G protein-coupled receptor diversity and evolution in the closest living relatives of Metazoa

Alain Garcia De Las Bayonas[1,2]* and Nicole King[1,2]

1 Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720
2 Howard Hughes Medical Institute, United States
*Corresponding author. Email: alaingdlb@berkeley.edu

ABSTRACT

G protein-coupled receptors (GPCRs) play a pivotal role in the perception of environmental cues across eukaryotic diversity. Although GPCRs have been relatively well characterized in metazoans, GPCR signaling is poorly understood in their sister group, the choanoflagellates, and in other close relatives of metazoans (CRMs). Here, we examine GPCR diversity and evolution in choanoflagellates by curating a catalog of 918 GPCRs, 141 G proteins, and 367 associated regulators from 23 choanoflagellate genomes and transcriptomes. We found that the repertoire of choanoflagellate GPCRs is larger and more diverse than previously anticipated, with 18 GPCR families found in choanoflagellates, of which 12 families are newly identified in these organisms. Comparative analyses revealed that most choanoflagellate GPCR families are conserved in metazoans and/or other eukaryotic lineages. Adhesion GPCRs and GPCR-TKL/Ks are the most abundant GPCRs in choanoflagellates. The identification of GPCR repertoires in CRMs and other non-metazoans refines our understanding of metazoan GPCR evolution and reveals the existence of previously unreported GPCR families in metazoans and at the root of the eukaryotic tree.

INTRODUCTION

G protein-coupled receptors (GPCRs) constitute one of the largest and oldest families of receptors used to sense extracellular cues in eukaryotes (Nordström et al. 2011; Krishnan et al. 2012; De Mendoza et al. 2014). A conserved seven transmembrane (7TM) domain is diagnostic for GPCRs, while the wide spectrum of extracellular and intracellular domains in some GPCRs reflects the diversification of the gene family and its functions (Schiöth and Lagerström 2008). For example, the extracellular N-terminus and the three extracellular loops of the 7TM domain respond to a wide range of cues, including photons, odorant molecules, peptides, amines, lipids, nucleotides, and other molecules (Yang et al. 2021). Ligand binding triggers a conformational change in the intracellular loops, thereby activating heterotrimeric GTP binding proteins (G proteins) and additional regulators to shape downstream cellular responses (Latorraca et al. 2017). Thus, GPCRs are essential for the control of metazoan development and tissue homeostasis, and their dysregulation often leads to pathological conditions in humans, making them the most researched drug targets in the pharmaceutical industry (Hauser et al. 2017).

Based on pioneering phylogenetic analyses in vertebrates, most metazoan GPCRs have been classified into five major families: Glutamate, Rhodopsin, Adhesion, Frizzled, and Secretin (GRAFS) (Fredriksson et al. 2003; Schiöth and Fredriksson 2005; Schiöth and Lagerström 2008a). Later studies found that four of these families (Glutamate, Rhodopsin, Adhesion, and Frizzled), alongside cyclic AMP (cAMP) Receptors, GPR-107/108-like GPCRs, GPCR PIPKs, and GPR180, exist in most metazoans and in diverse other eukaryotic lineages (Bakthavatsalam et al. 2006; Kamesh et al. 2008; Nordström et al. 2008; Krishnan et al. 2012; Krishnan et al. 2014; De Mendoza et al. 2014; Van Den Hoogen et al. 2018; Mojib and Kubanek 2020; Hall et al. 2023; Luo et al. 2023). Nonetheless, the premetazoan ancestry of many metazoan GPCRs remains unclear.

With the increasing availability of genomic and transcriptomic datasets from diverse early branching metazoans (Srivastava et al. 2010; Guzman and Conaco 2016; Gold et al. 2019; Nong et al. 2020; Schultz et al. 2021; Francis et al. 2023; Santini et al. 2023; Steffen et al. 2023; Vargas et al. 2024), the closest living relatives of metazoans, the choanoflagellates (King et al. 2008; Fairclough et al. 2013; Richter et al. 2018; Brunet et al. 2019; Hake et al. 2024), and other holozoans, including members of Filasterea, Ichthyosporea, and Corallochytrea (Suga et al. 2013; Torruella et al. 2015; Grau-Bové et al. 2017; Hehenberger et al. 2017; Ocaña-Pallarès et al. 2022; Sarre et al. 2024) (Fig. 1A), we set out to refine our understanding of GPCR evolution.

To this end, we analyzed the GPCR repertoires of 23 choanoflagellate species and diverse other eukaryotes. Because of the structural modularity of GPCRs, the canonical 7TM domain on one side and the associated N-terminal and C-terminal regions on the other can follow distinct evolutionary trajectories. In addition, many receptors grouped under the same GPCR family (e.g. Rhodopsins and aGPCRs) are subdivided into numerous subfamilies. Therefore, we described and compared GPCRomes at three different levels in our analysis: (1) family- and (2)

subfamily-level, based on the 7TM sequences of diverse GPCRs, and (3) associated protein domain composition, based on their respective N- and C-terminal regions.

RESULTS

Identification of GPCRs and downstream signaling pathway components in the proteomes of 23 diverse choanoflagellates

To catalog the diversity of GPCRs and components of the downstream signaling pathway encoded by choanoflagellates, we analyzed the genome-derived and transcriptome-derived predicted proteomes of 23 choanoflagellate species (Supplementary Table 1; King et al. 2008; Fairclough et al. 2013; Richter et al. 2018; Brunet et al. 2019; Hake et al. 2024). We first surveyed for candidate choanoflagellate GPCRs by searching choanoflagellate proteomes with Hidden Markov Models (HMMs) from the 7TM domains of 54 previously identified GPCR families (https://www.ebi.ac.uk/interpro/set/pfam/CL0192/; Fig. S1.1; Supplementary Table 2). In a complementary approach, we used an HMM that reflects the topology of the 7TM domains of all GPCRs (GPCRHMM; (Wistrand et al. 2006). These two approaches recovered 1095 and 1070 putative choanoflagellate GPCRs, respectively, of which 381 sequences were shared between the two datasets, leaving 1784 unique GPCR candidates. To remove potential false positives, highly fragmented sequences, and isoforms, the GPCR candidates were then subjected to additional filtering through homology- and topology-based approaches; 1113 sequences were removed after these filtering steps (Fig. S1.2).

Because existing 7TM HMMs may be biased toward metazoan sequences, we used the recovered choanoflagellate GPCRs to generate new choanoflagellate-derived 7TM HMMs (Supplementary file 2). To this end, the 671 validated choanoflagellate GPCRs were sorted by sequence similarity, resulting in 18 clusters (Table 1, Methods); the 76 GPCRs that did not assort into any of these clusters were kept for downstream analyses. We then built 7TM HMMs for each of the 18 choanoflagellate clusters and 76 additional HMMs for the unclustered GPCRs and used these custom HMMs to re-screen the 23 choanoflagellate proteomes. To complement this approach, we also selected GPCR sequences from each GPCR cluster and used them as BLAST queries to search for additional GPCRs in the 23 choanoflagellate proteomes. After filtering for fragmented sequences, isoforms, and false positives, we recovered 247 additional GPCRs that were not detected during the first round of screening. Together, our bioinformatic analysis identified a total of 918 choanoflagellate GPCRs (Supplementary file 1), with an estimated number of GPCRs per species ranging from 16 in *Barroeca monosierra* to 122 GPCRs in *Acanthoeca spectabilis* (Fig. S2).

To search for downstream signaling components of GPCRs, we aligned the 23 choanoflagellate proteomes with HMMs specific to the three classes of heterotrimeric G proteins (Gα, Gβ, and Gγ) (Neves et al. 2002; Dupré et al. 2009; Wootten et al. 2018), positive regulators (Ric8, Phosducin) (Willardson and Howlett 2007; Srivastava et al. 2019), and negative regulators (GRK, Arrestin, RGS, GoLoco) (Willard et al. 2004; Rajagopal and Shenoy 2018; Bagnato and Rosanò 2019; Gurevich and Gurevich 2019; O'Brien et al. 2019) of G protein signaling (Fig. S3

105 and Supplementary Table 2). Our analysis recovered 141 heterotrimeric G proteins, 308 positive regulators, and 59 negative regulators of G protein signaling (Supplementary text and Supplementary files 3, 4, and 5). Together with the abundant GPCRs detected in choanoflagellates, the finding of nearly complete G protein signaling pathways suggests that choanoflagellates engage in metazoan-like G protein signaling.

110 To assess the predictive power of our protein-detection pipeline, we then compared the new GPCR and cytosolic signaling component datasets from two genome-enabled choanoflagellates – *Salpingoeca rosetta* and *Monosiga brevicollis* – with previously published GPCR and downstream GPCR signaling component counts for these two species (Nordström et al. 2009a; Krishnan et al. 2012; De Mendoza et al. 2014; Krishnan et al. 2015; Lokits et al. 2018). All 11
115 GPCRs previously described in *M. brevicollis* and all 14 GPCRs found in *S. rosetta* were present in our dataset, along with 10 and 9 newly identified GPCRs, respectively. Similarly, all previously identified heterotrimeric G proteins and diverse positive and negative regulators of G protein signaling from these two choanoflagellates were in our dataset, along with one newly identified Gβ subunit, one RGS, one Goloco, and one additional Arrestin in *M. brevicollis*
120 (Fig.S3; Supplementary file 3). Due to the stringency of our filtering approach and the use of transcriptome-derived proteomes for a majority of choanoflagellate species, these numbers still likely underestimate total GPCR and downstream component abundance and diversity in choanoflagellates. Moreover, in the case of transcriptome-derived proteomes, failure to detect homologs could reflect false negatives rather than a real absence of the gene families.
125 Conversely, sequence contamination and splice isoforms could possibly account for false positives.

Identification of 18 GPCR families in choanoflagellates

The conservation and alignability of the 7TM domain enable the straightforward categorization of GPCRs into families across eukaryotic diversity (Attwood and Findlay 1993; Fredriksson et al.
130 2003; Schiöth and Fredriksson 2005). Therefore, to investigate the relationships among choanoflagellate GPCRs, we started by extracting and clustering their 7TM sequences. In addition, we included 7TM sequences from diverse metazoans, amoebozoans, stramenopiles, alveolates, and chlorophytes in our analysis to aid in the identification of the choanoflagellate GPCR families (Supplementary file 6).

135 Through all-against-all pairwise comparisons of the 7TM domains of the 918 choanoflagellate GPCRs recovered in the previous analysis (Fig. S1.3c) with 7TMs from diverse metazoans, amoebozoans, stramenopiles, alveolates, and chlorophytes, we found that most choanoflagellate GPCRs belong to 18 classes of GPCRs (Fig. 1B). Six of these classes were previously detected in *S. rosetta* and/or *M. brevicollis*: Glutamate, Adhesion, GPR107/108,
140 cAMP, GPR137, and GPCR_PIPK (Figs. 1B, 2, and S4; Table 1; Krishnan et al. 2012; De Mendoza et al. 2014; Van Den Hoogen et al. 2018; Gan et al. 2019)).

This left twelve new GPCR families that had not, to our knowledge, been previously detected in choanoflagellates: Rhodopsin, TMEM145, GPR180, TMEM87, GPR155, GPR157, and six additional GPCR families that appear to fall outside all previously characterized GPCR families

145 in eukaryotes. We have named these families “RémiSans-Famille” (RSF), “Hidden Gold” (Hi-GOLD), GPCR-TKL/K, GPRch1, GPRch2, and GPRch3. **(**Fig. 1B; Table 1). An additional 76 choanoflagellate GPCRs did not cluster above threshold with any other GPCRs detected from choanoflagellates or other eukaryotes.

Overview of choanoflagellate GPCR family evolution

150 Of the 18 GPCR families identified in choanoflagellates (Table 1), 16 were found, based strictly on the similarity of their signature 7TM domains, in diverse holozoan and non-holozoan eukaryotes and likely evolved in stem eukaryotes (Fig. 2; Supplementary File 7). GPR155 was only detected in metazoans and close relatives of metazoans (CRMs; Fig. 1A), supporting an origin in stem holozoans. The GPCR-TKL/K family was only detected in acanthoecid
155 choanoflagellates, suggesting it evolved in the acanthoecid stem lineage.

Although Adhesion, GPR107/108, TMEM87, TMEM145, Hi-GOLD, RSF, cAMP, GPR137, GPR155, and GPRch3 GPCRs were detected in a wide range of choanoflagellates, other GPCR families appeared to be restricted to one or the other of the two choanoflagellate orders: craspedids (GPR157, Glutamate, GPCR PIPK, and GPRch1) or acanthoecids (GPR180 and
160 GPCR TKL/K; Fig.2). The phylogenetic distributions of these GPCR families are likely the result of secondary losses in choanoflagellates because homologs of these GPCRs are detected in other eukaryotes.

Two additional choanoflagellate GPCR families evolved in stem eukaryotes but were lost from metazoans and some CRMs: RSF GPCRs and GPRch1. Frizzled GPCRs, which possibly
165 originated in the last common ancestor of opisthokonts and amoebozoans (Krishnan et al. 2012; De Mendoza et al. 2014), and GPR143, which emerged in stem holozoans (Fig.2; De Mendoza et al. 2014), were both lost from stem choanoflagellates.

Metazoan-like Rhodopsins were detected in two choanoflagellate sister species – *Salpingoeca macrocollata* and *Salpingoeca punica* (Richter et al. 2018; López-Escardó et al. 2019; Ginés-
170 Rivas and Carr 2025a). Their evolution is difficult to interpret as metazoan-like Rhodopsins are absent from all other choanoflagellates analyzed, filastereans, corallochytreans, and most other eukaryotes except metazoans, ichthyosporeans, holomycotans and amoebozoans (Fig. 2).

Metazoan GPCR families that originated in stem eukaryotes

Glutamate Receptors

175 The Glutamate Receptor family, which includes important regulators of neuronal excitability, synaptic transmission, and taste recognition in metazoans (Pin et al. 2003; Schiöth and Lagerström 2008a; Chun et al. 2012), first evolved in stem eukaryotes (Fig. 2; (Krishnan et al. 2012; De Mendoza et al. 2014)). In addition to their 7TM domain, Glutamate Receptors have a large, structured extracellular region that varies in different members of the family (Fig. 3A;
180 (Ellaithy et al. 2020a)). For example, mGluR/T1R/CaSR receptors exhibit a bi-lobed ligand-binding domain (ANF) directly N-terminal to a cysteine-rich domain (NCD3G) (Koehl et al.

2019), both of which are essential for receptor activity (Hu et al. 2000; Jiang et al. 2004; Rondard et al. 2006). In $GABA_B$ receptors, the ANF domain is conserved while the NCD3G is absent (Evenseth et al. 2020; Papasergi-Scott et al. 2020). Finally, mGlyR/GPR179 receptors present a distinct domain organization, with a ligand-binding Cache domain sitting atop an EGF-like domain (Jeong et al. 2021; Laboute et al. 2023; Rosenkilde and Mathiesen 2023; Yun et al. 2024).

While members of the Glutamate family and their canonical domain architectures are conserved in a wide range of metazoans, including sponges and ctenophores (Krishnan et al. 2014; De Mendoza et al. 2014; Krishnan and Schiöth 2015), it is presently unclear how CRM Glutamate Receptors identified by clustering of their 7TM domains relate to their metazoan counterparts. To better understand the evolution of Glutamate Receptors, we conducted a phylogenetic analysis of the 7TM domains from metazoan and CRM Glutamate Receptors (Fig. 3B and Supplementary files 8 and 9). All choanoflagellate Glutamate Receptors identified in our study formed the sister group to metazoan mGluR/T1R/CaSR GPCRs. The Glutamate Receptors of filastereans and ichthyosporeans branched either at the base of metazoan mGluR/T1R/CaSR receptors, mGlyR/GPR179 receptors, or $GABA_B$ receptors; or at positions on the tree that were distinct from the metazoan Glutamate Receptor families. The Glutamate Receptors from the amoebozan *Dictyostelium discoideum*, which bind both GABA and Glutamate (Anjard and Loomis 2006; Taniura et al. 2006; Wu and Janetopoulos 2013), grouped separately from metazoan and CRM GPCRs in our analysis.

Next, we assessed the protein domain compositions of the extracellular regions of Glutamate Receptors from diverse opisthokonts (Fig. 3C and Supplementary file 10). Although earlier studies reported no extracellular domains in choanoflagellate Glutamate Receptors (Krishnan et al. 2012; De Mendoza et al. 2014), our analysis revealed the presence of 5'-nucleotidase C-terminal domains (5_nucleotid_C) and Ephrin receptor-like domains (Ephrin_rec-like) that are also found in Glutamate Receptors from other CRMs, but not in metazoans.

Two domains present in the N-termini of some metazoan Glutamate Receptors – ANF and EGF – were also found in Glutamate Receptors from other opisthokonts, suggesting an ancient association of these domains with the signature Glutamate 7TM. On the contrary, no non-metazoan GPCRs yet identified possess the canonical metazoan mGluR/T1R/CaSR architecture of an ANF_receptor domain coupled with an NDC3G domain or the mGlyR/GPR179 architecture of a Cache domain coupled with an EGF domain. Together, our data suggest that while the 7TM domains of the three main classes of metazoan Glutamate Receptors – mGluR/T1R/CaSR, mGlyR/GPR179, and $GABA_B$ receptors – have a pre-metazoan origin, the stereotypical combination of domains in their N-terminus was a stem metazoan innovation.

Rhodopsins

Rhodopsins are the largest GPCR family in most metazoans and include receptors for hormones, neuropeptides, neurotransmitters, nucleotides, and light, among others (Fredriksson et al. 2003; Schiöth and Lagerström 2008; De Mendoza et al. 2014; Pándy-Szekeres et al.

2018). Orthologs of bilaterian Rhodopsins have been found in Cnidaria and Ctenophores but have not been detected in sponges (Feuda et al. 2014; Krishnan et al. 2014; Thiel et al. 2023). While Rhodopsins expanded and diversified in metazoans (Fredriksson and Schiöth 2005; De Mendoza et al. 2014; Thiel et al. 2023), the origin of this family might have predated the emergence of metazoans, as homologs have been detected in Fungi and in other eukaryotes (Fig. 2 ; Krishnan et al. 2012; De Mendoza et al. 2014).

Upon searching the choanoflagellate genomes and transcriptomes with an HMM profile built solely on the aligned 7TM domains from metazoan Rhodopsin sequences (7TM_1; Supplementary Table 2), we identified two previously undetected choanoflagellate Rhodopsins. Additionally, we detected a metazoan-like Rhodopsin from the ichthyosporean *Pirum gemmata* (Fig. 2 and Supplementary file 11).

All-against-all pairwise comparison of the choanoflagellate and ichthyosporean Rhodopsins with the complete Rhodopsin repertoires of representative metazoans (6149 Rhodopsins in total) revealed that choanoflagellate Rhodopsins are most similar to metazoan opsins (including arthropsin, cephalochordate Go opsin, placozoan placopsin, vertebrate opsin 3, and Teleost multiple tissue opsins) and, to a lesser extent, SOG receptors (Fig. 4A, B and Supplementary files 11 and 12 ; Fredriksson et al. 2003; Kamesh et al. 2008; Nordström et al. 2008; McVeigh et al. 2018; Yañez-Guerra et al. 2022). No statistically significant metazoan or choanoflagellate hits were detected for the predicted ichthyosporean Rhodopsin (Supplementary file 12).

To complement our sequence-based clustering approach, we then predicted the structures of the two choanoflagellate Rhodopsins and the ichthyosporean Rhodopsin using Alphafold 3 (Abramson et al. 2024). We searched for structural matches within three Alphafold databases: AFDB-PROTEOME, AFDB-SWISSPROT, and AFDB50 (van Kempen et al. 2024). In these searches, metazoan opsins (a subclass of Rhodopsins; Fig. 4A) were recovered as the closest structural matches to choanoflagellate Rhodopsins, with e-values ranging from e^{-12} to e^{-14}, with Melanopsin/OPN4, an opsin with non-visual and visual functions in metazoans (Hankins et al. 2008; Koyanagi et al. 2013; Karthikeyan et al. 2023), being the top hit for the two choanoflagellate Rhodopsins (Fig. 4C, D). In contrast, aminergic or peptide Rhodopsin receptors more closely matched the predicted structure of the ichthyosporean Rhodopsin candidate (Fig. S5). Notably, the predicted structures of choanoflagellate and ichthyosporean Rhodopsins exhibit a cytoplasmic helix eight (H8) located immediately after the end of the seventh transmembrane domain, a common feature of metazoan Rhodopsins that is involved in trafficking of the receptor to the cell membrane and in binding to G protein and β-Arrestin (Fig. 4C, D ; Krishna et al. 2002; Kock et al. 2009; Knepp et al. 2012; Sensoy and Weinstein 2015; Dijkman et al. 2020).

In metazoan Rhodopsins, conserved sets of amino acids in transmembrane passes 2, 3, 6, and 7 (TM2, TM3, TM6, and TM7), along with extracellular loop 2 (ECL2), serve important structural roles and as determinants of signal transduction (Fig. 4E, Supplementary files 13 and 14 ; Davies et al. 2010; Hankins et al. 2014; Nagata and Inoue 2021b). We found that two cysteines (C110 and C187) whose disulfide bridges often connect ECL2 and TM3 in metazoan Rhodopsins (Rader et al. 2004), were conserved in both the choanoflagellate and

ichthyosporean receptors. Similarly, asparagine D83 was conserved in TM2 of choanoflagellate and holomycotan Rhodopsins. Among the sites shown to be important for controlling G protein signaling, the CWxPY and NpxxY$(x)_{5,6}$FR motifs were partially conserved in choanoflagellate and ichthyosporean sequences. In contrast, the E/DRY motif, which might regulate the activation of metazoan Rhodopsins (Rovati et al. 2007; Sandoval et al. 2016), was lost from TM3 in the choanoflagellate Rhodopsins but partially conserved in the ichthyosporean Rhodopsin. Finally, none of the residues mediating light sensing in opsins – namely the highly conserved Lysine K269, which mediates the binding to the chromophore retinal and the glutamic acid E181 that serves as a counterion to the protonated Schiff base (Davies et al. 2010; Hankins et al. 2014; Nagata and Inoue 2021) – were detected in CRM or holomycotan Rhodopsins, suggesting that these residues (and their functions in light sensing) evolved in stem metazoans.

Adhesion GPCRs

In metazoans, Adhesion GPCRs (aGPCRs) are generally the second most abundant GPCR family after the Rhodopsin family (Kamesh et al. 2008; Nordström et al. 2008; Krishnan et al. 2014; De Mendoza et al. 2014). They are critical for the multicellular biology of metazoans, in which they regulate epithelial morphogenesis, neuronal development, and immunity and are implicated in the progression of various cancers (Hamann et al. 2015a; Langenhan et al. 2016; Liebscher et al. 2022). The signature 7TM domain of aGPCRs has been detected in CRMs, holomycotans, and diverse other lineages (Fig. 2 ; (Krishnan et al. 2012; De Mendoza et al. 2014)).

Phylogenetic analysis of the 7TM domains of choanoflagellates uncovered at least 19 subfamilies of aGPCRs (subfamilies I-XIX ; Figs. S6A and S6B; Supplementary files 15 and 16), of which only one, class VIII, is orthologous to a metazoan aGPCR subfamily (Bootstrap support 83%; Fig. 5A and Supplementary files 17 and 18), the ADGRV family (Weston et al. 2004; Hamann et al. 2015a; Scholz et al. 2019; Kusuluri et al. 2021). The other choanoflagellate aGPCR families presumably evolved on the choanoflagellate stem lineage or within choanoflagellates, as they are not detected in other lineages. In addition, a set of uncharacterized cnidarian and cephalochordate GPCRs grouped with a subset of filasterean aGPCRs (Bootstrap support 98%, Fig. 5A).

Unlike other GPCRs, metazoan aGPCRs have a large N-terminal extracellular region with diverse protein domains that are linked to the 7TM domain by a conserved GPCR-autoproteolysis-inducing (GAIN) domain (Fig. 5B ; (Araç et al. 2012; Langenhan et al. 2013; Hamann et al. 2015a)). Ligand binding can cause the N-terminal fragment (NTF) to be released from the rest of the protein at the GPCR proteolytic site (GPS), exposing a tethered agonist peptide (TA) that activates downstream G protein signaling (Barros-Álvarez et al. 2022; Kleinau et al. 2022; Ping et al. 2022; Xiao et al. 2022; Seufert et al. 2023). The structural complexity of the metazoan NTF, including the GAIN domain and additional extracellular domains (ECDs), appears restricted to holozoans, as holomycotans and other eukaryotes have shorter, unstructured N-termini (Krishnan et al. 2012). Therefore, a detailed analysis of aGPCR

subfamilies and their protein domains in CRMs promised to illuminate the evolution of aGPCRs after the holozoan-holomycotan divergence.

Although the phylogeny of aGPCRs in choanoflagellates was inferred solely based on the comparison of their 7TM domains, the resulting groups tended to cluster receptors that also shared sequence similarity and protein domain architecture in their NTFs (Fig. 5C and Fig. S7A). The connection between the 7TM phylogeny and the N-terminal protein domain architectures of different aGPCRs, which was previously reported for metazoan aGPCR families (Bjarnadóttir et al. 2007; Hamann et al. 2015a), is exemplified by the choanoflagellate aGPCR groups I, V, VI, VIII, XIII, and XV, whose NTFs generally contain one or multiple conserved protein domains that are shared among members (Fig. 5C).Comparison of metazoan and CRM NTFs revealed that only ADGRV receptors share robust N-terminal sequence similarity among metazoan, choanoflagellate, and filasterean aGPCRs (Fig. S8 and Supplementary file 19).

