorhynchus anatinus* (Platypus)	78.17	84.02	82.53	**89.58**	80.18
7.	*Gallus gallus* (Chicken)	80.11	82.25	86.747	**87.91**	67.56
8.	*Alligator mississippiensis* (Alligator)	77.34	76.92	78.31	**81.66**	65.76
9.	*Xenopus laevis* (African clawed frog)	68.50	74.55	**81.32**	80.41	55.85
10.	*Latimeria chalumnae* (Coelacanth fish)	69.33	76.92	**85.54**	77.08	58.55
11.	*Danio rerio* (Zebra fish)	76.24	78.96	77.71	**80.83**	58.55
12.	*Petromyzon marinus* (Sea lamprey)	25.96	72.78	72.89	**78.33**	66.66
13.	*Branchiostoma floridae* (Florida lancelet)	30.93	**35.50**	30.72	31.25	28.82
14.	*Strongylocentrotus purpuratus* (Sea urchin)	25.13	30.76	28.91	**35.83**	27.92
15.	*Octopus bimaculoides* (California two spot Octopus)	*25.69*	29.58	**37.95**	33.75	34.23
16.	*Magallana gigas* (Oyster)	27.90	25.44	29.15	**34.58**	32.43
17.	*Drosophila melanogaster* (Fruit fly)	26.24	31.36	**36.74**	33.75	32.43

18.	*Anopheles gambiae* (Mosquito)	23.75	27.21	**42.16**	27.5	33.33
19.	*Apis mellifera* (Honey bee)	24.58	27.81	**31.92**	24.16	28.82
20.	*Danaus plexippus* (Monarch butterfly)	25.69	30.17	**38.55**	25	28.82
21.	*Schistosoma mansoni* (Blood fluke)	26.51	26.62	**28.91**	25	24.32
22.	*Acropora millepora* (staghorn Coral)	23.20	27.21`	30.12	27.5	**30.63**
23.	*Methanothermobacter marburgensis* (Archaea)	24.86	**29.58**	29.51	29.16	28.82
24.	*Natronomonas pharaonis DSM 2160* (Archaea)	26.51	**31.95**	31.32	29.58	31.53
25.	*Halobacterium salinarum* (Archaea)	24.30	**33.72**	28.31	27.91	30.63

2.4. Phylogenetic Tree Construction: A phylogenetic reconstruction was performed using the Neighbour-Joining implemented in Clustal W, based on the distance matrix generated from the multiple sequence alignment and tree visualization was done using ETE3 (Huerta et al., 2016). To provide evolutionary directionality among the different taxa, the resulting tree was midpoint rooted, and reliability of the tree was evaluated by bootstrap value. The branch length depicts the amount of change that occurred between the nodes. The resulting tree was then visualized and interpretations of results had been done.

Results:

Graphical representation of the Table 2.2 is shown in the Figure 1 illustrates that Mammalian cluster from *Pan troglodytes* to *Tursiops truncatus* gene similarity is lower comparative to the protein which is up to 100% within the closest relative like *Pan troglodytes*. Organisms like *Alligator mississippiensis*, *Xenopus laevis*, *Latimeria chalumnae* etc. shows gene sequence similarity between (20-50%) and protein level (≈ 80 %) , depicts that functional pressures preserve key residues across the animal kingdom while allowing neutral drift in the nucleotide sequence. *Danio rerio* display a surprisingly high gene percentage similarity (53 %) despite being a teleost; this reflects a compact rhodopsin coding region with fewer introns. *Drosophila melanogaster* and other arthropods also show progressive decline in both gene and protein sequence similarity. Archaeons like *Methanothermobacter marburgensis*, *Natronomonas*

pharaonis DSM 2160, *Halobacterium salinarum* show gene similarity (24-35%) and protein similarity as (> 28%) fall near the bottom.

The parallel bars for each species show that, despite many synonymous nucleotide changes, the protein similarity remains relatively high because functional constraints on rhodopsin's seven-transmembrane fold limit amino acid substitutions. This conservation across vertebrates explains why rhodopsin is a classic model for studying GPCR evolution and why even distant species retain the core phototransduction function.

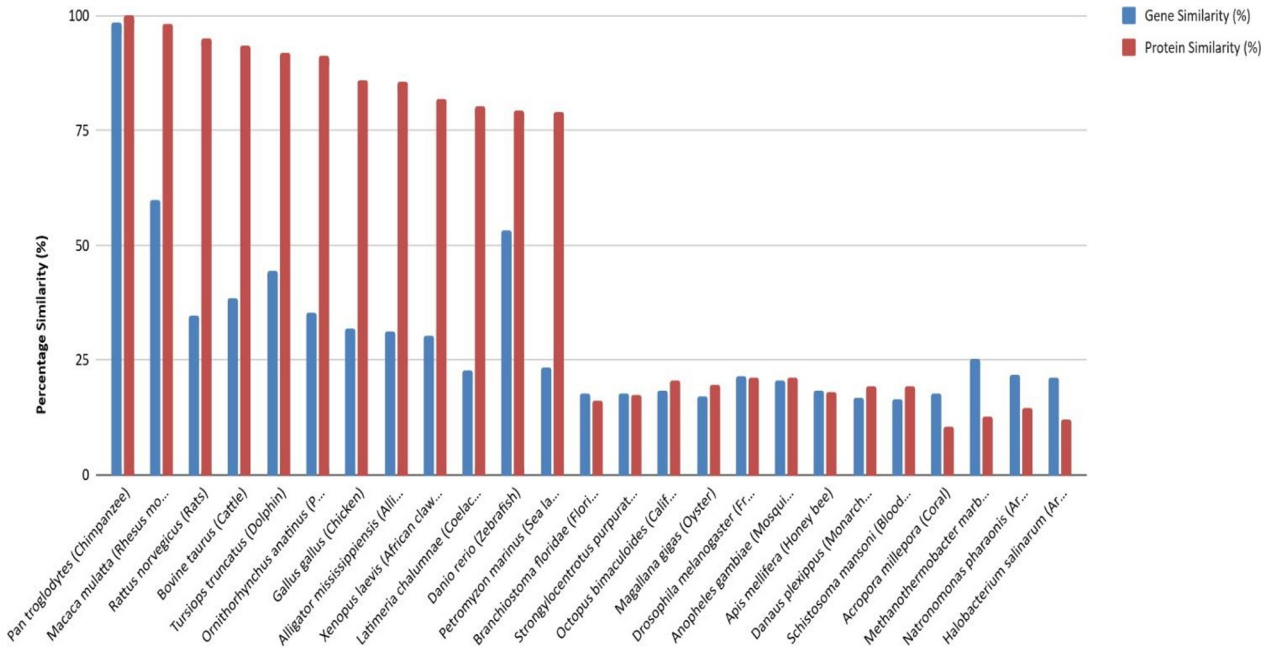

Figure 1. Sequence similarity of the selected organisms with *Homo sapiens* rhodopsin at the gene level and protein level.

Comparative analysis of *Homo sapiens* rhodopsin gene sequence with twenty-five taxonomically diverse organisms (Table 2.2) ranging from Archaea *Halobacterium salinarum* to primates *Pan troglodytes* represent percentage identity or similarity scores after multiple sequence alignment by Clustal W. It shows that *Homo sapiens* rhodopsin gene sequence shows highest similarity with *Pan troglodytes* of about 98.62% and 59.96% with *Macaca mulatta* signifies that gene sequence is highly conserved within the primate lineage. Lowest similarity is shown with invertebrates like *Schistosoma mansoni* of 16.64%, *Danaus plexippus* 16.82% and with archaea like *Halobacterium salinarum* 21.29% reflects the deep evolutionary distance and potential functional divergence.

Phylogenetic analysis with the phylogram mentioned in Figure 2 visualizes the evolutionary relationship based on the similarity score. Vertebrate cluster or lineage form a single large cluster includes jawed vertebrates like *Latimeria chalumnae* (coelacanth), *Petromyzon marinus* (lamprey), *Xenopus laevis* (amphibian), *Gallus gallus* (birds), *Danio rerio* (teleost) *and Ornithorhynchus anatinus* (monotreme) etc. and mammals like *Homo sapiens, Pan troglodytes, Macaca mullata, Bos taurus and Rattus norvegicus.* The branch length between *Pan troglodytes* and *Homo sapiens* is extremely short (0.0203 substitution/site,100% robustness,) least divergence reflects that their rhodopsin gene sequences are nearly identical while the mammal fish split indicates deeper divergence and evolution over the time.

Invertebrate lineage of *Drosophila melanogaster*, *Anopheles gambiae* (Arthropoda) and *Octopus bimaculoides* (Mollusca), *Strongylocentrotus purpuratus* (Echinodermata), *Acropora millepora* (cnidaria), and *Branchiostoma floridae* (cephalochordate) group together, separating from the vertebrate clade. This reflects the ancient split between bilaterian and non-bilaterian metazoans. Archaeon like *Halobacterium salinarum, Natronomonas pharaonis DSM 2160* and M*ethanothermobacter marburgensis* (methanogen) occupy the position closer to mammalian clade. Their inclusion provides a distant reference point for rooting the tree and for estimating absolute divergence between microbial and animal rhodopsin.