While the GAIN and aGPCR 7TM domains evolved before the origin of opisthokonts (Araç et al. 2012; Krishnan et al. 2012; De Mendoza et al. 2014), we detected the fusion of these two domains into a single module (GAIN/7TM) only in holozoan aGPCRs (Fig. 5D and Fig. S7B); therefore, the GAIN/7TM module likely evolved in stem holozoans. Furthermore, motifs necessary for self-proteolysis in metazoan GAIN domains (Araç et al. 2012; Stoveken et al. 2015; Kleinau et al. 2022; Ping et al. 2022; Xiao et al. 2022; Seufert et al. 2023) are also conserved in CRM aGPCRs, suggesting that these receptors may similarly undergo cleavage and activation (Fig. S9A and S9B; Supplementary file 20).

The presence of other aGPCR ECDs (Fig. 5B) likely evolved in the last common ancestor of filastereans and choanozoans. This is supported by the absence of ECDs from ichthyosporean, corallochytrean, or non-holozoan aGPCRs (Fig. S7B), coupled with the conservation of Laminin_G, EGF, LRR, FN3, Furin-like, Calxβ, PQQ, and RLD domains in the aGPCRs of filastereans with those of either choanoflagellates or metazoans (Fig. 5D, node “1”). We infer that the aGPCR ECD repertoire diversified in the choanozoan stem lineage (Fig. 5D, node “2), followed by lineage-specific acquisitions of additional ECDs in metazoans and choanoflagellates (Fig. 5D, node “3a” and “3b”; Fig. S7B). Interestingly, the HRM domain — previously identified only in the metazoan Secretin and aGPCR families (Fig. S10A and Schiöth and Lagerström 2008; Nordström et al., 2009b; Krishnan et al., 2012; Krishnan et al., 2014; Araç D, Sträter N 2016) — was also found in 22 choanoflagellate and 6 filasterean aGPCRs in our analysis (Fig. S10B). Like its positioning in metazoan aGPCRs (Araç et al. 2012; Prömel et al. 2012a), the HRM domain is located directly N-terminal to the GAIN domain in choanoflagellate and filasterean aGPCRs, with a comparable separation of ~280-330 amino acids between the C-terminus of the HRM domain and the N-terminus of the 7TM domain (Fig. S10B and S10C). These findings suggest that the HRM domain and its integration into an HRM-GAIN-7TM module likely originated in the last common ancestor of Filasterea and Choanozoa (Fig. 5D). Supporting this view, the HRM domain was absent from all non-holozoan proteomes analyzed, with one exception: the nucleariid *Fonticula alba* (Fig. S10D), a close relative of Fungi (Liu et al. 2009; Torruella et al. 2015). *Fonticula alba* encodes five non-GPCR transmembrane proteins that contain HRM domains (Supplementary file 21).

These findings clarify the evolution of key features of holozoan aGPCRs. While the
345 autoproteolytic GAIN domain (Araç et al. 2012), the aGPCR 7TM domain (Krishnan et al. 2012; De Mendoza et al. 2014), and possibly the HRM domain (this study) existed before the emergence of Holozoa, our analysis indicates that they were first combined into a single module in holozoans. The GAIN-7TM module originated in stem holozoans, followed by the positioning of the HRM domain N-terminal to the GAIN domain in the last common ancestor of Filasterea
350 and Choanozoa. This was followed by the acquisition of additional extracellular protein domains in choanozoans and filastereans (Fig. 5D).

cAMP Receptors

Twenty-eight choanoflagellate GPCRs grouped with cAMP Receptors from diverse other eukaryotes (Fig. 1B). These receptors possess short N- and C-termini and a unique 7TM
355 signature that distinguishes them from other GPCRs (Saxe et al. 1993; Krishnan et al. 2012; Greenhalgh et al. 2020a).

The cAMP Receptors were present in stem eukaryotes and are inferred to have given rise to GRAFS GPCRs (Nordström et al. 2011; Krishnan et al. 2012). cAMP Receptors were first described in the amoebozoan *Dictyostelium discoideum*, where they control chemotaxis,
360 aggregation, morphogenetic movements, cell growth, and many developmentally regulated genes (Klein et al. 1988; Johnson et al. 1993; Saxe et al. 1993; Louis et al. 1994). While cAMP Receptors bind to cAMP in *D. discoideum* (Juliani and Klein 1981; Caterina et al. 1995; Milne et al. 1997), the precise cAMP binding site(s) have yet to be experimentally identified (Greenhalgh et al. 2020b). Since then, cAMP Receptors have been identified in holozoans, holomycotans,
365 alveolates, plants, and other lineages (Fig. 2 ; (Krishnan et al. 2012; De Mendoza et al. 2014)). Notably, although present in diverse metazoans, cAMP Receptors have been lost from vertebrates. Previous studies found that the choanoflagellates *S. rosetta* and *M. brevicollis* each encode a single cAMP Receptor (Krishnan et al. 2012; De Mendoza et al. 2014). This study found cAMP Receptors in 17 additional choanoflagellate species, including craspedids and
370 acanthoecids (Fig. 2).

GPR137

GPR137 is a lysosomal receptor that controls mTORC1 translocation to the lysosomes and modulates the Hippo pathway (Gan et al. 2019). Moreover, GPR137 regulates autophagy, cell growth, and cell proliferation and is implicated in various cancers (Men et al. 2018; Iwasa et al.
375 2023; Li et al. 2024). GPR137 does not possess other protein domains outside the 7TM domain. The receptor is broadly conserved in metazoans, and homologs have been found in the choanoflagellate *Monosiga brevicollis* and the amoebozoan *D. discoideum* (Fig. 2 and (Gan et al. 2019)).

380 We detected GPR137 in a wide range of choanoflagellate species, as well as in filastereans, ichthyosporeans, and non-holozoans, including nucleariids and other eukaryotic lineages (Fig. 2). Thus, GPR137 probably evolved at the stem of the eukaryotic tree. Whether GPR137 controls cell growth and the cell cycle in non-metazoans remains to be tested.

GPR157

385 GPR157 Receptors promote neuronal differentiation of radial glial progenitors in mice, where they localize to the primary cilia and signal via a G(q)-protein that activates a phosphatidylinositol-calcium second messenger (Takeo et al. 2016). No additional domains outside the 7TM have been detected in GPR157.

GPR157 had previously been identified only in mammals (Takeo et al. 2016). In this study, we
390 detected GPR157 homologs in most metazoans, including sponges, as well as in choanoflagellates, filastereans, nucleariids, amoebozoans, rhizarians, and various other eukaryotes, suggesting that GPR157 likely evolved at the root of the eukaryotic tree (Figs. 2 and S11A; Supplementary File 22). Choanoflagellate and metazoan GPR157 share a level of sequence similarity with the 7TM domains of cAMP, aGPCRs, and GPCR PIPKs (Fig. 1B). In
395 addition, we noticed a cytoplasmic helix 8 (H8), better known from Rhodopsins, directly C-terminal to TM7 in choanoflagellate, metazoan, and other eukaryote GPR157 protein structure predictions (Fig. S11B). Striking conservation of motifs that have not previously been characterized was noted in TM2 ("LLxxLSL/V/IxD"), ECL2 ("WCWI/V/L"), and TM7 ("QGxxNxIxF") of the GPR157 TM domain (Fig. S11C).

400 ### GPR107/108, TMEM145, GPR180, and TMEM87 (GOST GPCRs)

We identified 63 choanoflagellate GPCRs that clustered with a group of metazoan GPCRs named "GOST" for Golgi-dynamics domain seven-transmembrane helix proteins (Fig. 1 and Table 1; (Hoel et al., 2022). All GOST members exhibit a Golgi dynamics (GOLD) domain in their extracellular region, which is possibly involved in ligand recognition, followed directly by the
405 7TM domain (Anantharaman and Aravind 2002; Hoel et al. 2022). GOST GPCRs include GPR107/108, TMEM145, GPR180, and TMEM87 – all found in choanoflagellates – along with GPR181 and Wntless, neither of which were detected in choanoflagellates.

The structural similarity between choanoflagellate and metazoan GOST receptors includes both the 7TM domain and the N-terminal GOLD domain (Fig. S12). Nonetheless, it is unclear
410 whether the GOST GPCRs are monophyletic, as GPR107/108/TMEM87 and GPR180/TMEM145 grouped separately in our analyses of the 7TM domains (Fig. 1B). Despite this uncertainty, all GOST members have homologs in holozoans and various other eukaryotes (Fig. 2), suggesting that they were present in stem eukaryotes.

GOST GPCRs are understudied, but the characterization of several members showed that they
415 localize at the Golgi and endocytic vesicles, suggesting a common function in trafficking membrane-associated cargo (Zhou et al. 2014; Hirata et al. 2015; Shin et al. 2020; Hoel et al. 2022; Mitrovic et al. 2024). In addition, a recent study found that TMEM145 also contributes to the structural integrity of hair cell stereocilia (Roh et al. 2025), expanding the functions associated with GOST GPCRs. Although ligands have been proposed for some GOST proteins
420 – neurostatin for GPR107 (Yosten et al. 2012; Yang et al. 2024), gambogic acid for GPR108 (Lyu et al. 2022), L-lactate (Mosienko et al. 2018), and CTHRC1 for GPR180 (Balazova et al. 2021) – it is presently unclear whether these receptors are competent for G protein signaling

since no molecular evidence has been reported for any member of this family (Balazova et al. 2021; Hoel et al. 2022).

425 Hi-GOLD

Seventeen choanoflagellate GPCRs clustered independently from other known GOST GPCRs but exhibit a similar topology with a conserved GOLD domain directly N-terminal to their 7TM (Fig. 1B and Fig. S12). We named this group of GPCRs “Hidden GOLD” (Hi-GOLD). Hi-GOLD GPCRs are conserved in both craspedids and acanthoecids (Fig. 2). Additionally, they have
430 homologs in metazoans, including sponges, ctenophores, placozoans, cnidarians, holothuroideans, and cephalochordates, but are absent from vertebrates (Supplementary File 23). Hi-GOLD GPCRs were also detected in other CRMs (filastereans and ichthyosporeans) and non-holozoans, such as amoebozoans. Thus, we have uncovered a new group of GPCRs bearing structural similarity with GOST GPCRs and having ancient origins in eukaryotes.

435 GPCR PIPK

Twenty-seven choanoflagellate GPCRs were identified as GPCR PIPKs based on their domain composition. These receptors comprise an N-terminal 7TM combined with a C-terminal phosphatidylinositol phosphate kinase (PIPK) catalytic domain (Bakthavatsalam et al. 2006; Van Den Hoogen et al. 2018; van den Hoogen and Govers 2018) (Fig. S13A and S13B). A
440 diagnostic “LR$(x)_9$GI” motif, previously found to be present in the linker region between the 7TM domain and the PIPK domain of these receptors (Van Den Hoogen et al. 2018) was also found in the choanoflagellate GPCR PIPKs (Fig. S13C).

Initially discovered in oomycetes, where the family greatly expanded, GPCR PIPKs have now been detected in diverse eukaryotes and sponges (Bakthavatsalam et al. 2006; Van Den
445 Hoogen et al. 2018). Sponges appear to be the only metazoans encoding GPCR PIPKs (Fig. 2 and (Van Den Hoogen et al. 2018)). Thus, genes encoding GPCR PIPKs were likely present in the last common eukaryotic ancestor and secondarily lost in diverse eukaryotes, including a loss after the divergence of sponges from the rest of metazoans.

A previous study found that one GPCR PIPK each was encoded in *S. rosetta* and *M. brevicollis*
450 species (Van Den Hoogen et al. 2018); our analysis revealed that GPCR PIPKs are also expressed by most craspedid choanoflagellates (Fig. 2). However, no acanthoecid choanoflagellates in our study appeared to possess these receptors (Fig. 2).

While GPCR PIPKs have previously been grouped into the same family because of their domain architecture – the 7TM domain combined with a PIPK domain – our analysis suggests that this
455 family is also supported by the sequence homology of their 7TM domain. Indeed, both choanoflagellate and sponge GPCR PIPKs analyzed in this study clustered together, forming a class distinct from other GPCR families (Fig. 1B). In addition, blasting the 7TM domains of choanoflagellate GPCR PIPKs against the full EUKPROT database (Richter et al. 2022) recovered GPCR PIPKS from various other eukaryotes (Filasterea, Discoba, Collodictyonida,
460 Telonemida, Rhodelphida, Ancyromonadida, Centroplasthelida) suggesting that the 7TM

domain is enough to find homologs, independently of the C-terminal PIPK domain. Thus, GPCR PIPKs likely constitute a family of GPCRs on their own, with a signature 7TM domain that might have evolved in stem eukaryotes.

Experimental data regarding the function of GPCR PIPKs is limited. Pioneer studies in oomycetes implicated the receptor in the regulation of sexual reproduction and virulence (Hua et al. 2013). In contrast, studies of *D. discoideum* GPCR PIPKs suggested a role in phagocytosis, cell density sensing, and bacterial defense (Bakthavatsalam et al. 2007; Riyahi et al. 2011). Although the catalytic domain of GPCR PIPKs is predicted to have a role in phospholipid signaling, experimental evidence for this hypothesis is still missing. Moreover, it is not known whether G protein signaling is involved downstream of GPCR-PIPKs (van den Hoogen and Govers 2018). Similarly, no ligands for GPCR PIKs have been identified.

GPRch3

The GPRch3 cluster contained 28 choanoflagellate GPCRs with short and unstructured N-terminal and C-terminal regions (Fig. S14A); they did not show any similarity to any other previously described GPCR families, either by sequence or structure (Fig. 1B). Nonetheless, these GPCRs appear to have homologs in diverse metazoans (sponges, ctenophores, cnidarians, protostomes and deuterostomes), corallochytreans, and non-holozoans (e.g. amoebozoans and alveolates; Fig. 2). Thus, the GPRch3 GPCRs likely represent a new family of GPCRs with ancient origin in eukaryotes that is retained in metazoans.

Holozoan GPCRs

GPR155

GPR155, also known as lysosomal cholesterol sensing (LYCHOS) protein, is an atypical GPCR that binds to lysosomal cholesterol and regulates mTORC1 activity (Shin et al. 2022; Bayly-Jones et al. 2024; Schöneberg 2024). It contains an unusual 17-transmembrane helix structure composed of an N-terminal transporter-like domain (10 transmembrane helices) fused to a 7TM, with Dishevelled, EGL-10 and Pleckstrin (DEP) domains located at the C terminus. The ability of GPR155 to transduce signals via G proteins is presently unclear. While it has been suggested that the 7TM of GPR155 shares similarities with the aGPCR family (Bayly-Jones et al. 2024), it appears to form a cluster of its own in our analysis (Fig. 1B).

GPR155 has been detected in diverse bilaterians (Vassilatis et al. 2003; Umeda et al. 2017; Wang et al. 2018), and its transporter domain has an ancient origin in eukaryotes (Dabravolski and Isayenkov 2022; Bayly-Jones et al., 2024). The finding that GPR155 homologs are encoded in CRMs, but are absent from non-holozoan lineages, suggests that GPR155 originated in the ancestor of holozoans (Fig. 2). Moreover, the conservation of GPR155 in CRMs extends beyond the 7TM as both the N-terminal transporter-like domain and the C-terminal DEP domain are detected in GPR155 homologs from CRMs (Fig. S15).

While choanoflagellates and close relatives can produce a variety of sterols (Kodner et al. 2008; Gold et al. 2016; Najle et al. 2016), they are not known to synthesize cholesterol, a vital lipid for metazoans (Zhang et al. 2019; Shamsuzzama et al., 2020). The ligand(s) of GPR155 in CRMs and whether it mediates sterol-induced signaling await further investigation.

GPR143

GPR143 homologs were detected in diverse metazoans, including sponges (Krishnan et al. 2014), and the filasterean *Capsaspora owczarzaki* (Fig. 2 and (De Mendoza et al. 2014)), but not in choanoflagellates. Therefore, GPR143 likely originated in holozoans and was lost in stem choanoflagellates (Fig. 2).

While little is known about GPR143, the receptor is expressed in pigment-producing cells in the skin and eyes of metazoans, where it controls melanosome biogenesis, organization, and transport (Bueschbell et al. 2022). The downstream signaling pathway is mostly uncharacterized, but GPR143 associates with several Gα and Gβ subunits, along with β-Arrestin (Schiaffino et al. 1999; Schiaffino and Tacchetti 2005; De Filippo et al. 2017). L-3,4-dihydroxyphenylalanine (L-DOPA) and dopamine have been suggested as ligands for GPR143 (Lopez et al. 2008; Goshima et al. 2019).

Stem eukaryote GPCRs that were lost from metazoans

Rémi-sans-famille (RSF)

28 choanoflagellate receptors clustered together and did not appear to share sequence similarities with other GPCR families previously described in eukaryotes. These GPCR candidates contained a 7TM domain but no additional domains (Fig. S16A). Their predicted structures revealed an additional short helix between TM6 and TM7 with a “NxLQxxMNxL” conserved motif (Fig. S16A and S16B). Structural similarity search suggested a weakly supported connection to the THH1/TOM1/TOM3 GPCR family (Yamanaka et al. 2000; Lu et al. 2018) (Fig. S16C).

While these GPCRs are absent from metazoans, we detected homologs in filastereans, as well as in other non-holozoans, including provorans (Tikhonenkov et al. 2022), rhodophytes, alveolates, amoebozoans, and ancyromonads (Fig. 2). Thus, it is likely that this family of GPCRs evolved early during the evolution of eukaryotes and was secondarily lost in various lineages, including metazoans. We named this new class of GPCRs “Rémi-sans-famille” (RSF), inspired by a fictional orphan (Malot 1878).

GPRch1

Clustering on the 7TM domains revealed a group of 16 craspedid choanoflagellate GPCRs, which we have named GPRch1 GPCRs. Although these proteins lack a PIPK domain, their 7TM domains resembled those of GPCR PIPKs from choanoflagellates and sponges (Figs. 1B and S14B). The GPRch1 7TMs also clustered with 7TMs from amoebozoans, stramenopiles,

535 alveolate, and apusomonadids; interestingly, the non-choanoflagellate GPCRs in this cluster all contained a PIPK domain in the C-terminus. Therefore, we infer that there was a duplication and divergence of GPCR PIPKs in stem eukaryotes, with one of the paralogs giving rise to GPRch1 GPCRs. Subsequently, the choanoflagellate GPRch1 GPCRs lost the PIPK domain. The absence of GPRch1 GPCRs from acanthoecid choanoflagellates, other CRMs, and metazoans
540 suggests that they were lost from these lineages (Fig. 2).

Choanoflagellate-specific GPCRs

GPCR-TKL/K and GPCR-TKs

The GPCR-TKL/K family was identified by examination of a cluster of 176 acanthoecid choanoflagellate GPCRs, 92 of which contain a kinase domain in their C-terminus (Fig. 1 and
545 Fig. 2; Fig. S17A, S17B, and S17C). Further analysis showed that the kinase domains of these GPCRs relate to Tyrosine Kinase-like (TKL) domains and other non-tyrosine kinase domains (Fig. S17B and Supplementary files 24, 25, and 26); we therefore named these receptors GPCR-Tyrosine Kinase-Like/Kinase (GPCR-TKL/K).

The fusion between a 7TM and a kinase domain has recently been reported in oomycetes and
550 amoebozoans but was not found in other eukaryotes (Judelson and Ah-Fong 2010; Van Den Hoogen et al. 2018). Thus, the acanthoecid GPCR-TKL/K represents the first observation, to our knowledge, of a GPCR-kinase fusion in holozoans. While the overall protein domain architecture of GPCR-TK/Ks is conserved between choanoflagellates, oomycetes, and amoebozoans, the 7TM domains of acanthoecid GPCR-TKL/K receptors did not cluster with any
555 other GPCRs in eukaryotes (Fig. 2), suggesting that this family probably originated through convergent evolution in stem acanthoecids (Fig. 2 and Fig. S17D).

Additionally, we detected 11 choanoflagellate GPCRs that contain a predicted tyrosine kinase domain in their C-terminal region, but that did not cluster with the GPCR-TKL/Ks mentioned above (Fig. S17B and Fig. S17C; Supplementary files 24, 25, 26, 27, and 28). These GPCRs
560 are found in both acanthoecids and craspedids but failed to form a unified family based on their 7TMs (Fig. S17C and S17D). Unlike their 7TM domains, the kinase domains of these GPCRs clustered together and separately from the GPCR TKL/K kinase domains; all recovered strong blast hits among curated metazoan tyrosine kinases (Fig. S17C and Supplementary file 26). Therefore, we identified these GPCRs as GPCR Tyrosine Kinases (GPCR-TKs). While no
565 metazoan homologs were found when using the 7TM domain of choanoflagellate GPCR-TKs as queries, blasting their conserved tyrosine kinase domains recovered GPCR-TKs in sponges but not in other metazoan lineages or other holozoans (Fig. S17E).

Together, our data suggests that the fusion of a C-terminal kinase domain to a 7TM domain likely evolved repeatedly and independently among eukaryotes and within the choanoflagellate
570 phylogeny. The evolution of GPCR-TKs seems to be restricted to choanoflagellates and early-branching metazoans. The evolutionary and molecular implications underlying the fusion of a 7TM domain with an intracellular kinase domain are unknown, and, to our knowledge, no GPCR-TK, GPCR-TKL, or GPCR-Ks have been functionally characterized.

Additional GPCRs

575 #### GPRch2

A cluster of twelve choanoflagellate GPCRs showed no similarity to previously identified GPCR families (Fig. 1). Although these GPCRs do not possess additional protein domains outside their signature 7TM domain (Fig. S18A), the GPRch2 GPCRs exhibit an unusual large intracellular loop 3 (ICL3) linking the TM5 to the TM6 (Fig. S18A and S18B).

580 While homologs of these GPCRs were also found in filastereans, ichthyosporeans, and corallochytreans, they are absent from metazoans and were only detected in green algae outside of holozoans (Fig. 2). Thus, it is likely that the GPRch2s evolved in stem holozoans either *de novo*, or via a horizontal transfer from green algae, and were subsequently lost in stem metazoans.

585 #### Frizzled

The Frizzled family, which encompasses Frizzled GPCRs and Smoothened, controls cell proliferation, cell fate, tissue polarity, and cell polarity during metazoan development (Taipale and Beachy 2001; Logan and Nusse 2004; Schulte 2024). These receptors have a highly conserved cysteine-rich domain called the Frizzled (Fz) domain, located in their extracellular N-
590 terminus (Schiöth and Lagerström 2008a; Zheng and Sheng 2024). This domain is connected to the seven-transmembrane (7TM) region by a linker segment. The Fz domain serves as the ligand-binding site for most Frizzled receptors. Several ligands have been found to activate members of the Frizzled family, of which the secreted glycoprotein Wingless/Int-1 (Wnt) has been the most studied (Schulte 2024).

595 Although initially characterized in *Drosophila*, Frizzled receptors have been detected in most metazoans, including sponges (Fig. 2 ; (Krishnan et al. 2014; Holzem et al. 2024)). While Frizzled receptors have not been previously detected in CRMs, homologs were found in fungi and amoebozoans, suggesting a probable origin of the family in the last common ancestor of opisthokonts and amoebozoans (Krishnan et al. 2012; De Mendoza et al. 2014). Notably, fungal
600 and amoebozoan Frizzled homologs also possess a Fz domain in their N-termini, supporting an overall conserved receptor topology.

While we could not detect members of the Frizzled family in any of the 23 choanoflagellate species used in our study (Fig. 2), we found four Frizzled/Smoothened receptors in the transcriptome of the corallochytrean *Syssomonas multiformis* (Fig. 2; Fig. S19A and S19B;
605 Supplementary File 29). However, contaminating sequences have been reported in this data set, and therefore, this observation should be treated with caution (Hehenberger et al. 2017).

DISCUSSION

Our transcriptomic and genomic survey demonstrates that choanoflagellate GPCRomes are richer than previously understood, both in the number and diversity of GPCRs represented. We

610 identified 18 distinct GPCR families in choanoflagellates, of which 12 – Rhodopsin, GPR155, GPR157, GPCR TKL/K, TMEM145, TMEM87, GPR180, Hi-GOLD, RSF, GPRch1, GPRch2, and GPRch3 – were newly described in these organisms (Table 1, Fig. 2). Among these families, we observed that aGPCRs generally constitute the largest class of GPCRs in choanoflagellates, apart from the GPCR TKL/K family in acanthoecids. Most choanoflagellate
615 GPCR families are conserved in metazoans or other eukaryotes. This supports the view that GPCRs are ancient gene families found across eukaryotic diversity (Krishnan et al. 2012; De Mendoza et al. 2014; Mojib and Kubanek 2020).

In addition, by assessing the conservation of choanoflagellate GPCRs in other eukaryotes, we uncovered GPCR families in metazoans that, to our knowledge, had not been reported in this
620 clade before. These include Hi-GOLD, GPRch3, and GPCR TKL/Ks. In addition, we found that five GPCR families – RSF, Hi-GOLD, GPRch3, GPR157, and GPRch1 – are conserved across diverse eukaryotes, thereby expanding the repertoire of GPCRs inferred at the root of the eukaryotic tree (De Mendoza et al. 2014). Future increases in the sequencing of non-metazoans will likely expand further the inferred diversity of GPCRs present in the progenitors of
625 metazoans and other eukaryotes.

The identification of Glutamate Receptors, Rhodopsins, and aGPCRs in choanoflagellates, three GPCR families with relatively well-characterized subfamilies in metazoans (Fredriksson et al. 2003; Pin et al. 2003; Bjarnadóttir et al. 2004; Chun et al. 2012; Scholz et al. 2019; Ellaithy et al. 2020b; Wittlake et al. 2021), led us to investigate whether these subfamilies are conserved in
630 choanoflagellates and close relatives. Based on analyses of the 7TM domains, we found relatives of the main subfamilies of metazoan Glutamate Receptors in CRMs (Fig. 3B). In contrast, most aGPCR subfamilies reported in metazoans have no obvious orthologs in CRMs, suggesting that metazoan and CRM aGPCRs diversified independently (Fig. 5). One exception was the ADGRV family (Weston et al. 2004; Hamann et al. 2015a; Kusuluri et al. 2021), for
635 which clear orthologs are found in choanoflagellates and filastereans (this study and (Krishnan et al. 2012; Peña et al. 2016)).

Whereas Rhodopsins are the most abundant GPCRs in metazoans, we identified only three metazoan-like Rhodopsins in CRMs: two in choanoflagellates and one in ichthyosporeans. However, Rhodopsins have also been identified in holomycotans and other non-CRM
640 eukaryotes, suggesting they may have evolved in stem eukaryotes (Krishnan et al. 2012). Sequence comparisons and structural predictions suggest that the two choanoflagellate Rhodopsins most closely resemble metazoan opsins, a Rhodopsin subfamily (Fredriksson et al. 2003; Mickael et al. 2016). Opsins, light-detecting receptors supporting vision, have only been described in metazoans (Feuda et al. 2012; Fleming et al. 2020; Wong et al. 2022; Hagen et al.
645 2023; Aleotti et al. 2025). Unlike metazoan opsins, the two choanoflagellate Rhodopsins lack canonical residues essential for light-sensing, including the highly conserved Lysine K296, which mediates binding to the chromophore retinal (Davies et al. 2010; Hankins et al. 2014; Nagata and Inoue 2021). While it is formally possible that Rhodopsins existed in stem choanoflagellates and were lost in a majority of modern choanoflagellate lineages, either
650 horizontal gene transfer or convergent evolution in the shared ancestor of *S. macrocollata* and *S. punica* are similarly plausible explanations for their presence in these species. Therefore,

ambiguities about the provenance and function of CRM Rhodopsins obscure the ancestry of metazoan Rhodopsins and opsins.

The N-terminal and C-terminal protein domains of choanoflagellate, other CRM, metazoan, and additional eukaryotic GPCRs revealed uneven evolutionary conservation between GPCR families. For example, we found that the C-terminal PIPK domain of the GPCR PIPK family, the N-terminal transporter domain and C-terminal DEP domain of GPR155, and the N-terminal GOLD domain of GOST GPCRs are generally conserved in CRMs, metazoans, and other eukaryotes. In contrast, the stereotypical association of N-terminal protein domains from metazoan Glutamate Receptors is absent from non-metazoan Glutamate Receptors identified by 7TM clustering. Moreover, we observed that the N-termini of Glutamate Receptors from filastereans and holomycotans appear to possess a larger diversity of protein domains than metazoan Glutamate Receptors, suggesting functional constraint exerted on the architecture of metazoan Glutamate Receptors. This likely has implications for both the ligands that non-metazoan Glutamate Receptors may recognize and their underlying mechanism of activation (Hu et al. 2000; Jiang et al. 2004; Rondard et al. 2006; Levitz et al. 2016; Ellaithy et al. 2020b; Laboute et al. 2023).

The aGPCR family shows the greatest level of protein domain composition diversity in metazoans (Schiöth and Lagerström 2008b; Krishnan et al. 2012; Hamann et al. 2015a). Similarly, we uncovered a large diversity of N-terminal protein domain combinations in CRM aGPCRs, with a clear diversification of the repertoire in choanoflagellates and, to a lesser extent, in filastereans (Figs. 1B, 5, S4, and S7). The aGPCRs in CRMs displayed all the core structural hallmarks of metazoan aGPCRs, including a long N-terminus containing a GAIN domain and many additional ECDs. Moreover, we found that the HRM/GAIN/7TM module, previously thought to be restricted to metazoans, is also found in CRM aGPCRs. Our data support an evolutionary scenario wherein the linkage between the GAIN and 7TM domains evolved first in stem holozoans, followed by the incorporation of the HRM into the HRM/GAIN/7TM module in the progenitors of Filasterea and Choanozoa.

Despite its conservation, the function of the HRM-GAIN-7TM module is presently unclear (Shima et al. 2004; Kimura et al. 2006; Araç et al. 2012; Prömel et al. 2012b; Araç and Leon 2019). Interestingly, the HRM domain binds peptide hormones and neuropeptides in the Secretin GPCR family (Schiöth and Lagerström 2008a; Pal et al. 2012; Zhao et al. 2023). The presence of HRM-containing aGPCRs in CRMs could suggest a pre-metazoan origin of peptide-based endocrine signaling. This is reinforced by the recent finding that sponges and choanoflagellates express sequences that resemble metazoan neuropeptides and peptide hormones (Steffen et al. 2021; Yañez-Guerra et al. 2022; Zhenjian Lin, Vinayak Agarwal, Ying Cong 2024).

Finally, the diversity of protein domain architectures in the aGPCRs of choanoflagellates and metazoans suggests that aGPCR evolution was shaped by extensive domain shuffling (Gilbert 1978; Patthy 1999; Smithers et al. 2019; Patthy 2021). While some discrete protein domain associations (GAIN/7TM and HRM/GAIN/TM) or, in rare cases, most of the receptor sequence (ADGRV/subfamily VIII) survived, presumably under selection, new protein domain

architectures arose, perhaps through a combination of gene duplication followed by domain shuffling. The diversification of NTF domains might have facilitated the recognition of a diversity of ligands (e.g., proteoglycans and proteins).

Of course, the aGPCRs were named for the presence of adhesion-related protein domains in the NTFs of metazoan aGPCRs, some of which support cell-cell and cell-matrix adhesion (Langenhan et al. 2013; Hamann et al. 2015b). The relevance of adhesion domains in CRM aGPCRs is intriguing and requires further investigation. Interestingly, diverse CRMs form facultative multicellular structures (Fairclough et al. 2010; Marshall and Berbee 2011; Glockling et al. 2013; Sebé-Pedrós et al. 2013; Carr et al. 2017a; Brunet et al. 2019; Dudin et al. 2019; Tikhonenkov et al. 2020; Ros-Rocher et al. 2021). Therefore, the diversification of adhesion-related protein domains in the NTFs of aGPCRs may have contributed to the evolution of multicellularity in holozoans. In addition, a recent study in sponges suggested that aGPCRs could be part of the immune response following exposure to microbial-associated molecular patterns (Pita et al. 2018).

A better appreciation of the GPCRs expressed in choanoflagellates and other CRMs is a first step to understanding cue sensing in these organisms. While environmental stimuli are known to play an essential role in controlling life history transitions in choanoflagellates, the receptors involved have yet to be characterized (Dayel et al. 2011; Alegado et al. 2012; Levin and King 2013; Woznica et al. 2017; Ros-Rocher and Brunet 2023). We anticipate that the identification of GPCR repertoires in the sister group of metazoans will open the way to the functional characterization of this superfamily of receptors in choanoflagellates, which could illuminate the origin and ancestral function(s) of key signaling pathways in metazoans.

MATERIAL AND METHODS

Taxon sampling

720 The taxa surveyed in this study are listed in Table 2.

Identification of choanoflagellate GPCRs

GPCR mining through sequence homology-based approaches

To identify putative choanoflagellate GPCR sequences in the 23 choanoflagellate species (Table 2 and Supplementary Table 1), the choanoflagellate predicted proteomes were searched
725 with Hidden Markov models (HMMs) unique to each of the 54 GPCR families belonging to the GPCR_A Pfam clan (CL0192 ; Supplementary Table 2) (Mistry et al. 2021) using the HMMER3.3 software package (cutoff threshold value: 0.01) (Eddy 2011). We recovered 1070 candidate GPCRs. In a complementary approach, we used a global HMM profile (GPCRHMM, local score) that mimics the common topology of GPCRs (Wistrand et al. 2006) and recovered
730 1095 candidates in our analysis. The comparison of the two sets of GPCRs revealed 381 duplicate candidates, which were merged, leaving a final set of 1784 candidates.

Validation of candidate GPCRs through domain composition and topology filtering

To reduce false positives and remove highly truncated sequences and possible isoforms, we submitted the candidate GPCRs to a round of filtering that combined complementary
735 approaches. All the candidates were used as queries to perform BLASTP searches against both the NCBI and GPCRdb databases (Altschul et al. 1990; Pándy-Szekeres et al. 2023) using an e-value cut-off of 1x 10^{-5}. The sequences that recovered a GPCR within the 10 top hits were kept. In addition, GPCR candidates were evaluated on the basis of their protein domain composition by combining the search tools InterProScan, CDvist, and the transmembrane
740 domain predictor TMHMM 2.0 (Krogh et al. 2001a; Adebali et al. 2015; Blum et al. 2025). The prediction of a 7TM domain, along with additional signature protein domains, was assessed for each candidate. Protein sequences derived from genomes and transcriptomes were retained if containing 6 to 8 transmembrane helices to account for the detection of possible N-terminal signal peptides (Rutz et al. 2015) and TMHMM prediction errors. Sequences that possessed a
745 7TM domain but that did recover GPCRs when used as BLAST queries were subjected to a comparative structural analysis by predicting their structure using Alphafold 3 (Abramson et al. 2024) and searching for structural homologs within the Alphafold databases (van Kempen et al. 2024)see “Protein structure prediction and search” below for more details on the analysis). Candidates with a predicted barrel shape conformation of their 7 alpha helices and that
750 recovered GPCR structural hits (E-value < $1e^{-9}$) were considered GPCRs and were kept in our dataset. Finally, we removed putative splice isoforms using CD-HIT with a 90% identity threshold (Li and Godzik 2006). A total of 1113 false positives, isoforms, and highly truncated sequences were identified and removed from our dataset, leaving 671 validated choanoflagellate GPCRs in total.

755 ### Recovering additional choanoflagellate GPCRs using choanoflagellate GPCR BLAST queries and custom choanoflagellate GPCR HMMs

Because the HMMs used to search for choanoflagellate GPCRs in our analysis were primarily based on seed alignments of metazoan sequences, we looked for additional choanoflagellate GPCRs by using the 671 validated choanoflagellate GPCRs as queries. To this end, we
760 simplified the choanoflagellate GPCRs dataset by clustering the 671 choanoflagellate GPCRs based on all-against-all pairwise sequence similarities (see "Clustering of the 918 validated choanoflagellate GPCRs" below to find more details about the protocol). The 671 choanoflagellate GPCRs were sorted into 18 clusters, with only 76 GPCRs that did not group with other GPCRs.

765 Next, we built cluster-specific GPCR HMMs for each of the 18 choanoflagellate GPCR clusters previously identified. We also built individual HMMs for the 76 GPCRs that did not fall into the 18 clusters. To this end, the 7TM domains of all the members of each of the 18 GPCR clusters were aligned independently using MAFFT (Version 7.4) with the E-INS-i algorithm (Katoh et al. 2018). The resulting 18 multisequence alignments were then used to build 18 corresponding
770 choanoflagellate cluster-specific GPCR HMMs using the hmmbuild module of HMMER v3.3. In the case of GPCRs that did not belong to any clusters, HMMs were built based on single 7TM domains. These HMMs were used to search the choanoflagellate proteomes again. In parallel to this approach, we also selected 3 choanoflagellate GPCR sequences per GPCR cluster or a single GPCR sequence in the case of isolated GPCRs, to use as BLASTP queries against the
775 23 choanoflagellate proteomes (E value: $1.0e^{-5}$). The resulting GPCR candidates were validated through the filtering approach previously described, leaving 247 new GPCRs that were not predicted with the metazoan-biased HMMs. Added to the original 671 this yielded a total of 918 GPCRs from the 23 choanoflagellate species. The complete list of validated choanoflagellate sequences in FASTA format and the choanoflagellate-specific GPCR HMMs are available in
780 Supplementary Files 1 and 2.

Identification of GPCR signaling pathway components

Heterotrimeric G proteins (Gα, Gβ, and Gγ), positive regulators (Phosducin and Ric8), and negative regulators (Arrestin, GRK, RGS, and GoLoco) of G protein signaling were identified by searching choanoflagellate proteomes with Pfam HMM profiles using the HMMER3.3 software
785 package (Eddy 2011) (cutoff threshold value: $1e^{-5}$) with the hmmsearch function. The following Pfam HMMs were used: PF00503 (G protein alpha subunit), PF00400 (WD domain, G-beta repeat), PF00631 (GGL domain), PF02752 (Arrestin (or S-antigen), C-terminal domain), PF00339 (Arrestin (or S-antigen), N-terminal domain), PF02188 (GoLoco motif), PF02114 (Phosducin), PF00615 (Regulator of G protein signaling domain), and PF10165 (Ric8)
790 (Supplementary Table 2). The complete protein domain architecture of the candidates was assessed with CDvist (Adebali et al. 2015) and choanoflagellate sequences from each category were then used as BLAST queries against the 23 choanoflagellate proteomes on the EukProt database (E value: $1.0e^{-5}$) (Richter et al. 2022) to look for additional sequences that might not have been detected in the previous round of screening. All the sequences identified in this
795 analysis are listed in Supplementary File 3.

Identification of HRM-containing proteins

To identify HRM-containing proteins across eukaryotes, we searched the proteomes of metazoans, choanoflagellates, filastereans, ichthyosporeans, corallochytreans, holomycotans, amoebozoans, viridiplantae, and ciliates with the Pfam HMM profile HRM (PF02793) using the HMMER3.3 software package (cutoff threshold value: $1e^{-5}$) with the hmmsearch function (Eddy 2011). The complete protein domain architecture of the candidates was then further assessed using Cdvis (Adebali et al. 2015). The five HRM-containing proteins identified in the nucleariid *F.alba* are provided in Supplementary File 21.

Identification of GPCR families in other eukaryotes

To assess the presence of Glutamate, Rhodopsin, Adhesion, and Frizzled GPCR families in other eukaryotes, we searched the relevant genomes and transcriptomes (Table 2) with their signature 7TM HMMs (Supplementary Table 2) using the HMMER3.3 software package (cutoff threshold value: 0.01) with the hmmsearch function. The candidates were then validated through the previously described domain composition and topology filtering pipeline (see "Validation of candidate GPCRs through domain composition and topology filtering") (Supplementary File 7).

To assess the presence of other GPCR families, we searched the entire Eukprot v3 dataset (993 species) with previously published metazoan and/or non-metazoan GPCR family members as queries (threshold E value: $1.0e^{-5}$). Both the full-length GPCR and the extracted 7TM domains were used as search queries in this analysis. We defined hits as those with at least 70% query coverage and at least 30% sequence identity. When no hits were recovered, an additional search using their signature 7TM HMMs (Supplementary Table 2) was performed on the genomes/transcriptomes dataset (Table 2). All the candidates were then validated through a similar domain composition and topology filtering approach.

Clustering analyses

Clustering of the 918 validated choanoflagellate GPCRs

To sort the choanoflagellate GPCRs into families, we created a dataset composed of all the 7TM regions extracted from the 918 choanoflagellate sequences identified in our study, along with the 7TM sequences of various metazoan, amoebozoan, stramenopile, and chlorophyte GPCRs to aid with the identification of the clusters (Supplementary File 6).To isolate the 7TM region, we aligned all the GPCR sequences on Geneious Prime v2024.07 using MAFTT with E-INS-I algorithm (Katoh et al. 2018), and predicted their transmembrane helices with TMHMM. All the sequences starting from the N-terminal end of the transmembrane helix 1 and finishing at the C-terminal end of the transmembrane helix 7 were kept.

Next, due to the inherent difficulties of analyzing a large set of proteins with mild sequence conservation without decreasing the accuracy of the alignment used for phylogenetic inference,

we opted for an all-against-all pairwise similarity-based clustering approach to assess the diversity of GPCR families in choanoflagellates using the clustering tool CLANS 2.2.2 (Frickey and Lupas 2004a; Gabler et al. 2020). Unlike phylogenetic reconstruction, this approach becomes more accurate with an increasing number of sequences (Frickey and Lupas 2004b)(Frickey and Lupas 2004b). First, we used the CLANS web utility in the MPI Bioinformatic toolkit (https://toolkit.tuebingen.mpg.de/tools/clans) to perform an all-against-all BLAST search (with scoring matrix BLOSUM62) to obtain a matrix of pairwise sequence similarities with a cut-off value of $1e^{-6}$ for BLAST E-values (High Scoring Pairs with E-value higher than $1e^{-6}$ were not extracted). We then opened the resulting matrix in the CLANS graphical interface, and CLANS was allowed to cycle 20,000 times to optimize the graph. Clusters were automatically detected using the convex clusters search of four or more proteins with the attraction value limit set at 0.5 standard deviation. A total of 18 choanoflagellate GPCR clusters were recovered. The clustering maps were further edited in Adobe Illustrator 2024 to modify the colors and symbols used.

Clustering of holozoan aGPCR N-termini

The extracellular region of the metazoan, choanoflagellate, and filasterean aGPCRs was extracted by combining the transmembrane domain prediction tool TMHMM-2.0 (Krogh et al. 2001b)(Krogh et al. 2001b) with the alignment tool MAFFT (E-INS-I algorithm) on Geneious Prime. All the regions starting at the N-terminal end of the GPCRs and finishing at the start of the first transmembrane helix were retained for downstream analysis (Supplementary File 19). The clustering analysis performed was comparable to the analysis described previously for GPCRs, with a $1e^{-20}$ cut-off value being used for the matrix pairwise sequence similarities.

Clustering of Rhodopsins

To compare the newly identified choanoflagellate Rhodopsins with the Rhodopsins found in a diversity of metazoans and other CRMs, we searched the proteomes of metazoans and CRMs (Table 2) with the Pfam HMM profile 7TM_1/Rhodopsin (PF00001) using the HMMER3.3 software package (cutoff threshold value: 1e-5) with the hmmsearch function. Candidates were then validated by assessing their InterProScan protein domain signature and BLASTing the sequences against the NCBI database. Additionally, we used the two choanoflagellate Rhodopsins – *S.macrocollata*_m.143379 and *Salpingoeca_punica*_m.44256 – to query the EukProt database (993 species), and the top blast hits with E values < $1x10^{-10}$ were included in our dataset. A total of 6149 validated Rhodopsins were then submitted for all-against-all pairwise comparison with a $1e^{-12}$ cut-off value being used for the matrix pairwise sequence similarities. All the validated Rhodopsins used in the analysis and the output of the analysis are found in Supplementary Files 11 and 12.

Clustering of GPCR Kinase and Kinase domains

To assess the diversity of choanoflagellate GPCR Kinase along with the diversity of their kinase domains, we extracted their 7TM domain and their kinase domain separately. To this end, we aligned all the choanoflagellate GPCR Kinase sequences on Geneious Prime using the alignment tool MAFFT (E-INS-I algorithm). The 7TM domains and the kinase domains were then predicted and extracted by running locally TMHMM-2.0 and Interproscan (PfamA) (Supplementary Files 24, 25, 27 and 28). The clustering analyses were similar to the ones described previously, with cut-off values of $1e^{-6}$ and $1e^{-20}$ for the 7TM and kinase domains, respectively. In parallel, each kinase domain was also blasted against KinBase (Manning et al. 2002; Bradham et al. 2006; Goldberg et al. 2006)), the curated protein kinase dataset from www.kinase.com, for identification (Supplementary File 26).

Sequence alignment and phylogenetic analyses

Phylogenetic analyses of holozoan aGPCRs, Glutamate Receptors, and Gα subunits were performed in this study. To construct the holozoan aGPCR and Glutamate Receptor phylogenies, we first extracted the aGPCR and Glutamate Receptor 7TM domains by combining the transmembrane domain prediction tool TMHMM-2.0 with the alignment tool MAFFT (E-INS-I algorithm) on Geneious Prime v2024.07; only the protein sequences spanning the 7TM region were retained for downstream analysis (Supplementary Files 8, 15 and 17). To reduce the size of the large GPCR datasets while preserving their diversity, we systematically clustered similar sequences using CD-HIT (threshold 0.8; word size = 5). A total of 257 Glutamate Receptor 7TM sequences and 659 aGPCR 7TM sequences were kept for downstream analyses. We then aligned the aGPCR and Glutamate Receptor 7TMs or the full-length Gα sequences using MAFFT with the E-INS-I algorithm. The resulting alignments were then used for Maximum-likelihood and/or Bayesian inference of phylogenies (Fig. 3B, Fig. 5A, Fig. S3D, and Fig. S6A; Supplementary Files 5, 9, 16 and 18). We built Maximum-likelihood phylogenies with IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/; (Trifinopoulos et al. 2016)) using ModelFinder (Kalyaanamoorthy et al. 2017) and 2000 Ultrafast Bootstraps (UF-boot) (Minh et al. 2013) or 2000 iterations of SH-aLRT. Bayesian inference was performed using MrBayes v3.2.7a (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012) on CIPRES Science Gateway (Miller et al. 2010). The posterior probability of the trees was estimated using Markov Chain Monte Carlo (MCMC) analysis and a fixed LG amino acid substitution model (Aamodelpr = Fixed(LG)). The gamma-shaped model was used to estimate the variation of evolutionary rates across sites (set rates = Gamma). MCMC analysis was set to run for 6,000,000 generations and every 120th tree was sampled. Diagnostics were calculated for every 1000 generations (diagnfreq = 1000) to analyze the convergence of the two independent runs starting from different random trees. To terminate the MCMC generations, a stop rule was applied, and the convergence was analyzed until the average standard deviation of split frequencies dropped below 0.01. To ensure that the parameter estimates were only made from data drawn from distributions derived after the MCMCs had converged, we discarded the first 25% of the sampled trees in the burnin phase using relative burnin setting (relburnin = yes and burninfrac = 0.25). A consensus tree was built from the remaining 75% of the sampled trees with the MrBayes sumt command using the 50% majority rule method. Trees were visualized using iTOL (https://itol.embl.de/) (Letunic and Bork 2024) and further edited in Adobe Illustrator 2024.

Protein Domain Search

To identify and compare the protein domains found in the extracellular region of aGPCRs and Glutamate Receptors across a range of eukaryotes, we first assessed the aGPCR and Glutamate Receptor repertoires from various metazoans, filastereans, ichthyosporeans, corallochytreans, holomycotans, amoebozoans, viridiplantae, and ciliates (see Table 2). Briefly, the proteomes were searched for aGPCR and Glutamate Receptors with the Pfam HMM profiles – 7TM_2/Adhesion (PF00002) and 7TM_3/Glutamate (PF00003) – respectively, using the HMMER3.3 software package (cutoff threshold value: $1e^{-5}$) with the hmmsearch function. Candidates were further verified by assessing their InterProScan protein domain signature and BLASTing the sequences against the NCBI database. The validated aGPCR and Glutamate Receptors from each clade (choanoflagellates, metazoans, filastereans, ichthyosporeans, corallochytreans, holomycotans, amoebozoas, viridiplantae, and ciliates), were then analyzed in batch using CDvist with default parameters to assess their protein domain composition (Fig. 3C and Fig. S7; Supplementary Files 10 and 19).

Protein structure prediction and search

The protein structures of full-length GPCRs and/or extracted 7TM domains were predicted using the AlphaFold 3 server with default parameters (https://alphafoldserver.com/; (Abramson et al. 2024)). Five models were generated per input, of which only the top-ranked prediction (based on the ranking_score metric) was selected for downstream analyses. The quality of the predicted structure was then assessed based on the pLDDT confidence score. Only models with pLDDT scores above 70 were retained. Structures were analyzed and figures were prepared with ChimeraX (Meng et al. 2023).

To find structural homologs, we searched the Alphafold-predicted structures against the AFDB-PROTEOME, AFDB-SWISSPROT, and AFDB50 databases using the search program Foldseek (https://search.foldseek.com/search ; (van Kempen et al. 2024)) with default parameters. Structural top hits with a pLDDT confidence score above 70 in the aligned region and an associated E-value < $1e^{-9}$ were considered putative structural homologs.

Logo analysis

To build GAIN domain sequence logos (Fig. S9B), we isolated the C-terminal region of the GAIN domain from 301 choanoflagellate, 99 filasterean, 1 corallochytrean, and 30 murine aGPCRs (Supplementary file 20). To do so, we aligned separately the aGPCR sequences from the four datasets using MUSCLE v5 (Edgar 2022) on Geneious Prime, and the region of the C-terminal GAIN domain starting with the first N-terminal conserved Cysteine and finishing with the TA consensus “TxFAVLM” was extracted and kept for downstream analysis. We then trimmed the aligned sequences with ClipKIT (Steenwyk et al. 2020) using the GAPPY mode (with the default value:0.9). The resulting processed alignments were analyzed with WebLogo3 (https://weblogo.threeplusone.com/) (Crooks et al. 2004).