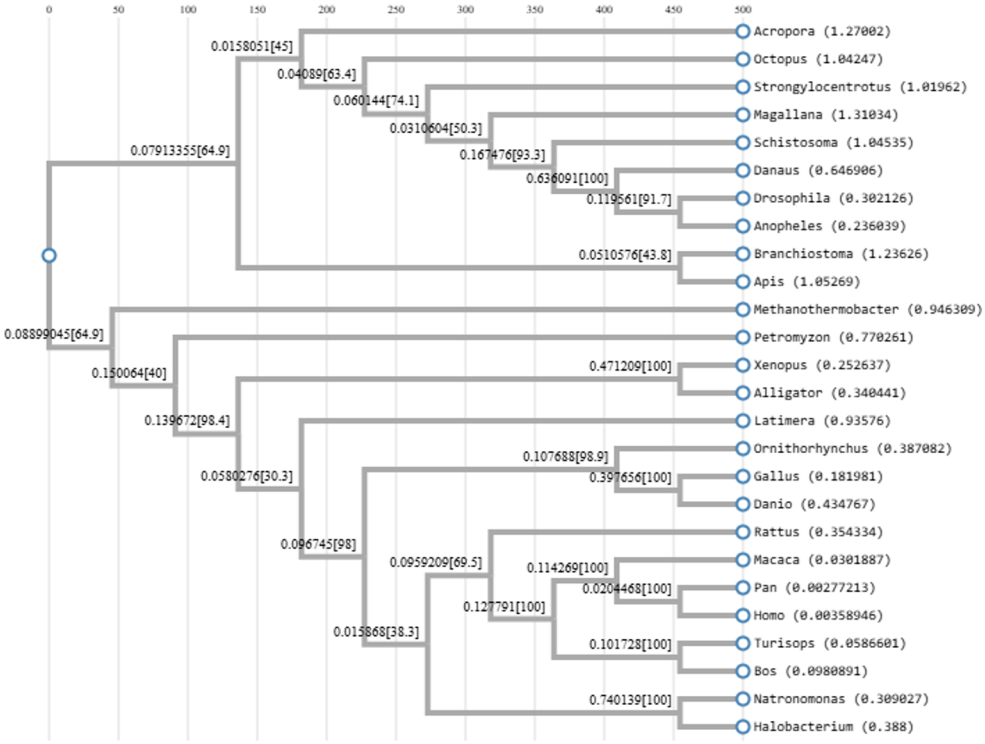

Figure 2: Phylogram (mid-point rooted tree, without branch length) showing evolutionary relationship of Rhodopsin gene sequences in twenty-six different organisms.

Comparative analysis of protein sequences of Homo sapiens with 25 different organisms (Table 2.1) on the basis of their similarity score by Clustal W depicts that *Homo sapiens* Rhodopsin protein sequence shows highest similarity with *Pan troglodytes* of about 100% which shows that the protein structure has remained virtually unchanged since the last common ancestor. Vertebrate mammals cluster includes organisms *Macaca mulatta*, *Bos taurus*, *Rattus norvegicus*, *Tursiops truncatus* shows ≥ 90 % similarity, reflects strong conservation of 7-transmembrane scaffold and retinal binding residues which are vital for phototransduction. Organisms like *Danio rerio*, *Xenopus laevis* shows (≈ 80 %) moderate similarity, while *Gallus gallus, Alligator mississippiensis* shows about (≈ 85 %) similarity

varying and adapting to their ecological niches. Non-vertebrates chordate like *Branchiostoma floridae, Petromyzon marinus* shows about 30–40 % similarity, indicates their divergence after the emergence of vertebrate visual system. *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Danaus plexippus* display similarity scores between 20–38 %, expressed the functional shift of rhodopsin in Arthropoda. Archaea like *Halobacterium salinarum, Natronomonas pharaonis DSM 2160, Methanothermobacter marburgensis* exhibits lowest similarity approx. 12-15% which shows that there is some kind of similarity between type-1 and type-2 rhodopsin like having some protein structure similarity, (retinal binding pocket and seven transmembrane) but differ in overall topology, molecular function. (Kojima, K.et al.,2023)

Phylogenetic analysis (Fig 2) interprets that Mammalian cluster of *Macaca mulatta*, *Bos taurus*, *Rattus norvegicus*, *Tursiops truncatus* display very short internal branches length (≤0.06), reflecting recent divergence; *Homo sapiens* and *Pan troglodytes* share a zero-length branch, indicating essentially identical rhodopsin protein sequences Bootstrap support for these vertebrate nodes is uniformly high (≥92 % for most, 100 % for the *Pan* node), confirming the robustness of the inferred relationships. A large vertebrate and invertebrate cluster (≈0.42) branch length includes invertebrates like *Acropora millepora,* depicting functional divergence. *Strongylocentrotus purpuratus* reflects a very large branch length (≈ 4.42) means that its sequence accumulated more changes due to different visual environment, distinct functional context etc. than others. Within the archaeon branch *Methanothermobacter marburgensis* (≈2.89) diverges earliest followed by *Natronomonas pharaonis DSM 2160* (≈1.18) and *Halobacterium salinarum* (≈1.41).

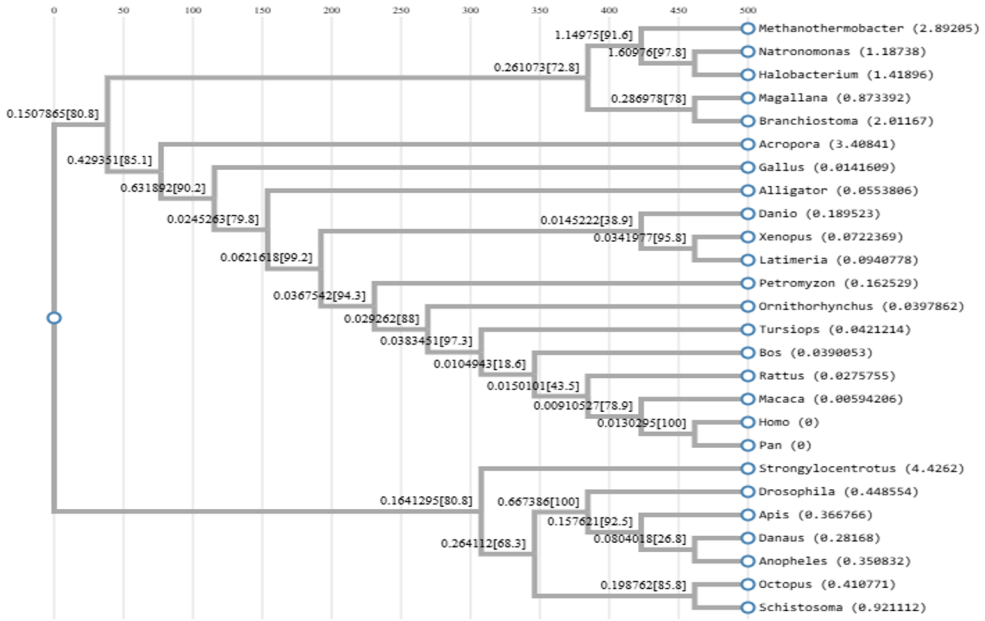

Figure 3: Phylogram (mid-point rooted tree, without branch length) showing evolutionary relationship of Rhodopsin protein sequences in twenty-six different organisms.