DECLARATION OF INTERESTS

The authors declare no competing interests.

ACKNOWLEDGEMENTS

We are grateful to Demet Araç, Thibaut Brunet, Daniel Richter, and Iñaki Ruiz-Trillo for feedback on the manuscript. We thank the whole King lab for valuable feedback throughout the project and Flora Rutaganira for advice about categorizing kinases. This work was supported by a Human Frontier Science Program long-term fellowship (LT 000919/2020-L) to A.G.D.L.B and an Investigator Award from the Howard Hughes Medical Institute (N.K.).

DATA AVAILABILITY

Supplementary Table 1: Choanoflagellate genome and transcriptome accession numbers.
Supplementary Table 2: List of identifiers of all Pfam HMMs used in this study.

The Supplementary Files below are all available on Figshare:
https://doi.org/10.6084/m9.figshare.28801289.v1

Supplementary File 1: 918 choanoflagellate GPCR sequences identified in this study.
Supplementary File 2: Choanoflagellate cluster-specific GPCR HMMs.
Supplementary File 3: Choanoflagellate heterotrimeric G proteins, positive and negative regulators of G protein signaling identified in this study.
Supplementary File 4: List of all Gα sequences used for phylogenetic reconstruction in Figure S3D.
Supplementary File 5: Maximum-likelihood phylogeny of choanoflagellate and metazoan Gα sequences. Related to Figure S3D.
Supplementary File 6: Complete set of 7TM sequences used for the clustering analysis related to Figure 1B.
Supplementary File 7: Glutamate, Rhodopsin, Adhesion, and Frizzled receptors identified in opisthokonts and other eukaryotes; related to Figure 2.
Supplementary File 8: Glutamate Receptor 7TM sequences used in the phylogenetic analysis related to Figure 3B.
Supplementary File 9: Fully annotated versions of the Glutamate Receptor phylogeny depicted in Figure 3B.
Supplementary File 10: List of opisthokont Glutamate Receptor sequences used for the analysis related to Figure 3C.
Supplementary File 11: Set of 6152 Rhodopsin sequences used in the clustering analysis related to Figure 4B.
Supplementary File 12: Raw cluster analysis CLANS file; related to Figure 4B.
Supplementary File 13: Seed alignment of the Rhodopsin sequences related to Figure 4E.

Supplementary File 14: Seed alignments downloaded from GPCRdb used to build human consensus sequences for the Aminergic, Peptide, and Lipid Rhodopsin subfamilies. Related to Figure 4E.
Supplementary File 15: List of all the aGPCR 7TM sequences used to build the two phylogenies depicted in Figure S6A and B.
Supplementary File 16: Fully annotated versions of the two phylogenies depicted in Figure S6A and B.
Supplementary File 17: List of all the aGPCR 7TM sequences used to build the phylogeny depicted in Figure 5A.
Supplementary File 18: Fully annotated versions of the aGPCR phylogeny depicted in Figure 5A.
Supplementary File 19: List of all the aGPCR NTF sequences used in Figure S7 and in the clustering analysis depicted in Figure S8.
Supplementary File 20: Extracted C-term GAIN sequences and seed alignments used to create the logos related to Figure S9B.
Supplementary File 21: Sequences of the five HRM-containing proteins identified in *Fonticula alba*; related to Figure S10D.
Supplementary File 22: GPR157 homologs identified in diverse eukaryotes; related to Figure S11A.
Supplementary File 23: Hi-GOLD homologs identified in diverse eukaryotes.
Supplementary File 24: Extracted kinase domain sequences from choanoflagellate GPCR Kinase used for the clustering analysis in Figure S17B.
Supplementary File 25: Raw cluster analysis CLANS file; related to Figure S17B.
Supplementary File 26: Top Blast hits of the choanoflagellate kinase domains analyzed in Fig. S17B searched against KinBase.
Supplementary File 27: Extracted 7TM domain sequences from choanoflagellate GPCR TKs, TKLs, and Ks used in the clustering analysis in Figure S17C.
Supplementary File 28: Raw cluster analysis CLANS file; related to Figure S17C.
Supplementary File 29: Frizzled/Smoothened sequences identified in the corallochytrean *S. multiformis*; related to Figure S19.

REFERENCES

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500.

Abu Obaid A, Ivandic I, Korsching SI. 2024. Deciphering the function of the fifth class of Gα proteins: regulation of ionic homeostasis as unifying hypothesis. *Cellular and Molecular Life Sciences* 81.

Adebali O, Ortega DR, Zhulin IB. 2015. CDvist: A webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* 31:1475–1477.

Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, Clardy J, King N. 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *Elife*:1–16.

Aleotti A, Giorgini F, Feuda R. 2025. A Pre-Bilaterian Origin of Phototransduction Genes and Photoreceptor Cells. *bioRxiv preprint*:1–30.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.

Anantharaman V, Aravind L. 2002. The GOLD domain, a novel protein module involved in Golgi function and secretion. *Genome Biol* 3:1–7.

Anjard C, Loomis WF. 2006. GABA induces terminal differentiation of Dictyostelium through a GABAB receptor. *Development* 133:2253–2261.

Ansel M, Ramachandran K, Dey G, Brunet T. 2024. Origin and evolution of microvilli. *Biol Cell*:1–23.

Araç D, Boucard AA, Bolliger MF, Nguyen J, Soltis SM, Südhof TC, Brunger AT. 2012. A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO Journal* 31:1364–1378.

Araç D, Leon K. 2019. Structure, function and therapeutic potential of adhesion GPCRs. *GPCRs: Structure, Function, and Drug Discovery*:23–41.

Araç D, Sträter N SE. 2016. Understanding the Structural Basis of Adhesion GPCR Functions. *Handb Exp Pharmacol.* [Internet]:221–247. Available from: https://link.springer.com/content/pdf/10.1007%2F978-3-319-41523-9.pdf

Attwood TK, Findlay JBC. 1993. Design of a discriminating fingerprint for g-protein-coupled receptors. *Protein Engineering, Design and Selection* 6:167–176.

Bagnato A, Rosanò L. 2019. New routes in GPCR/β-arrestin-driven signaling in cancer progression and metastasis. *Front Pharmacol* 10.

Bakthavatsalam D, Brazill D, Gomer RH, Eichinger L, Rivero F, Noegel AA. 2007. A G Protein-Coupled Receptor with a Lipid Kinase Domain Is Involved in Cell-Density Sensing. *Current Biology* 17:892–897.

Bakthavatsalam D, Meijer HJG, Noegel AA, Govers F. 2006. Novel phosphatidylinositol phosphate kinases with a G-protein coupled receptor signature are shared by Dictyostelium and Phytophthora. *Trends Microbiol* 14:378–382.

Balazova L, Balaz M, Horvath C, Horváth Á, Moser C, Kovanicova Z, Ghosh A, Ghoshdastider U, Efthymiou V, Kiehlmann E, et al. 2021. GPR180 is a component of TGFβ signalling that promotes thermogenic adipocyte function and mediates the metabolic effects of the adipocyte-secreted factor CTHRC1. *Nat Commun* 12:1–18.

Barros-Álvarez X, Nwokonko RM, Vizurraga A, Matzov D, He F, Papasergi-Scott MM, Robertson MJ, Panova O, Yardeni EH, Seven AB, et al. 2022. The tethered peptide activation mechanism of adhesion GPCRs. *Nature* 604:757–762.

Bayly-Jones C, Lupton CJ, Keen AC, Dong S, Mastos C, Luo W, Qian C, Jones GD, Venugopal H, Chang YG, et al. 2024. LYCHOS is a human hybrid of a plant-like PIN transporter and a GPCR. *Nature* [Internet] 634. Available from: http://dx.doi.org/10.1038/s41586-024-08012-9

Bjarnadóttir TK, Fredriksson R, Höglund PJ, Gloriam DE, Lagerström MC, Schiöth HB. 2004. The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics* 84:23–33.

Bjarnadóttir TK, Fredriksson R, Schiöth HB. 2007. The Adhesion GPCRs: A unique family of G protein-coupled receptors with important roles in both central and peripheral tissues. *Cellular and Molecular Life Sciences* 64:2104–2119.

Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, Pinto BL, Orr A, Paysan-Lafosse T, Ponamareva I, et al. 2025. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res* [Internet] 53:D444–D456. Available from: http://www.ncbi.nlm.nih.gov/pubmed/39565202%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC11701551

Bradham CA, Foltz KR, Beane WS, Arnone MI, Rizzo F, Coffman JA, Mushegian A, Goel M, Morales J, Geneviere AM, et al. 2006. The sea urchin kinome: A first look. *Dev Biol* 300:180–193.

Brunet T, Larson BT, Linden TA, Vermeij MJA, McDonald K, King N. 2019. Light-regulated collective contractility in a multicellular choanoflagellate. *Science* 366:326–334.

Bueschbell B, Manga P, Schiedel AC. 2022. The Many Faces of G Protein-Coupled Receptor 143, an Atypical Intracellular Receptor. *Front Mol Biosci* 9:1–16.

Cardoso JCR, Félix RC, Fonseca VG, Power DM. 2012. Feeding and the rhodopsin family G-protein coupled receptors in nematodes and arthropods. *Front Endocrinol (Lausanne)* 3:1–22.

Cardoso JCR, Mc Shane JC, Li Z, Peng M, Power DM. 2024. Revisiting the evolution of Family B1 GPCRs and ligands: Insights from mollusca. *Mol Cell Endocrinol* 586.

Carr M, Richter DJ, Fozouni P, Smith TJ, Jeuck A, Leadbeater BSC, Nitsche F. 2017a. A six-gene phylogeny provides new insights into choanoflagellate evolution. *Mol Phylogenet Evol* [Internet] 107:166–178. Available from: http://dx.doi.org/10.1016/j.ympev.2016.10.011

Carr M, Richter DJ, Fozouni P, Smith TJ, Jeuck A, Leadbeater BSC, Nitsche F. 2017b. A six-gene phylogeny provides new insights into choanoflagellate evolution. *Mol Phylogenet Evol* [Internet] 107:166–178. Available from: http://dx.doi.org/10.1016/j.ympev.2016.10.011

Caterina MJ, Hereld D, Devreotes PN. 1995. Occupancy of the Dictyostelium cAMP receptor, cAR1, induces a reduction in affinity which depends upon COOH-terminal serine residues. *Journal of Biological Chemistry* 270:4418–4423.

Chadha A, Paniagua AE, Williams DS. 2021. Comparison of ciliary targeting of two rhodopsin-like gpcrs: Role of c-terminal localization sequences in relation to cilium type. *Journal of Neuroscience* 41:7514–7531.

Chun L, Zhang WH, Liu JF. 2012. Structure and ligand recognition of class C GPCRs. *Acta Pharmacol Sin* 33:312–323.

Crooks G, Hon G, Chandonia J, Brenner S. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190.

Crupi R, Impellizzeri D, Cuzzocrea S. 2019. Role of metabotropic glutamate receptors in neurological disorders. *Front Mol Neurosci* 12:1–11.

Dabravolski SA, Isayenkov S V. 2022. Evolution of the Membrane Transport Protein Domain. *Int J Mol Sci* 23:1–13.

Davies WL, Hankins MW, Foster RG. 2010. Vertebrate ancient opsin and melanopsin: Divergent irradiance detectors. *Photochemical and Photobiological Sciences* 9:1444–1457.

Dayel MJ, Alegado RA, Fairclough SR, Levin TC, Nichols SA, Mcdonald K, King N. 2011. Cell differentiation and morphogenesis in the colony-forming choano fl agellate Salpingoeca rosetta. *Dev Biol* 357:73–82.

Devine EL, Oprian DD, Theobald DL. 2013. Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proc Natl Acad Sci U S A* 110:13351–13355.

Dijkman PM, Muñoz-García JC, Lavington SR, Kumagai PS, dos Reis RI, Yin D, Stansfeld PJ, Costa-Filho AJ, Watts A. 2020. Conformational dynamics of a G protein–coupled receptor helix 8 in lipid membranes. *Sci Adv* 6:1–13.

Doktorgrades E, Obaid AA. 2022. Deciphering the function and evolutionary context of a novel G alpha protein , Gv.

Dubovski N, Fierro F, Margulis E, Ben Shoshan-Galeczki Y, Peri L, Niv MY. 2022. Taste GPCRs and their ligands. 1st ed. Elsevier Inc. Available from: http://dx.doi.org/10.1016/bs.pmbts.2022.06.008

Dudin O, Ondracka A, Grau-Bové X, Haraldsen AAB, Toyoda A, Suga H, Bråte J, Ruiz-Trillo I. 2019. A unicellular relative of animals generates a layer of polarized cells by actomyosin-dependent cellularization. *Elife* 8:1–26.

Dupré DJ, Robitaille M, Rebois RV, Hébert TE. 2009. The role of Gβγ subunits in the organization, assembly, and function of GPCR signaling complexes. *Annu Rev Pharmacol Toxicol* 49:31–56.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7.

Edgar RC. 2022. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun* 13:1–9.

Ellaithy A, Gonzalez-Maeso J, Logothetis DA, Levitz J. 2020a. Structural and Biophysical Mechanisms of Class C G Protein-Coupled Receptor Function. *Trends Biochem Sci* [Internet] 45:1049–1064. Available from: https://doi.org/10.1016/j.tibs.2020.07.008

Ellaithy A, Gonzalez-Maeso J, Logothetis DA, Levitz J. 2020b. Structural and Biophysical Mechanisms of Class C G Protein-Coupled Receptor Function. *Trends Biochem Sci* [Internet] 45:1049–1064.

Elliott GRD, Leys SP. 2010. Evidence for glutamate, GABA and NO in coordinating behaviour in the sponge, Ephydatia muelleri (Demospongiae, Spongillidae). *Journal of Experimental Biology* 213:2310–2321.

Evenseth LSM, Gabrielsen M, Sylte I. 2020. The GABAB Receptor—Structure, Ligand Binding and Drug Development. *Molecules*:157–180.

Fairclough S, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, Begovic E, Richter DJ, Russ C, Westbrook MJ, et al. 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. *Genome Biol* 14:1–15.

Fairclough SR, Dayel MJ, King N. 2010. Multicellular development in a choanoflagellate. *Current Biology* 20:875–876.

Feng Z, Sun R, Cong Y, Liu Z. 2022. Critical roles of G protein-coupled receptors in regulating intestinal homeostasis and inflammatory bowel disease. *Mucosal Immunol* [Internet] 15:819–828. Available from: http://dx.doi.org/10.1038/s41385-022-00538-3

Feuda R, Hamilton SC, McInerney JO, Pisani D. 2012. Metazoan opsin evolution reveals a simple route to animal vision. *Proc Natl Acad Sci U S A* 109:18868–18872.
Feuda R, Rota-Stabelli O, Oakley TH, Pisani D. 2014. The comb jelly opsins and the origins of animal phototransduction. *Genome Biol Evol* 6:1964–1971.
De Filippo E, Manga P, Schiedel AC. 2017. Identification of novel G Protein-Coupled Receptor 143 ligands as pharmacologic tools for investigating X-linked ocular albinism. *Invest Ophthalmol Vis Sci* 58:3118–3126.
Fleming JF, Feuda R, Roberts NW, Pisani D. 2020. A Novel Approach to Investigate the Effect of Tree Reconstruction Artifacts in Single-Gene Analysis Clarifies Opsin Evolution in Nonbilaterian Metazoans. *Genome Biol Evol* 12:3906–3916.
Francis WR, Eitel M, Vargas S, Garcia-Escudero CA, Conci N, Deister F, Mah JL, Guiglielmoni N, Krebs S, Blum H, et al. 2023. The genome of the reef-building glass sponge Aphrocallistes vastus provides insights into silica biomineralization. *R Soc Open Sci* 10.
Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB. 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* 63:1256–1272.
Fredriksson R, Schiöth HB. 2005. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol Pharmacol* 67:1414–1425.
Frickey T, Lupas A. 2004a. CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20:3702–3704.
Frickey T, Lupas A. 2004b. CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20:3702–3704.
Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* 72:1–30.
Galindo LJ, Torruella G, Moreira D, Eglit Y, Simpson AGB, Völcker E, Clauß S, López-García P. 2019. Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.
Gan L, Seki A, Shen K, Iyer H, Han K, Hayer A, Wollman R, Ge X, Lin JR, Dey G, et al. 2019. The lysosomal GPCR-like protein GPR137B regulates Rag and mTORC1 localization and activity. *Nat Cell Biol* [Internet] 21:614–626. Available from: http://dx.doi.org/10.1038/s41556-019-0321-6
Gilbert W. 1978. Why genes in pieces? *Nature* [Internet] 271. Available from: https://www.nature.com/articles/271501a0#citeas
Ginés-Rivas J, Carr M. 2025a. Slowly evolving proteins support the monophyly of Craspedida (Choanoflagellatea) and a marine origin of choanoflagellates. *Protist* 176.
Ginés-Rivas J, Carr M. 2025b. Slowly evolving proteins support the monophyly of Craspedida (Choanoflagellatea) and a marine origin of choanoflagellates. *Protist* [Internet]:126085. Available from: https://doi.org/10.1016/j.protis.2025.126085.
Glockling SL, Marshall WL, Gleason FH. 2013. Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). *Fungal Ecol* [Internet] 6:237–247. Available from: http://dx.doi.org/10.1016/j.funeco.2013.03.005

Gold DA, Grabenstatter J, De Mendoza A, Riesgo A, Ruiz-Trillo I, Summons RE. 2016. Sterol and genomic analyses validate the sponge biomarker hypothesis. *Proc Natl Acad Sci U S A* 113:2684–2689.

Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, Holstein T, Steele RE, Jacobs DK, Greenspan RJ. 2019. The genome of the jellyfish Aurelia and the evolution of animal complexity. *Nat Ecol Evol* [Internet] 3:96–104. Available from: http://dx.doi.org/10.1038/s41559-018-0719-8

Goldberg JM, Manning G, Liu A, Fey P, Pilcher KE, Xu Y, Smith JL. 2006. The Dictyostelium kinome - Analysis of the protein kinases from a simple model organism. *PLoS Genet* 2:0291–0303.

Goshima Y, Masukawa D, Kasahara Y, Hashimoto T, Aladeokin AC. 2019. l-DOPA and its receptor GPR143: Implications for pathogenesis and therapy in Parkinson's disease. *Front Pharmacol* 10:1–9.

Grau-Bové X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, Ruiz-Trillo I. 2017. Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* 6:1–35.

Greenhalgh JC, Chandran A, Harper MT, Ladds G, Rahman T. 2020a. Proposed model of the Dictyostelium cAMP receptors bound to cAMP. *J Mol Graph Model* 100.

Greenhalgh JC, Chandran A, Harper MT, Ladds G, Rahman T. 2020b. Proposed model of the Dictyostelium cAMP receptors bound to cAMP. *J Mol Graph Model* 100.

Gühmann M, Porter ML, Bok MJ. 2022. The Gluopsins: Opsins without the Retinal Binding Lysine. *Cells* 11.

Gurevich V V., Gurevich E V. 2019. GPCR signaling regulation: The role of GRKs and arrestins. *Front Pharmacol* 10:1–11.

Guzman C, Conaco C. 2016. Comparative transcriptome analysis reveals insights into the streamlined genomes of haplosclerid demosponges. *Sci Rep* 6:1–10.

Hagen JFD, Roberts NS, Johnston RJ. 2023. The evolutionary history and spectral tuning of vertebrate visual opsins. *Dev Biol* [Internet] 493:40–66. Available from: https://doi.org/10.1016/j.ydbio.2022.10.014

Hake KH, West PT, McDonald K, Laundon D, Reyes-Rivera J, Garcia De Las Bayonas A, Feng C, Burkhardt P, Richter DJ, Banfield JF, et al. 2024. A large colonial choanoflagellate from Mono Lake harbors live bacteria. *mBio* 15:e0162324.

Hall G, Kelly S, Schaap P, Schilde C. 2023. Phylogeny-wide analysis of G-protein coupled receptors in social amoebas and implications for the evolution of multicellularity. *Open Research Europe* 2.

Hamann J, Aust G, Araç D, Engel FB, Formstone C, Fredriksson R, Hall RA, Harty BL, Kirchhoff C, Knapp B, et al. 2015a. International union of basic and clinical pharmacology. XCIV. adhesion G protein-coupled receptors. *Pharmacol Rev* 67:338–367.

Hamann J, Aust G, Araç D, Engel FB, Formstone C, Fredriksson R, Hall RA, Harty BL, Kirchhoff C, Knapp B, et al. 2015b. International union of basic and clinical pharmacology. XCIV. adhesion G protein-coupled receptors. *Pharmacol Rev* 67:338–367.

Hankins MW, Davies WIL, Foster RG. 2014. The Evolution of Non-visual Photopigments in the Central Nervous System of Vertebrates.

Hankins MW, Peirson SN, Foster RG. 2008. Melanopsin: an exciting photopigment. *Trends Neurosci* 31:27–36.

1250 Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. 2017. Trends in GPCR drug discovery: New agents, targets and indications. *Nat Rev Drug Discov* [Internet] 16:829–842. Available from: http://dx.doi.org/10.1038/nrd.2017.178

Hauser AS, Kooistra AJ, Munk C, Heydenreich FM, Veprintsev DB, Bouvier M, Babu MM, Gloriam DE. 2021. GPCR activation mechanisms across classes and macro/microscales.
1255 *Nat Struct Mol Biol* 28:879–888.

Hehenberger E, Tikhonenkov D V., Kolisko M, del Campo J, Esaulov AS, Mylnikov AP, Keeling PJ. 2017. Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. *Current Biology* [Internet] 27:2043-2050.e6. Available from: http://dx.doi.org/10.1016/j.cub.2017.06.006

1260 Hilger D, Masureel M, Kobilka BK. 2018. Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* [Internet] 25:4–12. Available from: http://dx.doi.org/10.1038/s41594-017-0011-7

Hirata T, Fujita M, Nakamura S, Gotoh K, Motooka D, Murakami Y, Maeda Y, Kinoshita T. 2015. Post-Golgi anterograde transport requires GARP-dependent endosome-to-TGN retrograde
1265 transport. *Mol Biol Cell* 26:3071–3084.

Hoel CM, Zhang L, Brohawn SG. 2022. Structure of the GOLD-domain seven-transmembrane helix protein family member TMEM87A. *Elife* 11:1–15.

Holzem M, Boutros M, Holstein TW. 2024. The origin and evolution of Wnt signalling. *Nat Rev Genet* [Internet] 25:500–512. Available from: http://dx.doi.org/10.1038/s41576-024-00699-
1270 w

Van Den Hoogen DJ, Meijer HJG, Seidl MF, Govers F. 2018. The ancient link between G-protein-coupled receptors and C-terminal phospholipid kinase domains. *mBio* 9:1–14.

van den Hoogen J, Govers F. 2018. GPCR-bigrams: Enigmatic signaling components in oomycetes. *PLoS Pathog* 14:1–7.

1275 Hu J, Hauache O, Spiegel AM. 2000. Human Ca2+ receptor cysteine-rich domain: Analysis of function of mutant and chimeric receptors. *Journal of Biological Chemistry* [Internet] 275:16382–16389. Available from: http://dx.doi.org/10.1074/jbc.M000277200

Hua C, Meijer HJG, De Keijzer J, Zhao W, Wang Y, Govers F. 2013. GK4, a G-protein-coupled receptor with a phosphatidylinositol phosphate kinase domain in Phytophthora infestans, is
1280 involved in sporangia development and virulence. *Mol Microbiol* 88:352–370.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Iwasa K, Yamagishi A, Yamamoto S, Haruta C, Maruyama K, Yoshikawa K. 2023. GPR137 Inhibits Cell Proliferation and Promotes Neuronal Differentiation in the Neuro2a Cells.
1285 *Neurochem Res* [Internet] 48:996–1008. Available from: https://doi.org/10.1007/s11064-022-03833-4

Jeong E, Kim Y, Jeong J, Cho Y. 2021. Structure of the class C orphan GPCR GPR158 in complex with RGS7-Gβ5. *Nat Commun* 12:1–11.

Jiang P, Ji Q, Liu Z, Snyder LA, Benard LMJ, Margolskee RF, Max M. 2004. The cysteine-rich
1290 region of T1R3 determines responses to intensely sweet proteins. *Journal of Biological Chemistry* [Internet] 279:45068–45075. Available from: http://dx.doi.org/10.1074/jbc.M406779200

Johnson RL, Saxe CL, Gollop R, Kimmel AR, Devreotes PN. 1993. Identification and targeted gene disruption of cAR3, a cAMP receptor subtype expressed during multicellular stages of Dictyostelium development. *Genes Dev* 7:273–282.

Judelson HS, Ah-Fong AMV. 2010. The kinome of Phytophthora infestans reveals oomycete-specific innovations and links to other taxonomic groups. *BMC Genomics* 11.

Juliani MH, Klein C. 1981. Photoaffinity labeling of the cell surface adenosine 3':5'-monophosphate receptor of Dictyostelium discoideum and its modification in down-regulated cells. *Journal of Biological Chemistry* 256:613–619.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.

Kamesh N, Aradhyam GK, Manoj N. 2008. The repertoire of G protein-coupled receptors in the sea squirt Ciona intestinalis. *BMC Evol Biol* 8:1–19.

Karthikeyan R, Davies WIL, Gunhaga L. 2023. Non-image-forming functional roles of OPN3, OPN4 and OPN5 photopigments. *J Photochem Photobiol* [Internet] 15:100177. Available from: https://doi.org/10.1016/j.jpap.2023.100177

Katoh K, Rozewicki J, Yamada KD. 2018. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160–1166.

van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 42:243–246.

Kawabe Y, Morio T, James JL, Prescott AR, Tanaka Y, Schaap P. 2009. Activated cAMP receptors switch encystation into sporulation. *Proc Natl Acad Sci U S A* 106:7089–7094.

Kimura H, Usui T, Tsubouchi A, Uemura T. 2006. Potential dual molecular interaction of the Drosophila 7-pass transmembrane cadherin Flamingo in dendritic morphogenesis. *J Cell Sci* 119:1118–1129.

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451:783–788.

Klein PS, Sun TJ, Saxe CL, Kimmel AR, Johnson RL, Devreotes PN. 1988. A Chemoattractant Receptor Controls Development in Dictyostelium discoideum. *Science (1979)* [Internet] 241:1467–1472. Available from: https://www.science.org/doi/10.1126/science.3047871

Kleinau G, Ali AH, Wiechert F, Szczepek M, Schmidt A, Spahn CMT, Liebscher I, Schöneberg T, Scheerer P, Lin HHH, et al. 2022. The tethered peptide activation mechanism of adhesion GPCRs. *Nature* [Internet] 604:757–762. Available from: https://doi.org/10.1016/j.jbc.2022.102594

Knepp AM, Periole X, Marrink SJ, Sakmar TP, Huber T. 2012. Rhodopsin forms a dimer with cytoplasmic helix 8 contacts in native membranes. *Biochemistry* 51:1819–1821.

Knierim AB, Röthe J, Çakir MV, Lede V, Wilde C, Liebscher I, Thor D, Schöneberg T. 2019. Genetic basis of functional variability in adhesion G protein-coupled receptors. *Sci Rep* 9:1–15.

Kock I, Bulgakova NA, Knust E, Sinning I, Panneels V. 2009. Targeting of Drosophila rhodopsin requires helix 8 but not the distal C-terminus. *PLoS One* 4.

Kodner RB, Summons RE, Pearson A, King N, Knoll AH. 2008. Sterols in a unicellular relative of the metazoans. *Proc Natl Acad Sci U S A* 105:9897–9902.

Koehl A, Hu H, Feng D, Sun B, Zhang Y, Robertson MJ, Chu M, Kobilka TS, Laermans T, Steyaert J, et al. 2019. Structural insights into the activation of metabotropic glutamate receptors. *Nature* [Internet] 566:79–84. Available from: http://dx.doi.org/10.1038/s41586-019-0881-4

Koyanagi M, Takada E, Nagata T, Tsukamoto H, Terakita A. 2013. Homologs of vertebrate Opn3 potentially serve as a light sensor in nonphotoreceptive tissue. *Proc Natl Acad Sci U S A* 110:4998–5003.

Krishna AG, Menon ST, Terry TJ, Sakmar TP. 2002. Evidence that helix 8 of rhodopsin acts as a membrane-dependent conformational switch. *Biochemistry* 41:8298–8309.

Krishnan A. 2015. Evolution of the G protein-coupled receptor signaling system Genomic and phylogenetic analyses.

Krishnan A, Almén MS, Fredriksson R, Schiöth HB. 2012. The origin of GPCRs: Identification of mammalian like rhodopsin, adhesion, glutamate and frizzled GPCRs in fungi. *PLoS One* 7:1–15.

Krishnan A, Dnyansagar R, Almén MS, Williams MJ, Fredriksson R, Manoj N, Schiöth HB. 2014. The GPCR repertoire in the demosponge Amphimedon queenslandica: Insights into the GPCR system at the early divergence of animals. *BMC Evol Biol* 14:1–16.

Krishnan A, Mustafa A, Almén MS, Fredriksson R, Williams MJ, Schiöth HB. 2015. Evolutionary hierarchy of vertebrate-like heterotrimeric G protein families. *Mol Phylogenet Evol* [Internet] 91:27–40. Available from: http://dx.doi.org/10.1016/j.ympev.2015.05.009

Krishnan A, Schiöth HB. 2015. The role of G protein-coupled receptors in the early evolution of neurotransmission and the nervous system. *Journal of Experimental Biology* 218:562–571.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001a. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305:567–580.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001b. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305:567–580.

Kusuluri DK, Güler BE, Knapp B, Horn N, Boldt K, Ueffing M, Aust G, Wolfrum U. 2021. Adhesion G protein-coupled receptor VLGR1/ADGRV1 regulates cell spreading and migration by mechanosensing at focal adhesions. *iScience* 24.

Laboute T, Zucca S, Holcomb M, Patil DN, Garza C, Wheatley BA, Roy RN, Forli S, Martemyanov KA. 2023. Orphan receptor GPR158 serves as a metabotropic glycine receptor: mGlyR. *Science (1979)* 379:1352–1358.

Langenhan T, Aust G, Hamann J. 2013. Sticky signaling - Adhesion class g protein-coupled receptors take the stage. *Sci Signal* 6:1–22.

Langenhan T, Piao X, Monk KR. 2016. Adhesion G protein-coupled receptors in nervous system development and disease. *Nat Rev Neurosci* 17:550–561.

Latorraca NR, Venkatakrishnan AJ, Dror RO. 2017. GPCR dynamics: Structures in motion. *Chem Rev* 117:139–155.

Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52:W78–W82.

Levin TC, King N. 2013. Evidence for sex and recombination in the choanoflagellate Salpingoeca rosetta. *Curr Biol* [Internet] 23:2176–2180. Available from: http://dx.doi.org/10.1016/j.cub.2013.08.061

Levitz J, Habrian C, Bharill S, Fu Z, Vafabakhsh R, Isacoff EY. 2016. Mechanism of Assembly and Cooperativity of Homomeric and Heteromeric Metabotropic Glutamate Receptors. *Neuron* [Internet] 92:143–159. Available from: http://dx.doi.org/10.1016/j.neuron.2016.08.036

Li L, Tang J, Cao B, Xu Q, Xu S, Lin C, Tang C. 2024. GPR137 inactivates Hippo signaling to promote gastric cancer cell malignancy. *Biol Direct* [Internet] 19:1–16. Available from: https://doi.org/10.1186/s13062-023-00449-8

Li W, Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.

Liebscher I, Cevheroğlu O, Hsiao CC, Maia AF, Schihada H, Scholz N, Soave M, Spiess K, Trajković K, Kosloff M, et al. 2022. A guide to adhesion GPCR research. *FEBS Journal* 289:7610–7630.

Liu S, Anderson PJ, Rajagopal S, Lefkowitz RJ, Rockman HA. 2024. G Protein-Coupled Receptors: A Century of Research and Discovery. *Circ Res* 135:174–197.

Liu Y, Steenkamp ET, Brinkmann H, Forget L, Philippe H, Lang BF. 2009. Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BMC Evol Biol* 9:1–11.

Logan CY, Nusse R. 2004. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 20:781–810.

Lokits AD, Indrischek H, Meiler J, Hamm HE, Stadler PF. 2018. Tracing the evolution of the heterotrimeric G protein α subunit in Metazoa. *BMC Evol Biol* 18:1–27.

Lopez VM, Decatur CL, Stamer WD, Lynch RM, McKay BS. 2008. L-DOPA is an endogenous ligand for OA1. *PLoS Biol* 6:1861–1869.

López-Escardó D, Grau-Bové X, Guillaumet-Adkins A, Gut M, Sieracki ME, Ruiz-Trillo I. 2019. Reconstruction of protein domain evolution using single-cell amplified genomes of uncultured choanoflagellates sheds light on the origin of animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.

Louis JM, Ginsburg GT, Kimmel AR. 1994. The cAMP receptor CAR4 regulates axial patterning and cellular differentiation during late development of Dictyostelium. *Genes Dev* 8:2086–2096.

Lu P, Magwanga RO, Lu H, Kirungu JN, Wei Y, Dong Q, Wang X, Cai X, Zhou Z, Wang K, et al. 2018. A novel G-protein-coupled receptors gene from upland cotton enhances salt stress tolerance in transgenic Arabidopsis. *Genes (Basel)* 9.

Luo S, Zhang P, Miao W, Xiong J. 2023. Genome-Wide Identification of G Protein-Coupled Receptors in Ciliated Eukaryotes. *Int J Mol Sci* 24.

Lv X, Liu J, Shi Q, Tan Q, Wu D, Skinner JJ, Walker AL, Zhao L, Gu X, Chen N, et al. 2016. In vitro expression and analysis of the 826 human G protein-coupled receptors. *Protein Cell* 7:325–337.

Lyu S, Zhang X, Tu Z, Zhou H, Ke X, Qu Y. 2022. GPR108 is required for gambogic acid inhibiting NF-κB signaling in cancer. *Pharmacol Res* 182.

Malot H. 1878. Sans famille. Paris, France: Edouard Dentu

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. *Science (1979)* 298:1912–1934.
Marinissen MJ, Gutkind JS. 2001. G-protein-coupled receptors and signaling networks: Emerging paradigms. *Trends Pharmacol Sci* 22:368–376.
Marshall WL, Berbee ML. 2011. Facing Unknowns: Living cultures (Pirum gemmata gen. nov., sp. nov., and Abeoforma whisleri, gen. nov., sp. nov.) from invertebrate digestive tracts represent an undescribed clade within the unicellular opisthokont lineage ichthyosporea (Mesomycetozoea). *Protist* [Internet] 162:33–57. Available from: http://dx.doi.org/10.1016/j.protis.2010.06.002
McCulloch KJ, Babonis LS, Liu A, Daly CM, Martindale MQ, Koenig KM. 2023. Nematostella vectensis exemplifies the exceptional expansion and diversity of opsins in the eyeless Hexacorallia. *Evodevo* [Internet] 14:1–19. Available from: https://doi.org/10.1186/s13227-023-00218-8
McVeigh P, McCammick E, McCusker P, Wells D, Hodgkinson J, Paterson S, Mousley A, Marks NJ, Maule AG. 2018. Profiling G protein-coupled receptors of Fasciola hepatica identifies orphan rhodopsins unique to phylum Platyhelminthes. *Int J Parasitol Drugs Drug Resist* [Internet] 8:87–103. Available from: https://doi.org/10.1016/j.ijpddr.2018.01.001
Men LJ, Liu JZ, Chen HY, Zhang L, Chen SF, Xiao TW, Wang JX, Li GY, Wu YP. 2018. Down regulation of G protein-coupled receptor 137 expression inhibits proliferation and promotes apoptosis in leukemia cells. *Cancer Cell Int* [Internet] 18:1–10. Available from: https://doi.org/10.1186/s12935-018-0507-1
De Mendoza A, Sebé-Pedrós A, Ruiz-Trillo I. 2014. The evolution of the GPCR signaling system in eukaryotes: Modularity, conservation, and the transition to metazoan multicellularity. *Genome Biol Evol* 6:606–619.
Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* 32:1–13.
Mickael ME, Rajput A, Steyn J, Wiemerslage L, Bürglin T. 2016. An optimised phylogenetic method sheds more light on the main branching events of rhodopsin-like superfamily. *Comp Biochem Physiol Part D Genomics Proteomics* [Internet] 20:85–94. Available from: http://dx.doi.org/10.1016/j.cbd.2016.08.005
Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010*.
Milne JLS, Caterina MJ, Devreotes PN. 1997. Random mutagenesis of the cAMP chemoattractant receptor, cAR1, of Dictyostelium. Evidence for multiple states of activation. *Journal of Biological Chemistry* 272:2069–2076.
Minh BQ, Nguyen MAT, Von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195.
Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* 49:D412–D419.
Mitrovic SA, Demalgiriya-Gamage C, Winter LM, Kiechle T, Ebenhoch R, Neubauer H, Stierstorfer B, Frego L, Wolfrum C, Reindl S, et al. 2024. GPR180 is a new member of the Golgi-dynamics domain seven-transmembrane helix protein family. *Commun Biol* [Internet] 7. Available from: http://dx.doi.org/10.1038/s42003-024-07260-9

Mojib N, Kubanek J. 2020. Comparative transcriptomics supports the presence of G protein-coupled receptor-based signaling in unicellular marine eukaryotes. *Limnol Oceanogr* 65:762–774.
Mosienko V, Rasooli-Nejad S, Kishi K, De Both M, Jane D, Huentelman MJ, Kasparov S, Teschemacher AG. 2018. Putative Receptors Underpinning l-Lactate Signalling in Locus Coeruleus. *Neuroglia* 1:365–380.
Nagata T, Inoue K. 2021. Rhodopsins at a glance. *J Cell Sci* 134.
Najle SR, Molina MC, Ruiz-Trillo I, Uttaro AD. 2016. Sterol metabolism in the filasterean Capsaspora owczarzaki has features that resemble both fungi and animals. *Open Biol* 6.
Nathans J, Hogness DS. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* 34:807–814.
Neves SR, Ram PT, Iyengar R. 2002. G protein pathways. *Science (1979)* 296:1636–1639.
Niswender CM, Conn PJ. 2010. Metabotropic glutamate receptors: Physiology, pharmacology, and disease. *Annu Rev Pharmacol Toxicol* 50:295–322.
Nong W, Cao J, Li Y, Qu Z, Sun J, Swale T, Yip HY, PY, Qiu JW, Kwan HS, et al. 2020. Jellyfish genomes reveal distinct homeobox gene clusters and conservation of small RNA processing. *Nat Commun* [Internet] 11:1–11.
Nordström KJV, Fredriksson R, Schiöth HB. 2008. The amphioxus (Branchiostoma floridae) genome contains a highly diversified set of G protein-coupled receptors. *BMC Evol Biol* 8:6–8.
Nordström KJV, Lagerström MC, Wallér LMJ, Fredriksson R, Schiöth HB. 2009a. The Secretin GPCRs descended from the family of Adhesion GPCRs. *Mol Biol Evol* 26:71–84.
Nordström KJV, Lagerström MC, Wallér LMJ, Fredriksson R, Schiöth HB. 2009b. The Secretin GPCRs descended from the family of Adhesion GPCRs. *Mol Biol Evol* 26:71–84.
Nordström KJV, Sällman Almén M, Edstam MM, Fredriksson R, Schiöth HB. 2011. Independent HHsearch, Needleman-Wunsch-based, and motif analyses reveal the overall hierarchy for most of the g protein-coupled receptor families. *Mol Biol Evol* 28:2471–2480.
O'Brien JB, Wilkinson JC, Roman DL. 2019. Regulator of G-protein signaling (RGS) proteins as drug targets: Progress and future potentials. *Journal of Biological Chemistry* 294:18571–18585.
Ocaña-Pallarès E, Williams TA, López-Escardó D, Arroyo AS, Pathmanathan JS, Bapteste E, Tikhonenkov D V., Keeling PJ, Szöllősi GJ, Ruiz-Trillo I. 2022. Divergent genomic trajectories predate the origin of animals and fungi. *Nature* 609:747–753.
Oka Y, Korsching SI. 2009. The fifth element in animal Gα protein evolution. *Commun Integr Biol* 2:227–229.
Oka Y, Saraiva LR, Kwan YY, Korsching SI. 2009. The fifth class of Gα proteins. *Proc Natl Acad Sci U S A* 106:1484–1489.
Pal K, Melcher K, Xu HE. 2012. Structure and mechanism for recognition of peptide hormones by Class B G-protein-coupled receptors. *Acta Pharmacol Sin* 33:300–311.
Pándy-Szekeres G, Caroli J, Mamyrbekov A, Kermani AA, Keseru GM, Kooistra AJ, Gloriam DE. 2023. GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Res* 51:D395–D402.

Pándy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, Bojarski AJ, Gloriam DE. 2018. GPCRdb in 2018: Adding GPCR structure models and ligands. *Nucleic Acids Res* 46:D440–D446.

Papasergi-Scott MM, Robertson MJ, Seven AB, Panova O, Mathiesen JM, Skiniotis G. 2020. Structures of metabotropic GABAB receptor. *Nature* 584:310–314.

Patthy L. 1999. Genome evolution and the evolution of exon-shuffling - A review. *Gene* 238:103–114.

Patthy L. 2021. Exon shuffling played a decisive role in the evolution of the genetic toolkit for the multicellular body plan of metazoa. *Genes (Basel)* 12:1–14.

Peña JF, Alié A, Richter DJ, Wang L, Funayama N, Nichols SA. 2016. Conserved expression of vertebrate microvillar gene homologs in choanocytes of freshwater sponges. *Evodevo* 7:1–15.

Pérez-Cerezales S, Boryshpolets S, Afanzar O, Brandis A, Nevo R, Kiss V, Eisenbach M. 2015. Involvement of opsins in mammalian sperm thermotaxis. *Sci Rep* 5:1–18.

Pin JP, Galvez T, Prézeau L. 2003. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol Ther* 98:325–354.

Ping YQ, Xiao P, Yang F, Zhao RJ, Guo SC, Yan X, Wu X, Zhang C, Lu Y, Zhao F, et al. 2022. Structural basis for the tethered peptide activation of adhesion GPCRs. Springer US

Pita L, Hoeppner MP, Ribes M, Hentschel U. 2018. Differential expression of immune receptors in two marine sponges upon exposure to microbial-associated molecular patterns. *Sci Rep* 8:1–15.

Prömel S, Frickenhaus M, Hughes S, Mestek L, Staunton D, Woollard A, Vakonakis I, Schöneberg T, Schnabel R, Russ AP, et al. 2012a. The GPS Motif Is a Molecular Switch for Bimodal Activities of Adhesion Class G Protein-Coupled Receptors. *Cell Rep* 2:321–331.

Prömel S, Frickenhaus M, Hughes S, Mestek L, Staunton D, Woollard A, Vakonakis I, Schöneberg T, Schnabel R, Russ AP, et al. 2012b. The GPS Motif Is a Molecular Switch for Bimodal Activities of Adhesion Class G Protein-Coupled Receptors. *Cell Rep* 2:321–331.

Qian X, Zhao X, Yu L, Yin Y, Zhang XD, Wang L, Li JX, Zhu Q, Luo JL. 2023. Current status of GABA receptor subtypes in analgesia. *Biomedicine and Pharmacotherapy* [Internet] 168:115800.

Rader AJ, Anderson G, Isin B, Khorana HG, Bahar I, Klein-Seetharaman J. 2004. Identification of core amino acids stabilizing rhodopsin. *Proc Natl Acad Sci U S A* 101:7246–7251.

Rajagopal S, Shenoy SK. 2018. GPCR desensitization: Acute and prolonged phases. *Cell Signal* [Internet] 41:9–16. Available from: https://doi.org/10.1016/j.cellsig.2017.01.024

Richter DJ, Berney C, Strassert JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, De Vargas C. 2022. EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community Journal* [Internet] 2 (2022). Available from: https://peercommunityjournal.org/articles/10.24072/pcjournal.173/

Richter DJ, Fozouni P, Eisen MB, King N. 2018. Gene family innovation, conservation and loss on the animal stem lineage. *Elife* 7.

Riyahi TY, Frese F, Steinert M, Omosigho NN, Glöckner G, Eichinger L, Orabi B, Williams RSB, Noegel AA. 2011. RpkA, a highly conserved GPCR with a lipid kinase domain, has a role in phagocytosis and anti-bacterial defense. *PLoS One* 6.

Roh JW, Oh KS, Lee J, Choi Y, Kim S, Hong JW, Kim Y, Lew H, Jang SH, Shin H-S, et al. 2025. TMEM145 is a key component in stereociliary link structures of outer hair cells. *bioRxiv* [Internet]:2025.02.10.637577. Available from: http://biorxiv.org/content/early/2025/02/12/2025.02.10.637577.abstract

Rondard P, Liu J, Huang S, Malhaire F, Vol C, Pinault A, Labesse G, Pin JP. 2006. Coupling of agonist binding to effector domain activation in metabotropic glutamate-like receptors. *Journal of Biological Chemistry* [Internet] 281:24653–24661. Available from: http://dx.doi.org/10.1074/jbc.M602277200

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542.

Rosenkilde MM, Mathiesen JM. 2023. Glycine: a long-sought novel ligand for GPR158. *Trends Pharmacol Sci* [Internet] 44:489–491. Available from: https://doi.org/10.1016/j.tips.2023.05.004

Ros-Rocher N, Brunet T. 2023. What is it like to be a choanoflagellate? Sensation, processing and behavior in the closest unicellular relatives of animals. *Anim Cogn* [Internet] 26:1767–1782. Available from: https://doi.org/10.1007/s10071-023-01776-z

Ros-Rocher N, Pérez-Posada A, Leger MM, Ruiz-Trillo I. 2021. The origin of animals: An ancestral reconstruction of the unicellular-to-multicellular transition. *Open Biol* 11.

Rovati GE, Capra V, Neubig RR. 2007. The highly conserved DRY motif of class A G protein-coupled receptors: Beyond the ground state. *Mol Pharmacol* 71:959–964.

Roy D, Levi K, Kiss V, Nevo R, Eisenbach M. 2020. Rhodopsin and melanopsin coexist in mammalian sperm cells and activate different signaling pathways for thermotaxis. *Sci Rep* [Internet] 10:1–10.

Rutz C, Klein W, Schülein R. 2015. N-terminal signal peptides of g protein-coupled receptors: Significance for receptor biosynthesis, trafficking, and signal transduction. *Prog Mol Biol Transl Sci* 132:267–287.

Sakurai T, Kamakura S, Hayase J, Kohda A, Nakamura M, Sumimoto H. 2022. GPR125 (ADGRA3) is an autocleavable adhesion GPCR that traffics with Dlg1 to the basolateral membrane and regulates epithelial apicobasal polarity. *Journal of Biological Chemistry* [Internet] 298:102475.

Sandoval A, Eichler S, Madathil S, Reeves PJ, Fahmy K, Böckmann RA. 2016. The Molecular Switching Mechanism at the Conserved D(E)RY Motif in Class-A GPCRs. *Biophys J* 111:79–89.

Santini S, Schenkelaars Q, Jourda C, Duchesne M, Belahbib H, Rocher C, Selva M, Riesgo A, Vervoort M, Leys SP, et al. 2023. The compact genome of the sponge Oopsacas minuta (Hexactinellida) is lacking key metazoan core genes. *BMC Biol* [Internet] 21:1–21. Available from: https://doi.org/10.1186/s12915-023-01619-w

Sarre LA, Kim I V., Ovchinnikov V, Olivetta M, Suga H, Dudin O, Sebé-Pedrós A, de Mendoza A. 2024. DNA methylation enables recurrent endogenization of giant viruses in an animal relative. *Sci Adv* 10:1–13.

1595 Sato M, Blumer JB, Simon V, Lanier SM. 2006. Accessory proteins for G proteins: Partners in signaling. *Annu Rev Pharmacol Toxicol* 46:151–187.

Saxe CL, Ginsburg GT, Louis JM, Johnson R, Devreotes PN, Kimmel AR. 1993. CAR2, a prestalk cAMP receptor required for normal tip formation and late development of Dictyostelium discoideum. *Genes Dev* 7:262–272.

1600 Schiaffino MV, D'Addio M, Alloni A, Baschirotto C, Valetti C, Cortese K, Puri C, Bassi MT, Colla C, De Luca M, et al. 1999. Ocular albinism: Evidence for a defect in an intracellular signal transduction system. *Nat Genet* 23:108–112.

Schiaffino MV, Tacchetti C. 2005. The ocular albinism type 1 (OA1) protein and the evidence for an intracellular signal transduction system involved in melanosome biogenesis. *Pigment*
1605 *Cell Res* 18:227–233.

Schiöth HB, Fredriksson R. 2005. The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen Comp Endocrinol* 142:94–101.

Schiöth HB, Lagerström MC. 2008a. Structural diversity of g proteincoupled receptors and significance for drug discovery. *Nat Rev Drug Discov* 7:339–357.

1610 Schiöth HB, Lagerström MC. 2008b. Structural diversity of g proteincoupled receptors and significance for drug discovery. *Nat Rev Drug Discov* 7:339–357.

Scholz N, Langenhan T, Schöneberg T. 2019. Revisiting the classification of adhesion GPCRs. *Ann N Y Acad Sci* 1456:80–95.

Schöneberg T. 2024. Beyond the classic GPCR: unraveling the role of GPR155 role in
1615 cholesterol sensing and signaling. *Signal Transduct Target Ther* [Internet] 9:2–4. Available from: http://dx.doi.org/10.1038/s41392-024-02059-w

Schulte G. 2024. International Union of Basic and Clinical Pharmacology. CXV: The Class F of G Protein-Coupled Receptors. *Pharmacol Rev* [Internet] 76:PHARMREV-INR-2023-001062. Available from: https://doi.org/10.1124/pharmrev.124.001062

1620 Schultz DT, Francis WR, McBroome JD, Christianson LM, Haddock SHD, Green RE. 2021. A chromosome-scale genome assembly and karyotype of the ctenophore Hormiphora californensis. *G3: Genes, Genomes, Genetics* 11.

Sebé-Pedrós A, Irimia M, del Campo J, Parra-Acero H, Russ C, Nusbaum C, Blencowe BJ, Ruiz-Trillo I. 2013. Regulated aggregative multicellularity in a close unicellular relative of
1625 metazoa. *Elife* 2013:1–20.