Comparative analysis of the exon sequences of *Homo sapiens* with twenty -five organisms (Table 2.3) in graph figure 4 shows highest percentage similarity interprets conserved sequences and phylogenetic proximity. Primates like *Pan troglodytes* show similarity (100 %) with Exon 5 sequence and (>98.79%) with other exon sequences as well, reflecting that the protein structure has remained unchanged till the last common ancestor. Other mammals like *Macaca mullata, Rattus norvegicus* and *Bovine taurus* maintain high scores (>85%), particularly in Exon 2 and Exon 4. *Gallus gallus*, *Alligator mississippiensis, Danio rerio, Petromyzon marinus* (>75%) shows highest similarity with exon 4 and depicts its critical functional domain of rhodopsin protein. *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Danaus plexippus* (23-42 %) shows highest similarity with Exon3. *Strongylocentrotus purpuratus* show the highest score with exon 4, (≈ 35%) highlighting divergence after the emergence of the vertebrate visual system Archaea *Halobacterium salinarum* display background similarity only with exon 2 (≈ 34%), underscoring the distant evolutionary relationship between type-1 microbial rhodopsin and animal rhodopsin. The gradient of identity from high (mammals) to low (invertebrates, microbes) reflects the functional constraints on the seven-transmembrane scaffold and the retinal-binding pocket. Positions with > 30 % identity across all groups likely correspond to structurally indispensable residues, whereas exon-specific variability particularly in Exon 3 may shows spectral adaptations to different light environments. These patterns play a vital role in detailed phylogenetic and functional analyses of rhodopsin evolution.

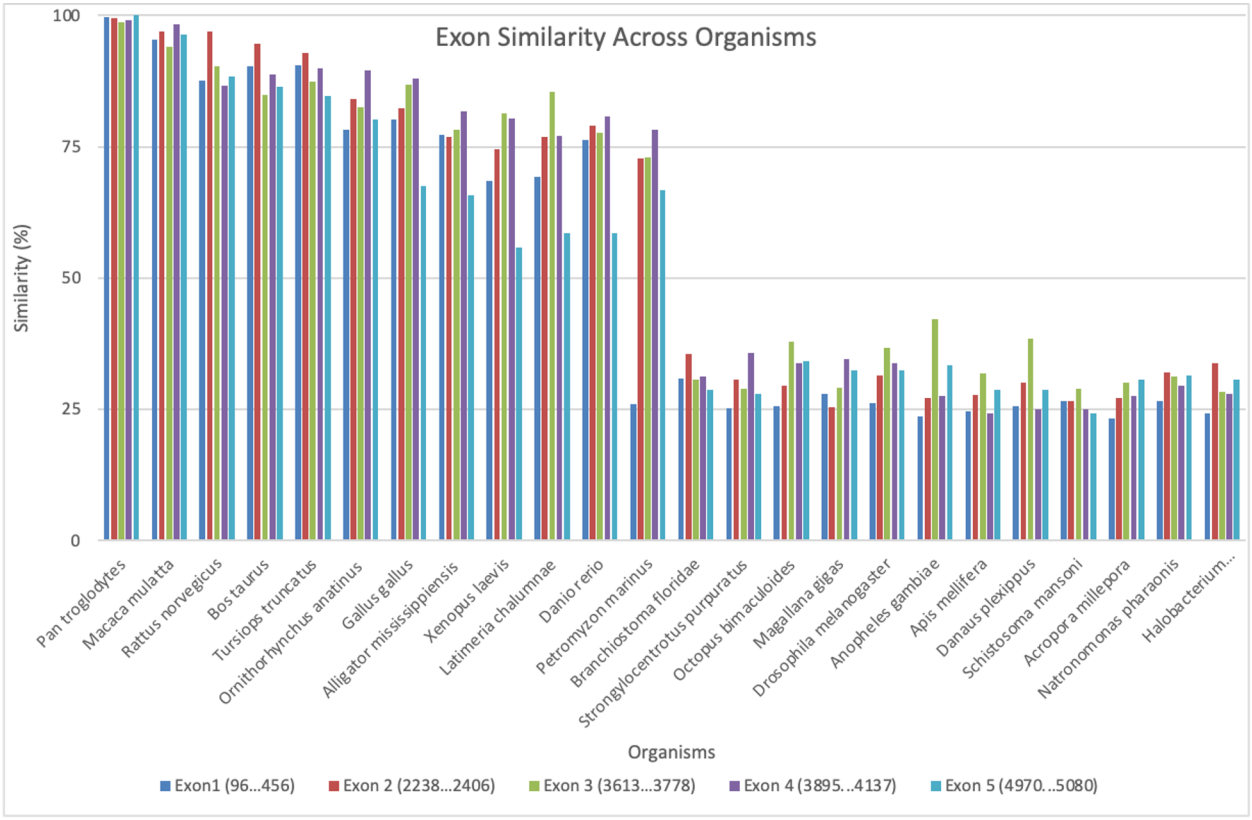

Figure 4: Graphical representation of sequence similarity of Rhodopsin gene at exon level.