Sensoy O, Weinstein H. 2015. A mechanistic role of Helix 8 in GPCRs: Computational modeling of the dopamine D2 receptor interaction with the GIPC1-PDZ-domain. *Biochim Biophys Acta Biomembr* [Internet] 1848:976–983. Available from: http://dx.doi.org/10.1016/j.bbamem.2014.12.002

1630 Seufert F, Chung YK, Hildebrand PW, Langenhan T. 2023. 7TM domain structures of adhesion GPCRs: what's new and what's missing? *Trends Biochem Sci* [Internet] 48:726–739. Available from: https://doi.org/10.1016/j.tibs.2023.05.007

Shamsuzzama, Lebedev R, Trabelcy B, Langier Goncalves I, Gerchman Y, Sapir A. 2020. Metabolic Reconfiguration in C. elegans Suggests a Pathway for Widespread Sterol
1635 Auxotrophy in the Animal Kingdom. *Current Biology* [Internet] 30:3031-3038.e7. Available from: https://doi.org/10.1016/j.cub.2020.05.070

Shima Y, Kengaku M, Hirano T, Takeichi M, Uemura T. 2004. Regulation of dendritic maintenance and growth by a mammalian 7-pass transmembrane cadherin. *Dev Cell* 7:205–216.

Shin HR, Citron YR, Wang L, Tribouillard L, Goul CS, Stipp R, Sugasawa Y, Jain A, Samson N, Lim CY, et al. 2022. Lysosomal GPCR-like protein LYCHOS signals cholesterol sufficiency to mTORC1. *Science (1979)* 377:1290–1298.

Shin JJH, Crook OM, Borgeaud AC, Cattin-Ortolá J, Peak-Chew SY, Breckels LM, Gillingham AK, Chadwick J, Lilley KS, Munro S. 2020. Spatial proteomics defines the content of trafficking vesicles captured by golgin tethers. *Nat Commun* 11.

Sigel E, Steinmann ME. 2012. Structure, function, and modulation of GABAA receptors. *Journal of Biological Chemistry* 287:40224–40231.

Silva L, Antunes A. 2017. Vomeronasal Receptors in Vertebrates and the Evolution of Pheromone Detection. *Annu Rev Anim Biosci* 5:353–370.

Simundza J, Cowin P. 2013. Adhesion G-Protein-Coupled Receptors: Elusive Hybrids Come of Age. *Cell Commun Adhes* [Internet] 20:213–225.

Smithers B, Oates M, Gough J. 2019. “Why genes in pieces?” - Revisited. *Nucleic Acids Res* 47:4970–4973.

Srivastava D, Gakhar L, Artemyev NO. 2019. Structural underpinnings of Ric8A function as a G-protein α-subunit chaperone and guanine-nucleotide exchange factor. *Nat Commun* [Internet] 10. Available from: http://dx.doi.org/10.1038/s41467-019-11088-x

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, et al. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466:720–726.

Steenwyk JL, Buida TJ, Li Y, Shen XX, Rokas A. 2020. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* [Internet] 18:1–17. Available from: http://dx.doi.org/10.1371/journal.pbio.3001007

Steffen K, Proux-Wéra E, Soler L, Churcher A, Sundh J, Cárdenas P. 2023. Whole genome sequence of the deep-sea sponge Geodia barretti (Metazoa, Porifera, Demospongiae). *G3: Genes, Genomes, Genetics* [Internet] 13:1–9. Available from: https://doi.org/10.1093/g3journal/jkad192

Stoveken HM, Hajduczok AG, Xu L, Tall GG. 2015. Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proc Natl Acad Sci U S A* 112:6194–6199.

Suga H, Chen Z, De Mendoza A, Sebé-Pedrós A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sánchez-Pons N, et al. 2013. The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nat Commun* 4:1–9.

Taipale J, Beachy PA. 2001. Pathways in Cancer. *Nature* 411.

Takeo Y, Kurabayashi N, Nguyen MD, Sanada K. 2016. The G protein-coupled receptor GPR157 regulates neuronal differentiation of radial glial progenitors through the Gq-IP3 pathway. *Sci Rep* 6:1–12.

Taniura H, Sanada N, Kuramoto N, Yoneda Y. 2006. A metabotropic glutamate receptor family gene in dictyostelium discoideum. *Journal of Biological Chemistry* [Internet] 281:12336–12343. Available from: http://dx.doi.org/10.1074/jbc.M512723200

1680 Thiel D, Guerra LAY, Kieswetter A, Cole AG, Temmerman L, Technau U, Jékely G. 2023. Large-scale deorphanization of Nematostella vectensis neuropeptide G protein-coupled receptors supports the independent expansion of bilaterian and cnidarian peptidergic systems. *Elife* 12:1–27.

Tikhonenkov D V., Hehenberger E, Esaulov AS, Belyakova OI, Mazei YA, Mylnikov AP, Keeling
1685 PJ. 2020. Insights into the origin of metazoan multicellularity from predatory unicellular relatives of animals. *BMC Biol* 18:1–24.

Tikhonenkov D V., Mikhailov K V., Gawryluk RMR, Belyaev AO, Mathur V, Karpov SA, Zagumyonnyi DG, Borodina AS, Prokina KI, Mylnikov AP, et al. 2022. Microbial predators form a new supergroup of eukaryotes. *Nature* 612:714–719.

1690 Torruella G, De Mendoza A, Grau-Bové X, Antó M, Chaplin MA, Del Campo J, Eme L, Pérez-Cordón G, Whipps CM, Nichols KM, et al. 2015. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology* 25:2404–2410.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online
1695 phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 44:W232–W235.

Umeda S, Kanda M, Sugimoto H, Tanaka H, Hayashi M, Yamada S, Fujii T, Takami H, Niwa Y, Iwata N, et al. 2017. Downregulation of GPR155 as a prognostic factor after curative resection of hepatocellular carcinoma. *BMC Cancer* 17:1–8.

Vargas AM, Debiasse MB, Dykes LL, Edgar A, Hayes TD, Groso DJ, Babonis LS, Martindale
1700 MQ, Ryan JF. 2024. Morphological and dietary changes encoded in the genome of Beroe ovata, a ctenophore-eating ctenophore. *NAR Genom Bioinform* 6.

Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, et al. 2003. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci U S A* 100:4903–4908.

1705 XC, Liu Z, Jin LH. 2018. Anchor negatively regulates BMP signalling to control Drosophila wing development. *Eur J Cell Biol* [Internet] 97:308–317.

Wang P, Jiang X, Zhu L, Zhou D, Hong M, He L, Chen L, Yao S, Zhao Y, Chen G, et al. 2022. A G-Protein-Coupled Receptor Modulates Gametogenesis via PKG-Mediated Signaling Cascade in Plasmodium berghei. *Microbiol Spectr* 10:1–20.

1710 Weis WI, Kobilka BK. 2018. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* 87:897–919.

Weston MD, Luijendijk MWJ, Humphrey KD, Möller C, Kimberling WJ. 2004. Mutations in the VLGR1 Gene Implicate G-Protein Signaling in the Pathogenesis of Usher Syndrome Type II. *Am J Hum Genet* 74:357–366.

1715 Willard FS, Kimple RJ, Siderovski DP. 2004. Return of the GDI: The GoLoco motif in cell division. *Annu Rev Biochem* 73:925–952.

Willardson BM, Howlett AC. 2007. Function of phosducin-like proteins in G protein signaling and chaperone-assisted protein folding. *Cell Signal* 19:2417–2427.

Wistrand M, Käll L, Sonnhammer ELL. 2006. A general model of G protein-coupled receptor
1720 sequences and its application to detect remote homologs. *Protein Science* 15:509–521.

Wittlake A, Prömel S, Schöneberg T. 2021. The evolutionary history of vertebrate adhesion GPCRs and its implication on their classification. *Int J Mol Sci* 22.

Wong E, Anggono V, Williams SR, Degnan SM, Degnan BM. 2022. Phototransduction in a marine sponge provides insights into the origin of animal vision. *iScience* [Internet] 25:104436. Available from: https://doi.org/10.1016/j.isci.2022.104436

Wootten D, Christopoulos A, Marti-Solano M, Babu MM, Sexton PM. 2018. Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat Rev Mol Cell Biol* [Internet] 19:638–653. Available from: http://dx.doi.org/10.1038/s41580-018-0049-3

Woznica A, Gerdt JP, Hulett RE, Clardy J, King N. 2017. Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase. *Cell* [Internet] 170:1175-1183.e11. Available from: http://dx.doi.org/10.1016/j.cell.2017.08.005

Wu Y, Janetopoulos C. 2013. Systematic analysis of γ-aminobutyric acid (GABA) metabolism and function in the social amoeba dictyostelium discoideum. *Journal of Biological Chemistry* [Internet] 288:15280–15290. Available from: http://dx.doi.org/10.1074/jbc.M112.427047

Xiao P, Guo S, Wen X, He QT, Lin H, Huang SM, Gou L, Zhang C, Yang Z, Zhong YN, et al. 2022. Tethered peptide activation mechanism of the adhesion GPCRs ADGRG2 and ADGRG4. *Nature* 604:771–778.

Yamanaka T, Ohta T, Takahashi M, Meshi T, Schmidt R, Dean C, Naito S, Ishikawa M. 2000. TOM1, an Arabidopsis gene required for efficient multiplication of a tobamovirus, encodes a putative transmembrane protein. *Proc Natl Acad Sci U S A* 97:10107–10112.

Yañez-Guerra LA, Thiel D, Jékely G. 2022. Premetazoan Origin of Neuropeptide Signaling. *Mol Biol Evol* 39:1–11.

Yang D, Zhou Q, Labroska V, Qin S, Darbalaei S, Wu Y, Yuliantie E, Xie L, Tao H, Cheng J, et al. 2021. G protein-coupled receptors: structure- and function-based drug discovery. *Signal Transduct Target Ther* [Internet] 6. Available from: http://dx.doi.org/10.1038/s41392-020-00435-w

Yang S, Tang Q, Zhang Y, Du Y, Zhao X, Mei F, Li Y. 2024. Neuronostatin regulates neuronal function and energetic metabolism in Alzheimer's disease in a GPR107-dependent manner. *Neuropharmacology* [Internet] 258:110090.

Yona S, Lin HH, Siu WO, Gordon S, Stacey M. 2008. Adhesion-GPCRs: emerging roles for novel receptors. *Trends Biochem Sci* 33:491–500.

Yoshida K, Yokota K, Watanabe K, Tsuda H, Matsumoto A, Mizukami H, Iwamoto S. 2023. Lack of GPR180 ameliorates hepatic lipid depot via downregulation of mTORC1 signaling. *Sci Rep* [Internet] 13:1–11.

Yosten GLC, Redlinger LJ, Samson WK. 2012. Evidence for an interaction of neuronostatin with the orphan G protein-coupled receptor, GPR107. *Am J Physiol Regul Integr Comp Physiol* 303:941–949.

Yun Y, Jeong H, Laboute T, Martemyanov KA, Lee HH. 2024. Cryo-EM structure of human class C orphan GPCR GPR179 involved in visual processing. *Nat Commun* [Internet] 15:8299. Available from: http://dx.doi.org/10.1038/s41467-024-52584-z

Zhang M, Chen T, Lu X, Lan X, Chen Z, Lu S. 2024. G protein-coupled receptors (GPCRs): advances in structures, mechanisms, and drug discovery. *Signal Transduct Target Ther* [Internet] 9:1–43. Available from: http://dx.doi.org/10.1038/s41392-024-01803-6

1765 Zhang T, Yuan D, Xie J, Lei Y, Li J, Fang G, Tian L, Liu J, Cui Y, Zhang M, et al. 2019. Evolution of the Cholesterol Biosynthesis Pathway in Animals. *Mol Biol Evol* 36:2548–2556.

Zhao L hua, Yuan Q ning, Dai A tao, He X heng, Chen C wei, Zhang C, Xu Y wei, Zhou Y, Wang M wei, Yang D hua, et al. 2023. Molecular recognition of two endogenous hormones
1770 by the human parathyroid hormone receptor-1. *Acta Pharmacol Sin* 44:1227–1237.

Zheng S, Sheng R. 2024. The emerging understanding of Frizzled receptors. *FEBS Lett* 598:1939–1954.

Zhou GL, Na SY, Niedra R, Seed B. 2014. Deficits in receptor-mediated endocytosis and recycling in cells from mice with Gpr107 locus disruption. *J Cell Sci* 127:3916–3927.

1775

1780

1785

1790

1795

1800

1805

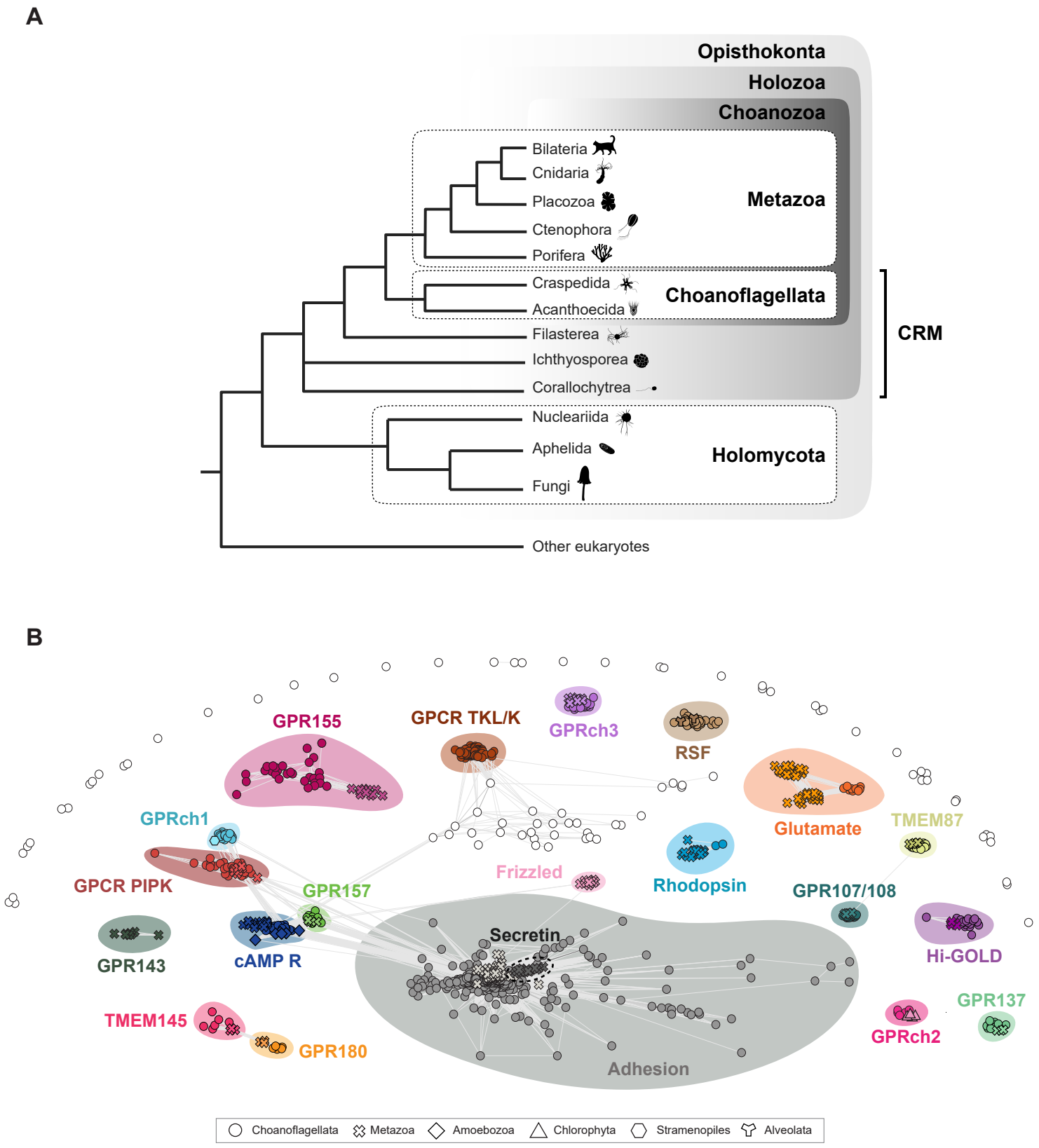

A
Opisthokonta
Holozoa
Choanozoa
Bilateria
Cnidaria
Placozoa
Ctenophora
Porifera
Metazoa
Craspedida
Acanthoecida
Choanoflagellata
Filasterea
Ichthyosporea
Corallochytrea
CRM
Nucleariida
Aphelida
Fungi
Holomycota
Other eukaryotes
B
GPR155
GPCR TKL/K
GPRch3
RSF
Glutamate
TMEM87
GPRch1
GPCR PIPK
GPR157
Frizzled
Rhodopsin
GPR107/108
GPR143
cAMP R
Secretin
Hi-GOLD
TMEM145
GPR180
GPRch2
GPR137
Adhesion
Choanoflagellata
Metazoa
Amoebozoa
Chlorophyta
Stramenopiles
Alveolata

Fig. 1

Figure 1: **Sequence clustering reveals similarities among the 7TMs of GPCRs from choanoflagellates, metazoans, and other eukaryotes. (A)** Choanoflagellates are the closest living relatives of metazoans. Shown is a consensus phylogeny of the major lineages analyzed in this study. We use the term "close relatives of metazoans" (CRM) to denote the paraphyletic group of non-metazoan holozoans that includes choanoflagellates, filastereans, ichthyosporeans, and corallochytreans. Organism silhouettes are from PhyloPic (http://phylopic.org/). **(B)** Most choanoflagellate GPCRs cluster with GPCRs from metazoans and other eukaryotes. The 918 choanoflagellate GPCRs (circles) identified in this study were sorted into clusters based on sequence similarity of their 7TM domains and the 7TM domains of metazoan, amoebozoan, chlorophyte, stramenopile, and alveolate GPCRs. Connecting lines (light grey) correspond to pairwise BLAST scores of p-value $<1e^{-6}$. With the exceptions of RSF, GPCR TKL/K, GPRch1, and GPRch2, all choanoflagellate GPCR clusters contained metazoan GPCRs and, in most cases, GPCRs from other eukaryotes. No choanoflagellate GPCRs clustered with metazoan Secretin GPCRs, Frizzled GPCRs, or GPR143 GPCRs. The collection of choanoflagellate GPCRs shown as open circles did not meet the statistical threshold for designation as clusters. All the GPCR sequences used in this analysis are provided in Supplementary File 6.

	Adhesion	GPR107/108	TMEM87 *	TMEM145 *	Hi-GOLD *	RSF *	cAMP	GPR137	GPRch2 *	GPR155 *	GPRch3 *	GPR157 *	Glutamate	GPCR PIPK	GPRch1 *	GPR180	GPCR TKL/K *	Rhodospsin *	Frizzled	GPR143
H. balthica	10	1	1	nd	1	2	1	nd	nd	1	nd	nd	nd	nd	2	nd	nd	nd	nd	nd
H. gracilis	12	nd	1	nd	1	2	1	nd	nd	nd	nd	nd	nd	1	1	nd	nd	nd	nd	nd
S. infusionum	11	1	1	nd	1	1	1	1	nd	1	1	nd	nd	1	1	nd	nd	nd	nd	nd
G *S. rosetta*	12	1	1	0	0	0	1	0	0	2	0	0	1	1	1	0	0	0	0	0
G *B. monosierra*	6	1	1	0	0	0	2	0	0	2	0	0	1	1	0	0	0	0	0	0
M. roanoka	7	1	1	nd	1	1	nd	nd	nd	1	nd	nd	1	1	nd	nd	nd	nd	nd	nd
G *M. brevicollis*	9	1	1	0	0	0	1	1	nd	3	2	0	1	1	1	0	0	0	0	0
C. perplexa	13	1	1	nd	1	1	1	1	nd	4	nd	nd	1	3	1	nd	nd	nd	nd	nd
C. flexa	10	1	1	nd	1	1	1	1	nd	5	1	nd	1	1	1	nd	nd	nd	nd	nd
S. kvevrii	19	2	1	2	2	1	1	2	1	nd	1	1	1	2	1	nd	nd	nd	nd	nd
S. urceolata	20	1	1	nd	1	1	2	1	1	1	1	1	nd	1	2	nd	nd	nd	nd	nd
S. macrocollata	18	1	1	nd	2	1	nd	2	1	nd	1	1	nd	nd	1	nd	nd	1	nd	nd
S. punica	18	1	1	nd	nd	1	nd	1	1	1	1	2	nd	nd	1	nd	nd	1	nd	nd
S. helianthica	16	1	2	1	nd	2	1	1	nd	nd	3	1	1	1	nd	nd	nd	nd	nd	nd
M. fluctuans	17	1	3	1	nd	1	1	1	nd	nd	3	1	1	1	2	nd	nd	nd	nd	nd
C. hollandica	24	1	2	nd	1	5	nd	nd	1	2	4	1	nd	5	nd	nd	nd	nd	nd	nd
S. dolichothecata	22	1	1	1	1	1	1	2	1	3	2	1	2	nd	1	nd	nd	nd	nd	nd
S. parva	23	1	1	nd	nd	1	2	nd	1	nd	1	nd	nd	nd	nd	1	10	nd	nd	nd
A .spectabilis	41	1	1	nd	1	1	3	nd	1	nd	2	nd	nd	nd	nd	2	69	nd	nd	nd
H. nana	35	1	1	nd	1	2	2	nd	1	nd	2	nd	nd	nd	nd	1	43	nd	nd	nd
D. costata	26	1	1	1	nd	1	1	1	1	1	1	nd	nd	nd	nd	1	nd	nd	nd	nd
D. grandis	23	1	1	1	1	1	2	nd	1	1	1	nd	nd	nd	nd	1	28	nd	nd	nd
S. diplocostata	25	1	1	1	1	1	2	nd	2	nd	1	nd	nd	nd	nd	nd	26	nd	nd	nd
Metazoa					(b)	nd			nd		(b)			(d)	nd		nd			
Filasterea											nd				nd		nd	nd	nd	
Ichthyosporea				nd		nd				nd	nd	nd			nd	nd	nd		nd	nd
Corallochytrea					nd	nd	nd	nd				nd		nd	nd	nd	nd	nd	(f)	nd
Holomycota		(a)		nd	nd	nd		(a)	nd	nd	nd	(a)		nd	nd	nd	nd			nd
Other euks									(c)	nd							nd	(e)		nd

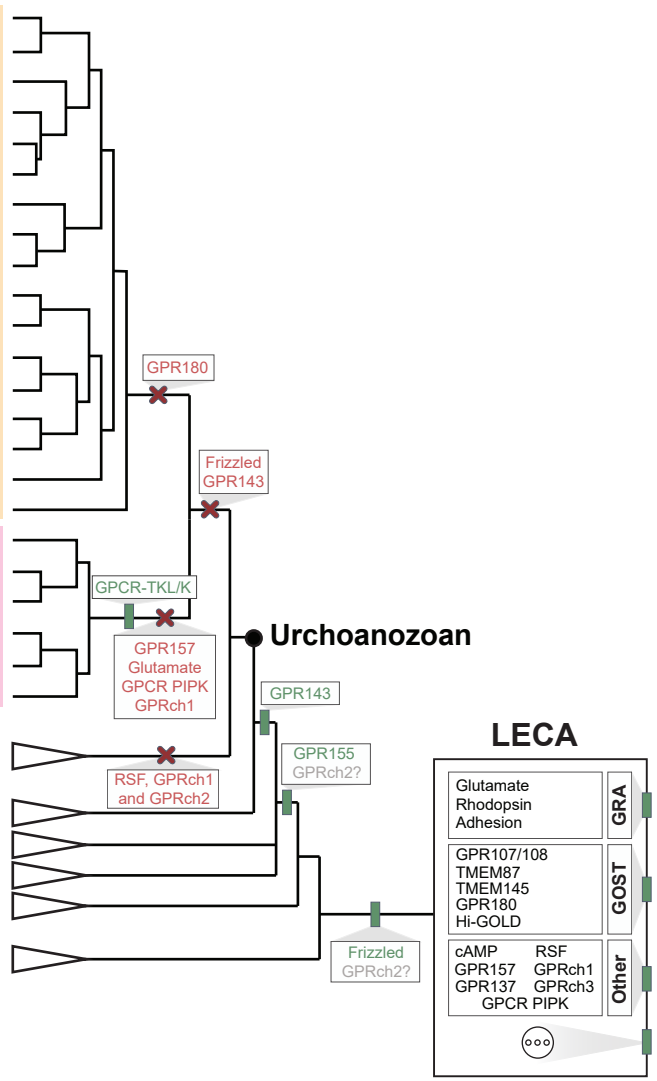


Fig. 2

Figure 2. Evolutionary history of GPCR families detected in choanoflagellates, metazoans, and other eukaryotes.