Phylogenetic analysis of all the exons, Figure 5(a-e) depicts that Exon 5 is the most conserved across all the taxon. In Exon 5 phylogram (Fig 5e) the branch length is extremely short and well supported clade is seen with bootstrap of (≈ 92 %) which shows only few nucleotide changes accumulated over the last common ancestor, a distinctive feature of purifying selection. The high conservation of Exon 5 reflects that it encodes a core functional domain that tolerates little change likely essential for protein stability or catalytic activity. Mutations would likely disrupt protein stability or activity, leading to strong negative selection. Exon 5 changes little, it provides a stable molecular anchor for rooting the other exon trees and for estimating relative rates. Researchers can use its branch lengths as a baseline to when assessing selective pressures on the more variable exons and highlighting where adaptive evolution may have occurred. While Exon 3 exhibits longer branches, indicating towards lineage-specific adaptation.

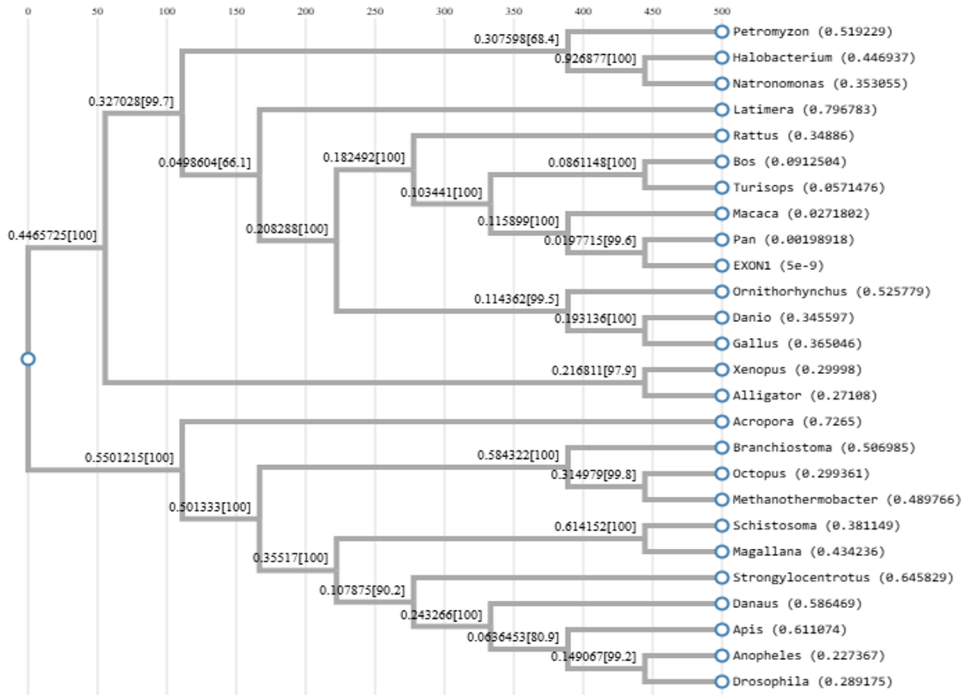

Figure 5(a): Phylogram (mid-point rooted tree, without branch length) of Exon 1 sequence of *Homo sapiens* with twenty -five different organisms.

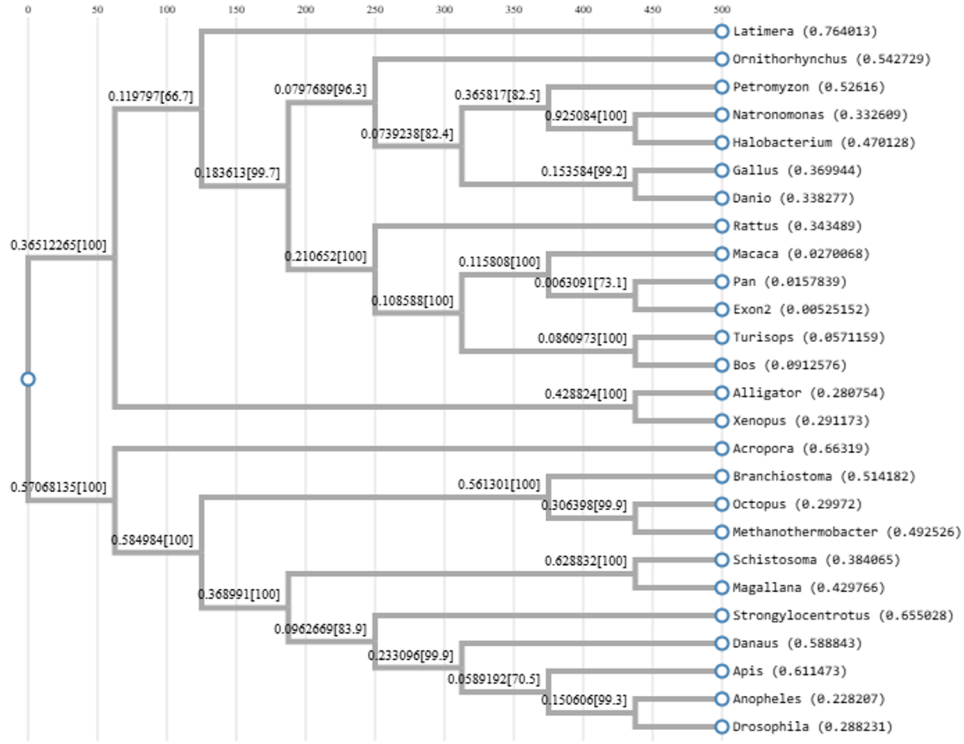

Figure 5(b): Phylogram (mid-point rooted tree, without branch length) of Exon 2 sequence of *Homo sapiens* with twenty-five different organisms.

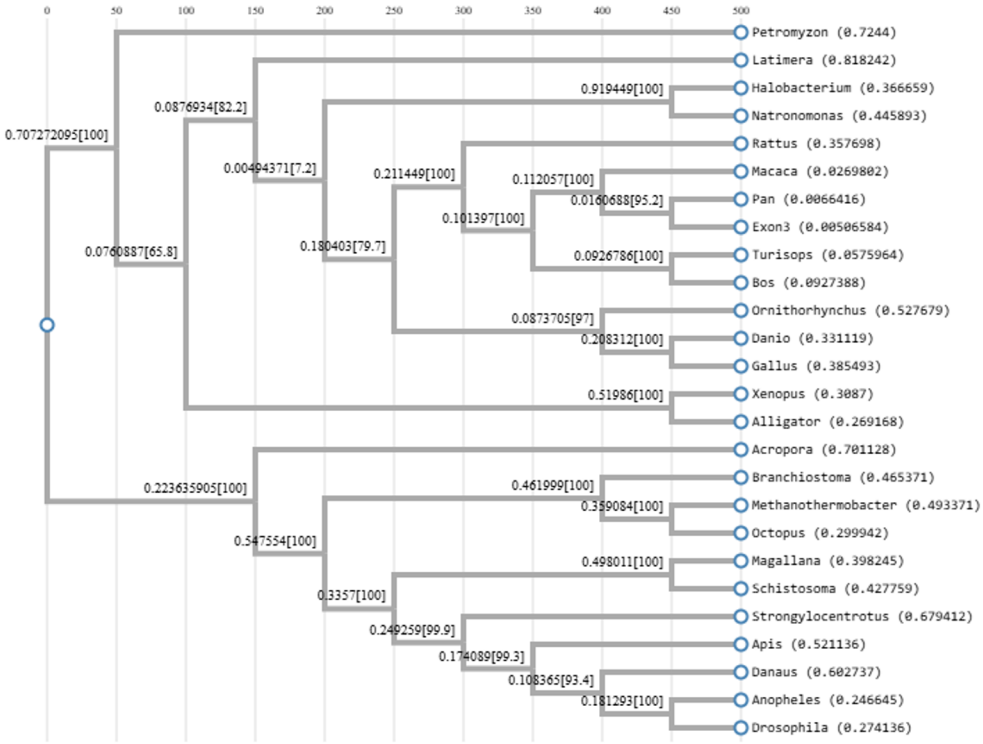

Figure 5(c): Phylogram (mid-point rooted tree, without branch length) of Exon 3 sequence of *Homo sapiens* with twenty -five different organisms.

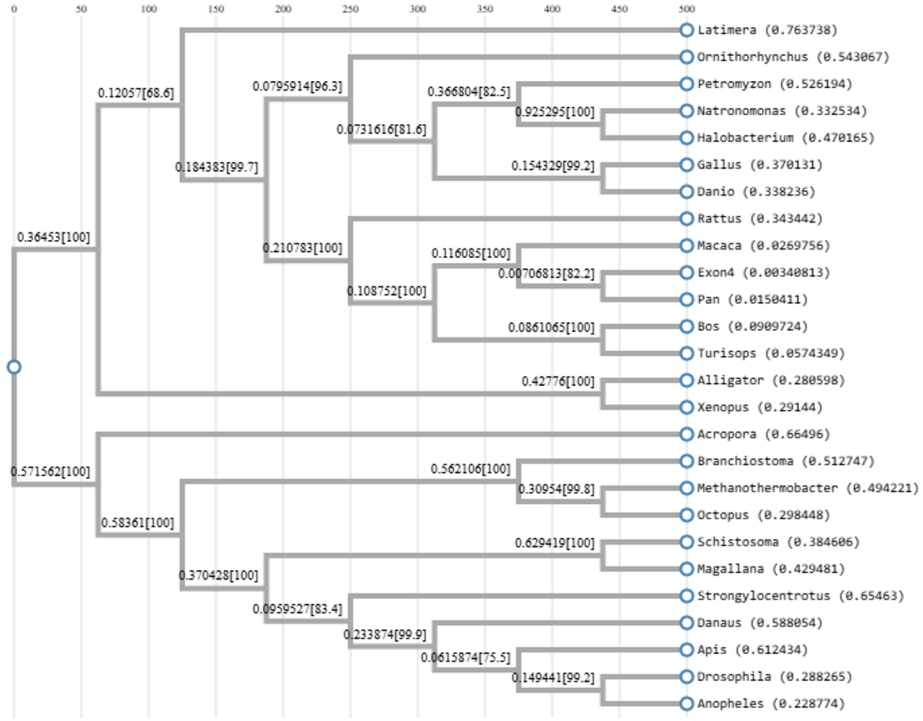

Figure 5(d): Phylogram (mid-point rooted tree, without branch length) of Exon 4 sequence of *Homo sapiens* with twenty- five different organisms.