(Left) Taxonomic distribution of GPCR families. Shown in the table are the numbers of GPCR family members (columns) detected in each lineage (rows). The GPCR families are grouped based on their inferred phylogenetic distributions in choanoflagellates: pan-choanoflagellate GPCRs (light blue), GPCR families found in craspedids but not acanthoecids (orange), GPCR families found in acanthoecids but not craspedids (pink), and GPCRs found only in *S. macrocollata* and *S. punica* (dark blue). Adhesion, Glutamate, Rhodopsin, and Frizzled, four of the five GRAFS GPCR families, are indicated in bold. Asterisks indicate GPCR families that were not known to exist in choanoflagellates prior to this study. For those species in which only transcriptomes are available, gene numbers represent a minimum. nd = not detected in lineages for which only transcriptome data are available. For species with both transcriptome and genome data (*S. rosetta*, *B. monosierra*, and *M. brevicollis*; G enclosed within a black circle) (King et al. 2008; Fairclough et al. 2013; Hake et al. 2024), failure to detect a GPCR subfamily member is indicated with a “0”. (a) only found in nucleariids, (b) lost in vertebrates, (c) only found in chlorophytes, (d) only found in sponges, (e) only found in amoebozoans, (f) only found in *Syssomonas multiformis*. **(Right)** A consensus phylogeny shows the relationships among the 23 choanoflagellates included in this study, metazoans, filastereans, ichthyosporeans, corallochytreans, holomycotans, and diverse other eukaryotes. The inferred origins (vertical green rectangle) and subsequent losses (red cross) of GPCR families are indicated at relevant branches on the consensus phylogeny. GPCR families inferred to have originated in the Last Eukaryotic Common Ancestor (LECA) are represented at the root of the phylogeny (box). The presence of additional GPCR families in LECA, not covered in our study, is depicted by three dots. “GRA” indicates three of the GRAFS GPCR families: Glutamate, Rhodopsin, and Adhesion. “GOST” indicates subfamilies of GOST GPCRs. Additional GPCR families are listed under “Other.” Uncertainty about the ancestry of GPRch2 is indicated with a question mark.

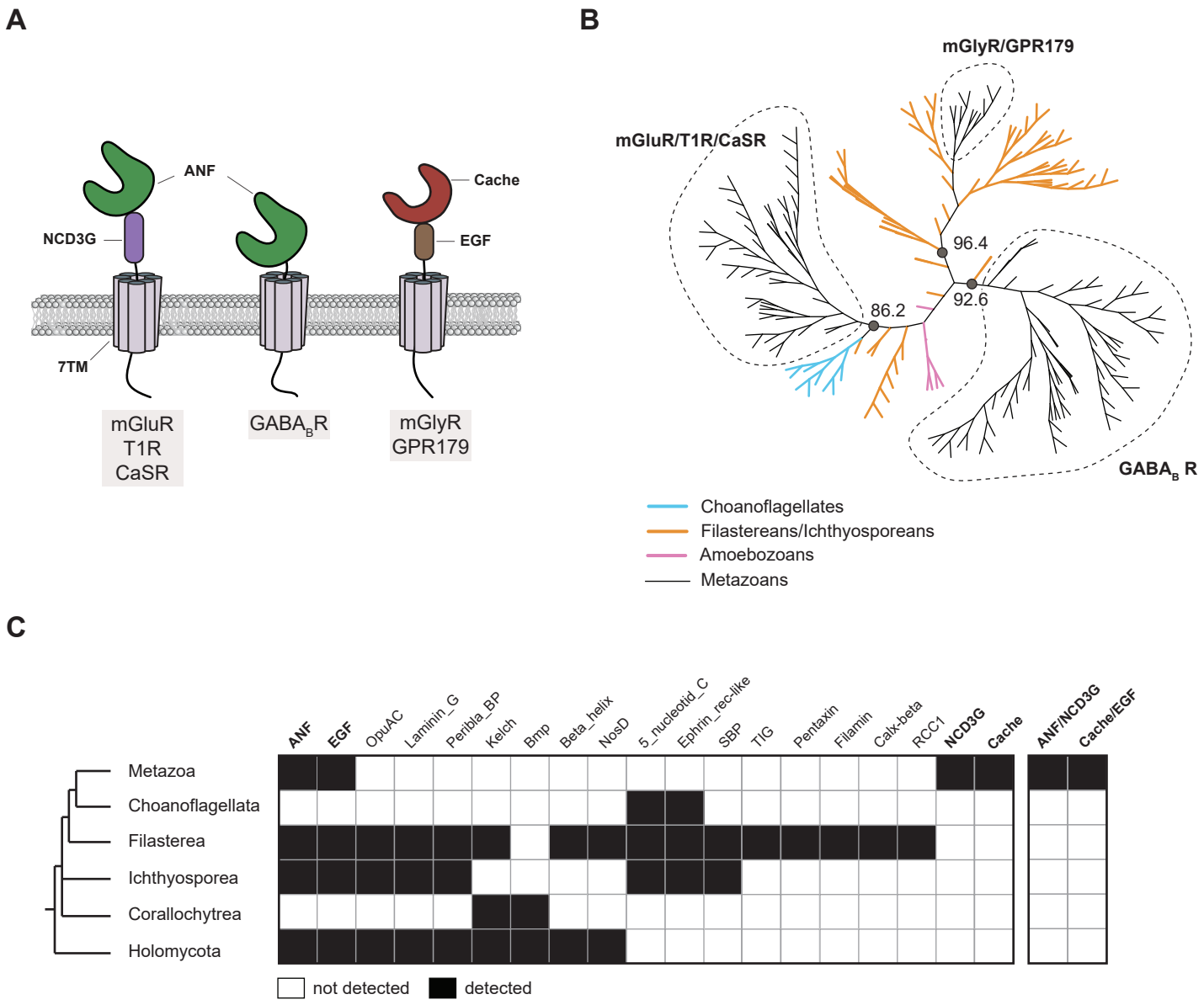

A
ANF
Cache
NCD3G
EGF
7TM
mGluR
T1R
CaSR
GABA$_B$R
mGlyR
GPR179
B
mGlyR/GPR179
mGluR/T1R/CaSR
96.4
92.6
86.2
GABA$_B$ R
Choanoflagellates
Filastereans/Ichthyosporeans
Amoebozoans
Metazoans
C
ANF
EGF
OpuAC
Laminin_G
Peribla_BP
Kelch
Bmp
Beta_helix
NosD
5_nucleotid_C
Ephrin_rec-like
SBP
TIG
Pentaxin
Filamin
Calx-beta
RCC1
NCD3G
Cache
ANF/NCD3G
Cache/EGF
Metazoa
Choanoflagellata
Filasterea
Ichthyosporea
Corallochytrea
Holomycota
not detected
detected

Fig. 3

Figure 3: Evolution of Glutamate Receptor protein domain architecture. (A) Schematic representation of the main protein domains found in metazoan the Glutamate Receptor family. GluR/T1R/CaSR receptors have a conserved extracellular module containing an ANF ligand-binding domain fused to a cysteine-rich NCD3G domain. $GABA_BR$ GPCRs contain only the ANF ligand binding domain. mGlyR/GPR179 GPCRs contain a Cache ligand-binding domain fused to an EGF-like domain. **(B)** Phylogenetic analysis of the 7TM domains of metazoan, choanoflagellate, filasterean, ichthyosporean, and amoebozoan Glutamate Receptors yielded three well-supported clades of GPCRs corresponding to the receptors depicted in panel A. Filasteran and ichthyosporean sequences (orange) clustered as the sister group to metazoan mGluR/CaSR/T1R, mGlyR/GPR179, and $GABA_B$ GPCRs, or branch separately from the rest of GPCRs used in this analysis. Choanoflagellate sequences (light blue) branch as the sister group to the metazoan mGluR/CaSR/T1R GPCRs. No amoebozoan sequence (pink) clustered with metazoan or CRMs. UFboot support values are indicated for each three major nodes of this unrooted maximum-likelihood phylogenetic tree. All the sequences used to build this phylogeny are listed in Supplementary Files 8 and 9. **(C)** Phylogenetic distribution of ECDs detected in Glutamate Receptors from diverse opisthokonts. While the ANF and EGF protein domains evolved before the diversification of opisthokonts, the presence of the NCD3G and Cache protein domains in GPCRs was not detected in any non-metazoan. Similarly, the ANF/NCD3G and Cache/EGF protein domain modules were only detected in metazoans. Glutamate Receptors from non-metazoans contain diverse extracellular protein domains that are not found in metazoan GPCRs. Protein domains (left) or modules containing pairs of protein domains (right) are indicated in the columns. Taxonomic distribution is indicated in the rows. Protein domains found in metazoan Glutamate Receptors – ANF, EGF, NCD3G, and Cache – are indicated in bold. See Supplementary File 10 for a list of all Glutamate Receptors screened in this analysis.

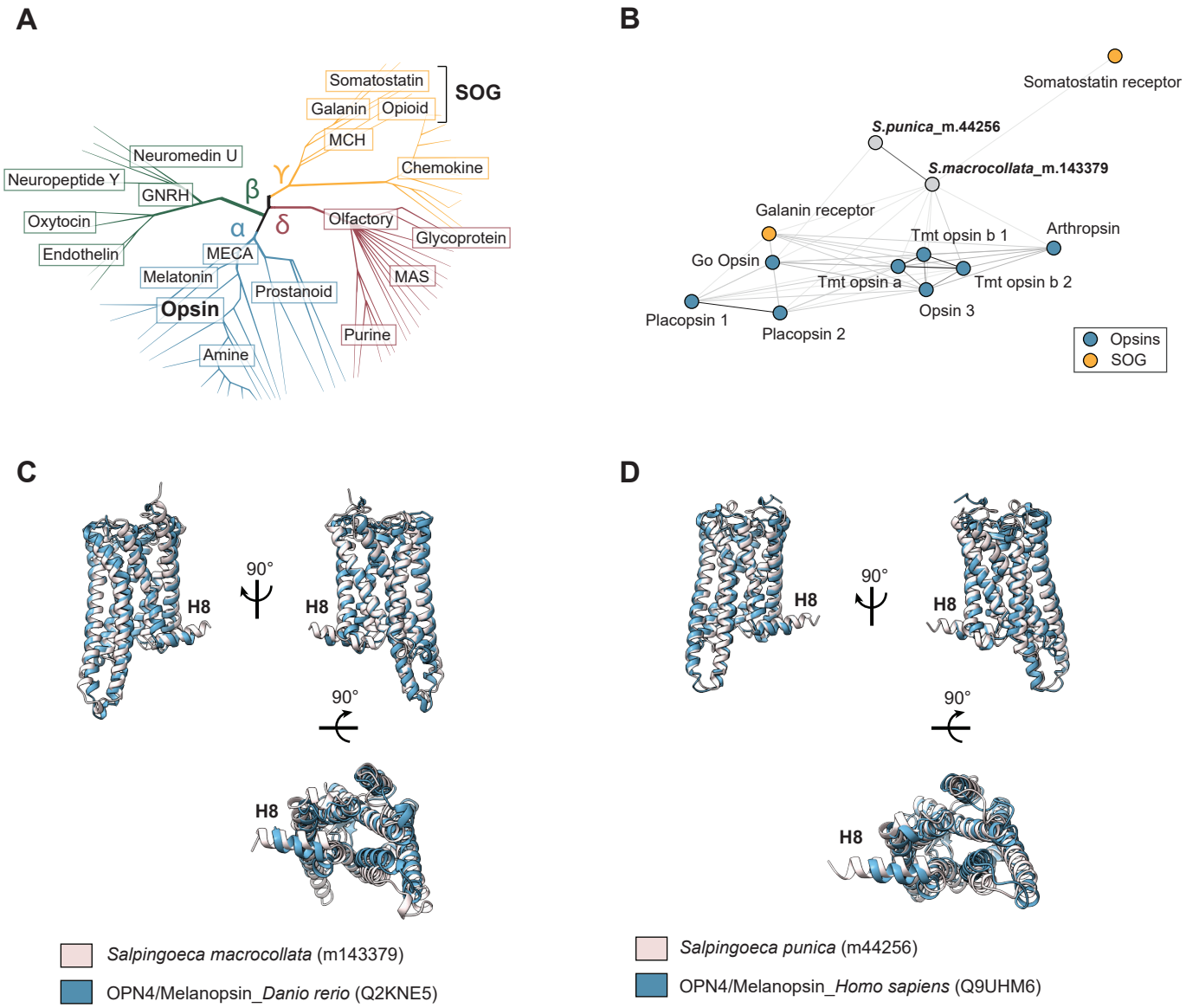

A
Somatostatin
Galanin
Opioid
SOG
MCH
Chemokine
Neuromedin U
Neuropeptide Y
GNRH
β
γ
Oxytocin
α
δ
Olfactory
Glycoprotein
Endothelin
MECA
Melatonin
Prostanoid
MAS
Opsin
Purine
Amine
B
Somatostatin receptor
S.punica_m.44256
S.macrocollata_m.143379
Galanin receptor
Arthropsin
Tmt opsin b 1
Go Opsin
Tmt opsin a
Tmt opsin b 2
Opsin 3
Placopsin 1
Placopsin 2
Opsins
SOG
C
H8
90°
H8
90°
H8
Salpingoeca macrocollata (m143379)
OPN4/Melanopsin_Danio rerio (Q2KNE5)
D
H8
90°
H8
90°
H8
Salpingoeca punica (m44256)
OPN4/Melanopsin_Homo sapiens (Q9UHM6)
E
TM2
TM3
ECL2
TM6
TM7
H8
83 110 134 135 136 181 186 187 264 265 267 268 296 302 303 304 305 306 310 311 312 313 314
Metazoa
Q6U736_OPN5_Neuropsin_Homo_sapiens
P02699_RHO_Bos_taurus
Q9Z2B3_RGR-opsin_Mus_musculus
P51475_Pinopsin_Gallus_gallus
Q2KNE5_Melanopsin_Danio_rerio
Q95P33_Ci-opsin1_Ciona_intestinalis
O15974_G0-opsin_Mizuhopecten_yessoensis
ASM47594.1_Xenopsin_Leptochiton_asellus
P04950_Arthropod_r_opsin_Drosophila_melanogaster
A0A059NTC8_Cnidopsin_Tripedalia_cystophora
A0A369RPG6_Placopsin_Trichoplax_sp
K9LLF4_Ctenopsin_Mnemiopsis_leidyi
Aminergic R
Peptide R
Lipid R
opsins
metazoan rhodopsins
Choanoflagellata
m.143379_Salpingoeca_macrocollata
m.44256_Salpingoeca_punica
Ichthyosporea
evm1s17431_Pirum_gemmata
Holomycota
KNE54739.1_Allomyces_macrogynus
A0A075AY71_Rozella_allomycis
A0A137NZX0_Conidiobolus_coronatus
Binding to retinal
G protein signaling
Disulfide bridge
Counterion
E/DRY motif
CWxPY motif
NPxxY(x)5,6FR motif

Fig. 4

Figure 4: Sequence and structural similarities and differences of choanoflagellate Rhodopsins and metazoan opsins. (A) Rhodopsin diversity in metazoans; adapted from (Fredriksson et al. 2003; Cardoso et al. 2012; Lv et al. 2016). Rhodopsins cluster into four main groups – α (light blue), β (green), γ (yellow), and δ (maroon). The Somatostatin/Opiod/Galanin group of Rhodopsins is indicated as SOG. Opsins and SOG receptors are indicated in bold for reference. **(B)** Choanoflagellate Rhodopsins cluster with opsins and SOG receptors from metazoans. All-against-all pairwise comparison of the choanoflagellate Rhodopsins with the complete Rhodopsin repertoire of representative metazoans revealed that choanoflagellate Rhodopsins are most similar to opsins and SOG. Shown is the local sequence similarity network, with nodes indicating Rhodopsin sequences and lines indicating BLAST connections of p-value < $1e^{-12}$. Choanoflagellate Rhodopsins are shown in grey (*S.macrocollata*_m.143379 and *S.punica*_m.44256), metazoan opsins in blue ((Placopsin 1_*T.adherens* (A0A369S8C3), Placopsin 2_*Trichoplax* sp (XP_002114592.1), Tmt opsin a_*D.rerio* (A0A2R8Q4C0), Tmt opsin b 1_*M.mola* (ENSMMOP00000004804.1), Tmt opsin b 2 _*D.rerio* (A0A2R8Q4C0) Opsin_3 *A.platyrhynchos* (XP_012958902.3), Arthropsin 8_*D.pulex* (EFX84032.1), Go Opsin_*B.floridae* (DAC74052.1)), and metazoan SOG receptors in orange (Somatostatin receptor_*B.floridae* (XP_032829507.1), Galanin 1_*H.sapiens* (P47211)).The complete dataset of 6149 Rhodopsins used in this analysis is provided in FASTA format in Supplementary File 11. See Supplementary File 12 for the full analysis. **(C and D)** Protein structure predictions link choanoflagellate Rhodopsins to metazoan opsins. **(C)** The predicted structure of *S. macrocollata* Rhodopsin most closely matches that of opn4a/Melanopsin-A from *Danio rerio*. Shown is the predicted structural similarity of *S. macrocollata* Rhodopsin (m.143379; pink) with Foldseek top hit (E-value: $6.26e^{-13}$) opn4a/Melanopsin-A from *Danio rerio* (AF-Q2KNE5-F1-model_v4; blue). Low confidence regions (>70 pLDDT) were removed for clarity. Shown are views of the superimposed models from the plane of the membrane (top) and from the extracellular perspective (bottom). **(D)** The predicted structure of *S. punica* Rhodopsin most closely matches that of opn4/Melanopsin from humans. Predicted structural similarity of *S. punica* Rhodopsin (m.44256; pink) with Foldseek top hit (E-value: $8.34e^{-13}$) opn4/Melanopsin from humans (AF-Q9UHM6-F1-model_v4; blue). Low confidence regions (>70 pLDDT) were removed for clarity. Shown are views of the superimposed models from the plane of the membrane (top) and from the extracellular perspective (bottom). **(E)** Alignment showing the conservation of functionally important motifs in metazoan, choanoflagellate, ichthyosporean, and holomycotan Rhodopsins. The alignment includes diverse metazoan opsins, three consensus sequences of human non-opsin Rhodopsins (Aminergic, Peptide, and Lipid Rhodopsins; Supplementary File 14), the two choanoflagellate Rhodopsins identified in this study (highlighted in bold), one ichthyosporean Rhodopsin, and three representative Rhodopsins from holomycotans. Residues identified as being critical for Rhodospin protein structure and function are shown. These include: a conserved Aspartic acid (D) at position 83 in the transmembrane helix 2 (TM2); two conserved Cysteines (C; orange) at positions 110 and 187 that are involved in disulfide bond formation; and the conserved E/DRY, CWxPY, and NpxxY$(x)_{5,6}$FR motifs (green), located in TM3, TM6, and TM7/H8, respectively (Davies et al. 2010; Nagata and Inoue 2021). These three motifs are essential for G protein interaction and to control the activity of the Rhodopsins. In addition, residues that are specific to opsins are also depicted: Glutamic acid (E)181, which acts as a counterion to the protonated Schiff base (Davies et al. 2010; Hankins et al. 2014; Nagata and

Inoue 2021), Serine (S)186 in extra-cellular loop 2 (EL2), and the highly conserved Lysine (K) at position 296 (blue) in TM7, that is almost universally found across all metazoan opsins (Gühmann et al. 2022; McCulloch et al. 2023). Lys(K)296 is required for covalent binding to the 11-cis retinal chromophore (Devine et al. 2013). Notably, the two choanoflagellate Rhodopsins show a Lys296Ser (*Salpingoeca_macrocollata*_m.143379) and a Lys296Val (*Salpingoeca_punica*_m.44256) substitutions, suggesting that these Rhodopsins may not have light-responsive functions. Canonically conserved functional residues and positions follow bovine Rhodopsin numbering (Nathans and Hogness 1983). The consensus sequences of the three non-opsin subfamilies of human Rhodopsins (Aminergic R, Peptide R, and Lipid R) were downloaded from GPCRdb (https://gpcrdb.org/). (+) symbol is used when consensus sequences cannot be resolved at a given position (ambiguity). The 36 human Aminergic receptors, 76 human Peptide receptors, and 36 human Lipid receptors aligned to build these consensus sequences are provided in Supplementary File 14.

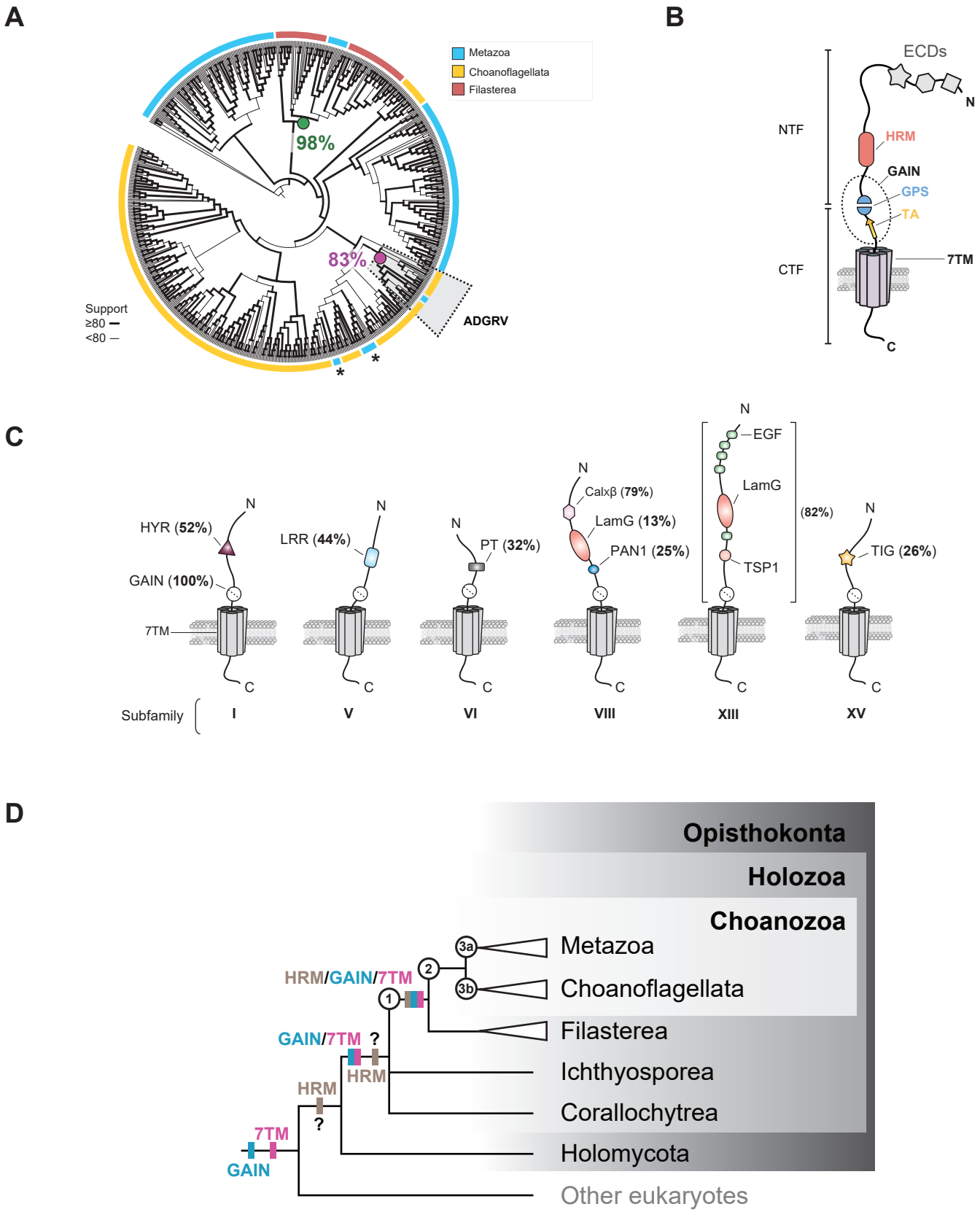

A
Metazoa
Choanoflagellata
Filasterea
98%
83%
ADGRV
Support
≥80
<80
B
ECDs
N
NTF
HRM
GAIN
GPS
TA
CTF
7TM
C
C
N
HYR (52%)
GAIN (100%)
7TM
C
N
LRR (44%)
C
N
PT (32%)
C
N
Calxβ (79%)
LamG (13%)
PAN1 (25%)
C
N
EGF
LamG
TSP1
(82%)
C
N
TIG (26%)
C
Subfamily
I
V
VI
VIII
XIII
XV
D
Opisthokonta
Holozoa
Choanozoa
Metazoa
Choanoflagellata
Filasterea
Ichthyosporea
Corallochytrea
Holomycota
Other eukaryotes
HRM/GAIN/7TM
GAIN/7TM
HRM
HRM
7TM
GAIN
1
2
3a
3b
ADGRV family
Laminin_G, EGF, LRR, FN3, Furin-like, Calxβ, PQQ, RLD
Lectin_C, Cadherin, EGF_CA, Somatomedin_B, Sushi, F5_F8_C, VWF, TSP1, PAN, Ephrin_rec-like, Collagen, Kringle, Gal_lectin
Ig, Ldl_recept_a, EAR, SEA, DSL, TILa, SRCR, Pentaxin, Epimerase
TIG, TNFR, HYR, Cohesin, PT