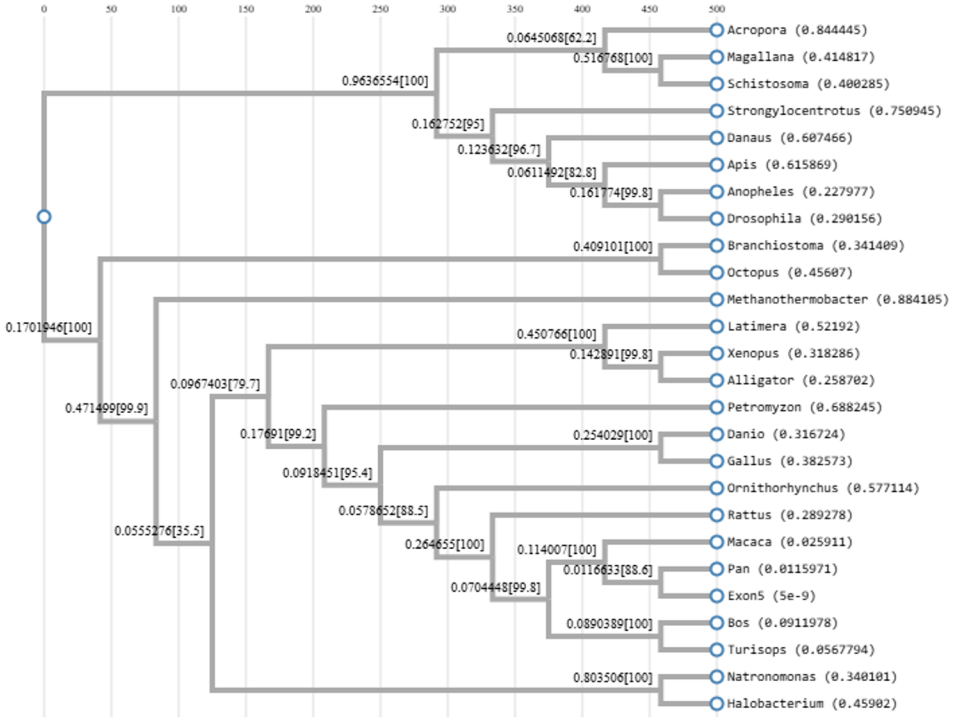

Figure 5(e): Phylogram (mid-point rooted tree, without branch length) of Exon 5 sequence of *Homo sapiens* with twenty- five different organisms.

Conclusion:

Based on our analysis, Comparison of *Homo sapiens* with twenty-five different organisms reflects that the gene sequence similarity of *Homo sapiens* with *Pan troglodytes* is 98.62% and protein sequence similarity of *Homo sapiens* with *Pan troglodytes* is 100% respectively which states that gene sequence is highly conserved and protein structure remains unchanged over the last common ancestor, act as distinctive feature of purifying selection. Similarly, it shows lowest similarity with other organisms like *Danaus plexippus* 16.82% and with archaea like *Halobacterium salinarum* 21.29% reflects the deep evolutionary distance and potential functional divergence. All organisms have exon-specific similarity but Exon 4 and 5 shows highest similarity among primates signifying that it may have some residues critical for retinal binding and G-protein coupling, maintain structural integrity of the protein. Variability in Exon 3 exhibits lineage-species adaptation. *Strongylocentrotus purpuratus* show the highest score with exon 4, (≈ 35%) highlighting divergence after the emergence of the vertebrate visual system. This analysis mainly focuses on how Rhodopsin gene evolved over the time to meet the species-species spectral tuning demand like *Danio rerio* and a group of teleost in which gene duplication in the rhodopsin gene shapes the biodiversity in the entire lineage and rooting the basis of low light condition and adapting according to ecological niche over the time. (Chen et al., 2018; Morrow et.al., 2011). Diving mammals like *Tursiops truncatus*, adapts to darker environments and adjusts to low light conditions (Dungan et al., 2022). The parallel bars for each species in the graphs mentioned above shows that despite many synonymous nucleotide changes, the protein similarity remains relatively high because functional constraints on rhodopsin's seven- transmembrane fold limit amino acid substitutions. This conservation across vertebrates explains why rhodopsin is a classic model for studying GPCR evolution and why even distant species retain the core phototransduction function. It focused on how comparative analysis play a vital role in finding phylogenetic proximity among different organisms.

References:

1. Blackshaw, S., & Snyder, S. H. (1999). Encephalopsin: a novel mammalian extraretinal opsin discretely localized in the brain. Journal of Neuroscience, 19(10), 3681-3690.

2. Bratanov, D., Kovalev, K., Machtens, J.-P., Astashkin, R., Chizhov, I., Soloviov, D., et al. (2019). Unique Structure and Function of Viral Rhodopsins. Nat. Commun. 10, 4939. doi:10.1038/s41467-019-12718-0

3. Chen, J. N., Samadi, S., & Chen, W. J. (2018). Rhodopsin gene evolution in early teleost fishes. *PloS one*, *13*(11), e0206918.

4. Dartnall, H. J. A. (1962a). "The Chemical Structure and Photochemistry of the Visual Pigments," in The Visual Process. Editor H. Davson. 1 ed (New York, U.S.A. Academic Press), 427–471

5. de Grip WJ and Ganapathy S (2022). Rhodopsins: An Excitingly Versatile Protein Species for Research, Development and Creative Engineering. Front. Chem. 10:879609. doi: 10.3389/fchem.2022.879609

6. Dungan, S. Z., & Chang, B. S. (2022). Ancient whale rhodopsin reconstructs dim-light vision over a major evolutionary transition: implications for ancestral diving behavior. *Proceedings of the National Academy of Sciences*, *119*(27), e2118145119.

7. Govorunova, E. G., Sineshchekov, O. A., and Spudich, J. L. (2022b). Emerging Diversity of Channel rhodopsins and Their Structure-Function Relationships. Front. Cell. Neurosci. 15, 800313. doi:10.3389/fncel.2021.800313.

8. Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular biology and evolution*, *33*(6), 1635-1638.

9. Jiang, M., Pandey, S., & Fong, H. K. (1993). An opsin homologue in the retina and pigment epithelium. Investigative ophthalmology & visual science, 34(13), 3669-3678.

10. Kojima, K., & Sudo, Y. (2023). Convergent evolution of animal and microbial rhodopsin. RSC advances, 13(8), 5367-5381.

11. Lamarche, L. B., Kumar, R. P., Trieu, M. M., Devine, E. L., Cohen-Abeles, L. E., Theobald, D. L., et al. (2017). Purification and Characterization of RhoPDE, a Retinylidene/Phosphodiesterase Fusion Protein and Potential Optogenetic Tool from the Choanoflagellate Salpingoeca Rosetta. Biochemistry 56, 5812–5822. doi:10.1021/acs.biochem.7b00519

12. Mackin, K. A., Roy, R. A., & Theobald, D. L. (2014). An empirical test of convergent evolution inrhodopsins. Molecularbiologyandevolution, 31(1),85–95. https://doi.org/10.1093/molbev/mst171

13. Morrow, J. M., Lazic, S., & Chang, B. S. (2011). A novel rhodopsin-like gene expressed in zebrafish retina. *Visual neuroscience*, *28*(4), 325-335.

14. . Morton, R. A., and Pitt, G. A. J. (1957). Visual Pigments. Fortschritte Chem. Org. Naturst. 14, 244–316. doi:10.1007/978-3-7091-7164-6_6

15. Nagata, T., and Inoue, K. (2022). Rhodopsins at a Glance. J. Cell. Sci. 134, jcs258989. doi:10.1242/jcs.258989

16. Oesterhelt, D. (1998). The Structure and Mechanism of the Family of Retinal Proteins from Halophilic Archaea. Curr. Opin. Struct. Biol. 8, 489–500. doi:10. 1016/s0959-440x(98)80128-0

17. Oesterhelt, D., and Hess, B. (1973). Reversible Photolysis of Purple Complex in Purple Membrane of Halobacterium Halobium. Eur. J. Biochem. 37, 316–326. doi:10.1111/j.1432-1033.1973.tb02990.x

18. Oesterhelt, D., and Stoeckenius, W. (1971). Rhodopsin-like Protein from the Purple Membrane of Halobacterium Halobium. Nat. New Biol. 233, 149–152. doi:10.1038/newbio233149a0

19. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., ... & Miyano, M. (2000). Crystal structure of rhodopsin: AG protein-coupled receptor. science, 289(5480), 739-745.

20. Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P., & Rollag, M. D. (1998). Melanopsin: An opsin in melanophores, brain, and eye. Proceedings of the National Academy of Sciences, 95(1), 340-345.

21. Rozenberg, A., Inoue, K., Kandori, H., and Béjà, O. (2021). Microbial Rhodopsins: The Last Two Decades. Annu. Rev. Microbiol. 75, 427–447. doi:10.1146/ annurev-micro-031721-020452

22. Shen, D., Jiang, M., Hao, W., Tao, L., Salazar, M., & Fong, H. K. (1994). A human opsin-related gene that encodes a retinaldehyde-binding protein. Biochemistry, 33(44), 13117-13125.

23. Smith, S. O., Myers, A. B., Mathies, R. A., Pardoen, J. A., Winkel, C., Vandenberg, E. M. M., et al. (1985). Vibrational Analysis of the All-Trans Retinal Protonated Schiff Base. Biophys. J. 47, 653–664. doi:10.1016/s0006-3495(85)83961-8

24. Sun, H., Gilbert, D. J., Copeland, N. G., Jenkins, N. A., & Nathans, J. (1997). Peropsin, a novel visual pigment-like protein located in the apical microvilli of the retinal pigment epithelium. Proceedings of the National Academy of Sciences, 94(18), 9893-9898.

25. Tansley, K. (1931). The regeneration of visual purple: its relation to dark adaptation and night blindness. The Journal of Physiology, 71(4), 442

26. Tarttelin, E. E., Bellingham, J., Hankins, M. W., Foster, R. G., & Lucas, R. J. (2003). Neuropsin (Opn5): a novel opsin identified in mammalian neural tissue. FEBS letters, 554(3), 410-416.

27. Wald, G. (1935). Carotenoids and the visual cycle. The Journal of general physiology, 19(2), 351-371.

28. Wald, G. (1953). The Biochemistry of Vision. Annu. Rev. Biochem. 22, 497–526. doi:10.1146/annurev.bi.22.070153.002433.

29. Zabelskii, D., Alekseev, A., Kovalev, K., Rankovic, V., Balandin, T., Soloviov, D., et al. (2020). Viral Rhodopsins 1 Are an Unique Family of Light-Gated Cation Channels. Nat. Commun. 11, 5707. doi:10.1038/s41467-020-19457-