Fig. 5

Figure 5: **Evolution of aGPCR protein domain architecture. (A)** Although aGPCRs were
2025 present in LECA (**Fig. 2**), phylogenetic analysis of holozoan aGPCRs revealed that they diversified independently in metazoans and CMRs. Most metazoan (blue), choanoflagellate (orange), and filasterean (red) aGPCRs formed distinct clades in our analysis, suggesting an absence of orthologous relationships between the 7TM region of the aGPCRs from these three clades. A notable exception was the metazoan **AD**hesion **G** protein-coupled **R**eceptor **V**
2030 (ADGRV) GPCRs that grouped with choanoflagellate 7TM sequences (dotted grey box; 83% bootstrap support for ancestral node, magenta circle), suggesting they are orthologous. In addition, we also observed that members of the metazoan ADhesion G protein-coupled Receptor A (ADGRA) subfamily (asterisks) tended to group with choanoflagellate aGPCRs but either lacked reliable confidence value support or were not systematically recovered in all the
2035 inferred phylogenies. A subset of filasterean aGPCRs clustered with a set of uncharacterized cnidarian and cephalochordate receptors (bootstrap support 98%, green circle). This maximum-likelihood phylogenetic tree infers the evolutionary history of the 7TM domain of 329 choanoflagellate aGPCRs identified in this study, along with 76 filasterean aGPCRs and 253 metazoan aGPCRs. The 7TM sequence of a ciliate aGPCR
2040 (Stentor_coeruleus_OMJ80129.1_19670) was also included in the analysis and used as an outgroup to root the tree. The width of branches scales with UFboot support for the ancestral node. Branch lengths do not scale with evolutionary distance in this rendering. All the 7TM sequences used in this phylogenetic reconstruction are found in Supplementary File 17, and the fully annotated version of this phylogenetic tree, including bootstrap values, branch lengths, and
2045 all species names, is found in Supplementary File 18. **(B)** Protein domain architecture of metazoan aGPCRs. Like all GPCRs, aGPCRs contain a 7TM domain (represented by seven barrels) that anchors the protein in a lipid bilayer. What separates aGPCRs from other GPCRs is their possession of a diagnostic autoproteolytic GAIN domain (dotted oval) containing a proteolysis site (GPS, blue) and a hydrophobic tethered agonist peptide (TA, orange). The
2050 cleavage site in the GPS represents the boundary between the N-terminal fragment (NTF) and the C-terminal fragment (CTF). Many, but not all, aGPCRs also contain an HRM domain (salmon) within a few hundred amino acids of the GPS (**Fig. S10C**) (Prömel et al. 2012b; Araç D, Sträter N 2016)In metazoans, the NTF often contains a diversity of additional extracellular protein domains (ECDs, represented by grey star, hexagon, and square) that likely contribute to
2055 the diversity of ligands bound by aGPCRs (Araç and Leon 2019; Knierim et al. 2019). (**C**) Diversity and conservation of protein domains in the NTFs of different choanoflagellate aGPCR families. Like metazoan aGPCRs, nearly all choanoflagellate aGPCRs contain a GAIN domain with conserved GPS and TA motif (**Fig. S9**). Many aGPCR subfamilies identified through their 7TM were also distinguishable through their characteristic combinations of N-terminal protein
2060 domains (**Fig. S7A**). Shown here are six subfamilies – I, V, VI, VIII, XIII, and XV – whose members frequently share one or more protein domains. While subfamilies I, V, VIII, XIII, and XIV are seemingly unique to choanoflagellates, subfamily VIII is notable because the full-length proteins (both the NTFs and CTFs) are homologous to those of the ADGRV subfamily in metazoans (**Fig. 5A** and **Fig. S8**). Percentages indicate the number of members within a given
2065 subfamily possessing a conserved protein domain (subfamilies I,V,VI,VIII, and XV) or a conserved combination of protein domains (subfamily XIII). All the choanoflagellate aGPCR NTF sequences used in these analyses are found in Supplementary File 19 **(D)** Hypothesized

evolution of the HRM/GAIN/7TM module and additional ECDs in aGPCRs. For four nodes on the eukaryotic tree of life – (1) the last common ancestor of filastereans and choanozoans, (2) stem choanozoans, (3a) stem metazoans, and (3b) stem choanoflagellates – we reconstructed the phylogenetic distribution of diverse protein domains in the NTFs of aGPCRs. Most aGPCRs in metazoans, choanoflagellates, and filastereans contain a protein domain module composed of a GAIN domain and a 7TM. Additionally, the HRM/GAIN/7TM module, which is less common than the GAIN/7TM module in aGPCRs, is nonetheless conserved across most metazoan, choanoflagellate, and filasterean lineages. The linkage of the GAIN and 7TM domains likely occurred in stem holozoans, as today the two domains are restricted to the aGPCRs of extant holozoans, although the two domains are encoded in other genes in diverse non-holozoans (Araç et al. 2012; Krishnan et al. 2012; De Mendoza et al. 2014), suggesting that they originated in earlier branches of the eukaryotic tree. The linkage of the HRM domain with the GAIN/7TM module likely occurred in the last common ancestor of filastereans and choanozoans. The HRM domain is largely restricted to holozoans, although we found six non-aGPCR proteins containing HRM domains in the proteome of the nucleariid *Fonticula alba* (Fig. S10D). Therefore, HRMs either evolved in stem holozoans and were incorporated into the nucleariids by horizontal gene transfer, they are homologous in the two lineages and were lost from most holomycotans, or the similarities of HRMs in *F. alba* and holozoans are the result of convergent evolution. The repertoires of ECDs inferred for each node – (1), (2), (3a), and (3b) – are shown in grey boxes labeled accordingly. Finally, ADGVR likely evolved in the last common ancestor of filastereans and choanozoans.

Table 1. List of 18 GPCR clusters identified in choanoflagellates.

	GPCR family	Additional domain(s)	Biological function(s)	Subcellular localization	Natural ligand(s)	References
GRAFS	Rhodopsin	LRR, LDLa	light sensing, chemosensing, thermotaxis, energy metabolism, taste ✖ ★	plasma membrane (outer segment, primary cilia), disc membranes, vacuoles ✖ ★	light/retinal, hormones (GnRH, oestrogen, melanocortin,...), neurotransmitter (dopamine,acetylcholine,...), lipids ✖ ★	(Pérez-Cerezales et al, 2015) (Chadha et al, 2021) (Karthikeyan et al, 2023) (Roy et al, 2020) (Cardoso et al, 2012)
GRAFS	Adhesion	GPS, EGF, LamG, Calx-beta, HRM, TSP1, EGF_CA, LRR, FN3, Kringle, CUB, MANEC, OLF, SEA, PT, etc	cell adhesion, PCP, cell migration, tissue morphogenesis, neurodevelopment, etc ✖	plasma membrane (adherens junction, FAs, dendrite, primary cilia, stereocilia), spindle, nucleus, ER ✖	proteins of the extracellular matrix (glycosaminoglycans, collagens, laminin, …), other membrane proteins (FLRT, neurexins, teneurins,...) ✖	(Kusuluri et al, 2021) (Sakurai et al, 2022) (Yona et al, 2008) (Langenhan et al, 2013) (Hamann et al, 2015) (Araç and Leon, 2019)
GRAFS	Glutamate	ANF, NCD3G, EGF, Cache	synaptic transmission and neuronal excitability, cell proliferation, cell migration, phototaxis, calcium homeostasis, taste ✖	plasma membrane (dendrite) ✖	L-Glutamate, GABA, Glycine, sugars, peptide pheromones, ... ✖	(Elliott and Leys, 2010) (Crupi et al, 2019) (Niswender and Conn, 2010) (Wong et al, 2022) (Sigel and Steinmann, 2012) (Qian et al, 2023) (Dubovski et al, 2022) (Silva and Antunes, 2017) (Laboute et al, 2023)
	cAMP	none	cell chemotaxis, cell aggregation, cell differentiation, cyst formation ★	plasma membrane ★	cAMP ★	(Kawabe et al, 2009) (Louis et al, 1994) (Saxe et al, 1993) (Klein et al, 1988)
	GPR137	none	cell proliferation and apoptosis, neuronal differentiation ✖	lysosome ✖	unknown	(Gan et al, 2019) (Men et al, 2018)
	GPR155	PIN-like transporter, DEP	nutrient sensing, energy metabolism, tissue patering ✖	lysosome ✖	cholesterol ✖	(Shin et al, 2022) (Bayly-Jones et al, 2023) (Schöneberg, 2024) (Wang et al, 2018)
	GPR157	none	neuronal differentiation ✖	plasma membrane (primary cilia) ✖	unknown	(Takeo et al, 2016)
	GPCR-TKL/K	Tyrosine kinase-like / kinase	unknown	unknown	unknown	this study
GOST	TMEM145	GOLD	structural integrity of hair cell stereocilia ✖	plasma membrane (stereocilia) ✖	unknown	(Roh et al, 2025)
GOST	TMEM87	GOLD	protein trafficking, Golgi-Ph maintenance ✖	Golgi ✖	unknown	(Hoel et al, 2022) (Hirata et al, 2015) (Shin et al, 2020)
GOST	GPR107/108	GOLD	protein trafficking, recycling of receptors, immunity ✖	Golgi, nucleus, clathrin vesicles ✖	neurostatin, gambogic acid ✖	(Yang et al, 2024) (Yosten et al, 2012) (Zhou et al, 2014) (Lyu et al, 2022) (Balazova et al, 2021)
GOST	GPR180	GOLD	gametogenesis ★, lipid metabolism ✖	plasma membrane, vesicles ✖	L-lactate, CTHRC1 ✖	(Yoshida et al, 2023) (Wang et al, 2022) (Mosienko et al, 2018)
GOST	Hi-GOLD	GOLD	unknown	unknown	unknown	this study
	GPCR PIPK	PIPK	cell density sensing, bacterial defense, phagocytosis, asexual development, chemotaxis ★	endosomal vesicles, phagosome ★	unknown	(Van Den Hoogen et al, 2018) (Bakthavatsalam et al, 2006) (Bakthavatsalam et al, 2007) (Hua et al, 2013) (Riyahi et al, 2011)
	RSF	none	unknown	unknown	unknown	this study
	GPRch1	none	unknown	unknown	unknown	this study
	GPRch2	none	unknown	unknown	unknown	this study
	GPRch3	none	unknown	unknown	unknown	this study

✖ Metazoa ▲ Embryophyta ★ Protozoa

Table 2: Taxon sampling and data sources for genomes and transcriptomes analyzed in this study

Group	Species	Source
Choanoflagellate Proteomes	*Diaphanoeca grandis, Acanthoeca spectabilis, Salpingoeca punica, Salpingoeca urceolata, Hartaetosiga gracilis, Salpingoeca macrocollata, Stephanoeca diplocostata, Helgoeca nana, Salpingoeca kvevrii, Didymoeca costata, Choanoeca perplexa, Salpingoeca infusionum, Microstomoeca roanoka, Savillea parva, Hartaetosiga balthica Salpingoeca dolichothecata, Codosiga hollandica, Salpingoeca helianthica, Mylnosiga fluctuans, Salpingoeca rosetta, Monosiga brevicollis, Barroeca monosierra, Choanoeca flexa*	NCBI BioProject PRJNA419411, Figshare DOI: 10.6084/m9.figshare.5686984.v2 (Richter et al. 2018), NCBI BioProject PRJNA37927 (Fairclough et al. 2013) ,GenBank GCA_000002865.1 (King et al. 2008), Figshare DOI: 10.6084/m9.figshare.8216291 (Brunet et al. 2019), *Barroeca monosierra* genome-derived proteome (Hake et al. 2024), EukProt: https://evocellbio.com/eukprot/
Filasterean Proteomes	*Pigoraptor vietnamica, Pigoraptor chileana, Capsaspora owczarzaki, Ministeria vibrans*	MulticellGenome: https://multicellgenome.com/genomes
Ichthyosporean Proteomes	*Sphaeroforma arctica, Creolimax fragrantissima, Ichthyophonus hoferi, Abeoforma whisleri, Pirum gemmata, Chromosphaera perkinsii*	MulticellGenome: https://multicellgenome.com/genomes
Corallochytrean Proteomes	*Syssomonas multiformis, Corallochytrium limacisporum*	MulticellGenome: https://multicellgenome.com/genomes ; https://doi.org/10.5061/dryad.26bv4
Metazoan Proteomes	*Trichoplax adhaerens, Amphimedon queenslandica, Corticium candelabrum, Dysidea avara, Ephydatia muelleri, Halichondria panicea, Oscarella lobularis, Oopsacas minuta, Tethya wilhelma, Mnemiopsis leidyi, Bolinopsis microptera, Danio rerio, Gallus gallus, Drosophila melanogaster, Ciona intestinalis, Branchiostoma floridae,*	UniProt Proteomes: https://www.uniprot.org/

Table 2: Taxon sampling and data sources for genomes and transcriptomes analyzed in this study

Group	Species	Source
	Nematostella vectensis, Acropora millepora, Caenorhabditis elegans, Mus musculus, Homo sapiens	
Holomycota Proteomes	*Paraphelidium tribonemae, Neurospora crassa, Emericella nidulans, Ustilago maydis, Yarrowia lipolytica, Cryptococcus neoformans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Fonticula alba, Rozella allomycis, Allomyces macrogynus, Conidiobolus coronatus, Batrachochytrium dendrobatidis, Rhizophagus irregularis, Rhizopus delemar*	UniProt Proteomes https://www.uniprot.org/
Amoebozoan Proteomes	*Dictyostelium purpureum, Heterostelium pallidum, Entamoeba histolytica, Dictyostelium discoideum, Cavenderia fasciculata, Acanthamoeba castellanii, Planoprotostelium fungivorum, Polysphondylium violaceum, Dictyostelium firmibasis*	UniProt Proteomes: https://www.uniprot.org/
Viridiplantae Proteomes	*Volvox carteri, Chlamydomonas reinhardtii, Chlorella variabilis, Ostreococcus tauri, Gonium pectorale, Bathycoccus prasinos, Chlamydomonas eustigma, Raphidocelis subcapitata, Chloropicon primus, Chlamydomonas schloesseri, Pycnococcus provasolii, Volvox reticuliferus, Coccomyxa sp., Pleodorina starrii, Arabidopsis thaliana, Ricinus communis, Oryza sativa, Rosa chinensis, Vigna angularis, Rhododendron simsii, Papaver atlanticum, Cryptomeria japonica*	UniProt Proteomes: https://www.uniprot.org/

Table 2: Taxon sampling and data sources for genomes and transcriptomes analyzed in this study

Group	Species	Source
Ciliate Proteomes	*Paramecium tetraurelia, Ichthyophthirius multifiliis, Tetrahymena thermophila, Stylonychia lemnae, Pseudocohnilembus persalinus, Stentor coeruleus, Halteria grandinella, Euplotes crassus*	UniProt Proteomes: https://www.uniprot.org/

Supplementary materials for:

G protein-coupled receptor diversity and evolution in the closest living relatives of Metazoa

Alain Garcia De Las Bayonas[1,2]* and Nicole King[1,2]

1 Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720
2 Howard Hughes Medical Institute, United States
*Corresponding author. Email: alaingdlb@berkeley.edu

Supplemental Text

Conservation of the GPCR signaling pathway in choanoflagellates

In Metazoa, the binding of a ligand to the N-terminus or extracellular loops of a GPCR triggers a
2145 conformational change that initiates the rest of the GPCR signaling pathway (Hilger et al. 2018; Wootten et al. 2018; Hauser et al. 2021). Ligand-bound GPCRs act as guanine nucleotide exchange factors (GEFs) for the GDP-bound Gα subunit, catalyzing the exchange of GDP for GTP and, thereby, promoting the dissociation of activated Gα-GTP from Gβγ dimers (Fig. S3A, (Hilger et al. 2018; Weis and Kobilka 2018; Wootten et al. 2018)). Both activated Gα-GTP and
2150 free Gβγ dimers can transduce the signal independently to downstream effectors (Marinissen and Gutkind 2001; Neves et al. 2002; Dupré et al. 2009). Ric8 proteins and Phosducin facilitate the folding and targeting at the membrane of Gα subunit and Gβγ dimers, respectively (Willardson and Howlett 2007; Srivastava et al. 2019). Overstimulation of GPCR signaling is detrimental to the cell, and various mechanisms quickly turn off GPCR activity in Metazoa.
2155 Signal termination is controlled in part by the hydrolysis of Gα-GTP to Gα-GDP, which reconstitutes the inactive Gαβγ heterotrimer. The rate of Gα-GTP hydrolysis is a consequence of the intrinsic GTPase activity of Gα and can be enhanced by the action of regulators of G protein signaling (RGS), proteins that act to accelerate the GTPase activity (O'Brien et al. 2019). In addition, GoLoco motif proteins are negative regulators of Gα-GTP signaling, as they slow the
2160 spontaneous release of GDP from the Gα subunit (Willard et al. 2004; Sato et al. 2006). Another major axis to terminate GPCR signaling involves the rapid phosphorylation of the ligand-bound receptor by GPCR-Kinases (GRKs) and the subsequent binding of Arrestin proteins that selectively recognize active phosphorylated receptors (Rajagopal and Shenoy 2018; Gurevich and Gurevich 2019). Arrestins not only abolish G protein-mediated signaling but also act as a
2165 scaffold to facilitate multiple downstream signaling pathways and trigger the endocytosis of the receptor (Fig. S3A, (Bagnato and Rosanò 2019)).

We found that the heterotrimeric G proteins Gα, Gβ, and Gγ, along with positive regulators (Ric8, Phosducin) and negative regulators (GRK, Arrestin, RGS, GoLoco) of GPCR signaling, were detected in most choanoflagellates, with some exceptions in a few species (Fig. S3B,
2170 Supplementary File 3). Notably, we failed to detect Gγ and GRK in nearly half of the choanoflagellates, which could either suggest real losses, false negatives, or a divergence beyond recognition of these genes in the species concerned.

We decided to assess further the diversity of Gα subunits, the best-characterized transducers of metazoan GPCR signaling, in choanoflagellates. In metazoans, Gα subunits are classified into
2175 five main families (Gαs, Gαi/o, Gαq/11, Gα12/13, and Gαv), and members of each family interact with different effectors to produce distinct cellular responses (Fig. S3C, (Marinissen and Gutkind 2001; Neves et al. 2002; Oka et al. 2009; Oka and Korsching 2009; Doktorgrades and Obaid 2022; Zhang et al. 2024)). Although all major G protein Gα classes ($Gα_s$, $Gα_{q/11}$, $Gα_{12/13}$, $Gα_v$, and $Gα_{i/o}$) likely originated prior to the diversification of metazoans from the rest of
2180 holozoans (De Mendoza et al. 2014; Krishnan 2015; Lokits et al. 2018), our phylogenetic analysis recovered only three Gα classes in choanoflagellates ($Gα_s$, $Gα_{q/11}$, and $Gα_v$) (Fig. S3D

and Supplementary Files 4 and 5). Interestingly, $G\alpha_s$ subunits, previously thought to be lost in choanoflagellates, were detected in four closely related species (*S. kvevrii*, *S. urceolata*, *S. macrocollata*, *S. punica*). Based on their functions in metazoans, the activation of these Gα subunits could modulate protein phosphorylation (Gαs and Gαq/11), calcium signaling (Gαq/11), and ion homeostasis (Gαv) downstream of the GPCRs in choanoflagellates (Fig. S3C and S3D) (Marinissen and Gutkind 2001; Neves et al. 2002; Abu Obaid et al. 2024).

Overall, our results suggest that the GPCR signaling pathway is conserved in most choanoflagellate species, providing additional evidence that metazoan-like G protein signaling is part of the repertoire of signaling activities in choanoflagellates (De Mendoza et al. 2014; Krishnan et al. 2015; Lokits et al. 2018).

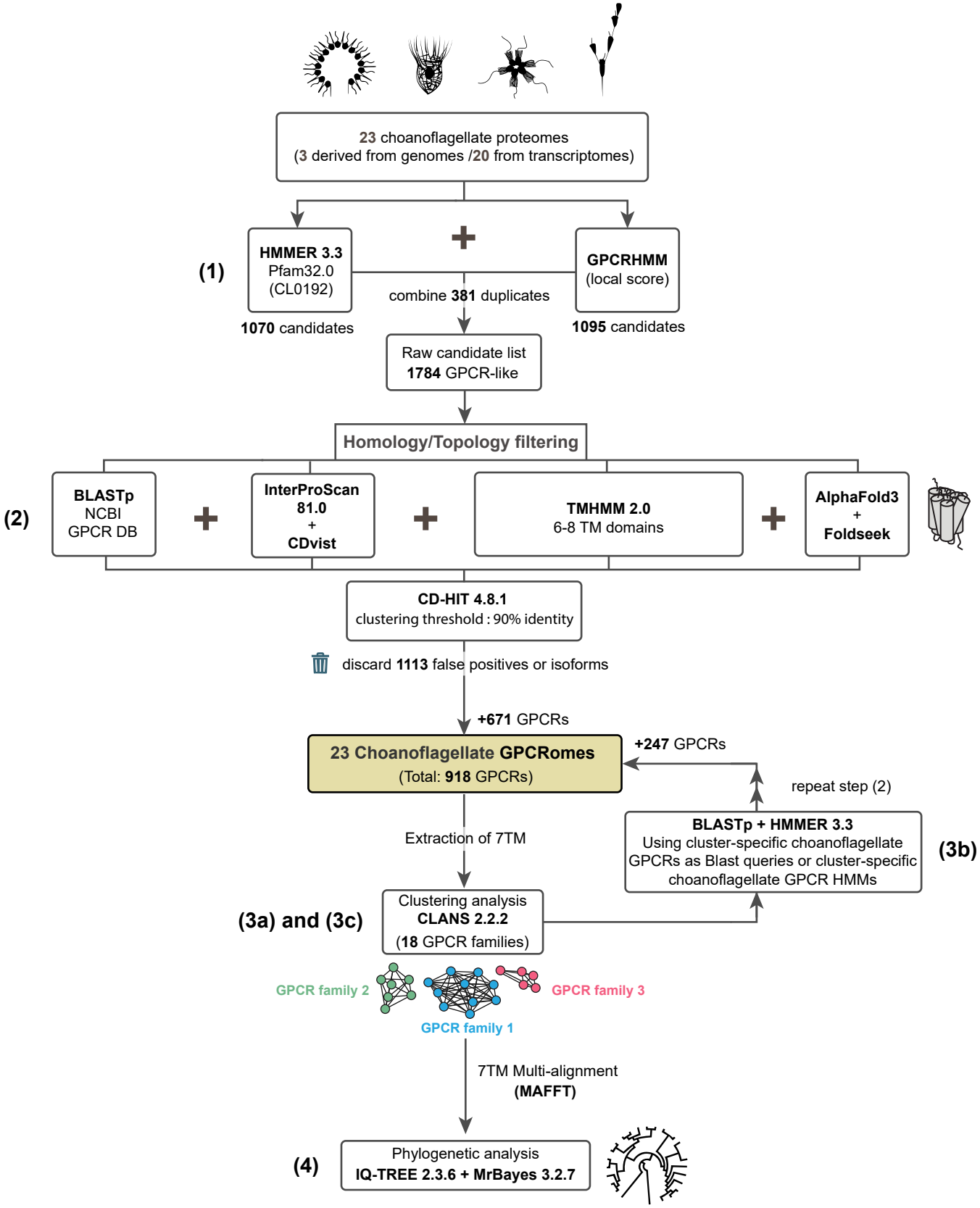

23 choanoflagellate proteomes
(3 derived from genomes /20 from transcriptomes)
(1)
HMMER 3.3
Pfam32.0
(CL0192)
GPCRHMM
(local score)
combine 381 duplicates
1070 candidates
1095 candidates
Raw candidate list
1784 GPCR-like
Homology/Topology filtering
(2)
BLASTp
NCBI
GPCR DB
InterProScan
81.0
+
CDvist
TMHMM 2.0
6-8 TM domains
AlphaFold3
+
Foldseek
CD-HIT 4.8.1
clustering threshold : 90% identity
discard 1113 false positives or isoforms
+671 GPCRs
23 Choanoflagellate GPCRomes
(Total: 918 GPCRs)
+247 GPCRs
repeat step (2)
BLASTp + HMMER 3.3
Using cluster-specific choanoflagellate GPCRs as Blast queries or cluster-specific choanoflagellate GPCR HMMs
(3b)
Extraction of 7TM
(3a) and (3c)
Clustering analysis
CLANS 2.2.2
(18 GPCR families)
GPCR family 2
GPCR family 3
GPCR family 1
7TM Multi-alignment
(MAFFT)
(4)
Phylogenetic analysis
IQ-TREE 2.3.6 + MrBayes 3.2.7

Fig. S1

2210 **Figure S1: Pipeline for identifying GPCRs in choanoflagellate genomes and transcriptomes. (1)** GPCR candidates were searched in the genome- and transcriptome-derived proteomes of 23 choanoflagellate species using Hidden Markov model(s) (HMMs) unique to 54 GPCR families in eukaryotes (GPCR_A Pfam clan CL0192) or a global HMM profile that matches the common topology of GPCRs irrespective or their families (GPCRHMM).
2215 The comparison of the two sets of candidate GPCRs obtained through these two independent approaches revealed 381 duplicate candidates, which were merged, leaving a final set of 1784 candidates. **(2)** The 1784 candidate choanoflagellate GPCRs were then subjected to a validation step where false negatives and highly truncated sequences were filtered out using the search tools BLASTp, Interproscan, CDvist, and the transmembrane domain predictor TMHMM
2220 (see Methods). Alphafold-predicted structures of the GPCR candidates and subsequent structure homology search complemented the analysis. Finally, we used CD-HIT to discard isoforms. A total of 1113 sequences were removed from our dataset, leaving 671 validated choanoflagellate GPCRs. **(3a, 3c)** To assess the diversity of choanoflagellate GPCRs, we then clustered the 671 GPCRs based on all-against-all pairwise sequence similarities using the
2225 clustering tool CLANS. Most GPCRs (629 receptors) were sorted into 18 clusters, with a minority of them (76 receptors) sharing no sequence similarities with any other GPCRs in the dataset. The same protocol (3a) was re-iterated in (3c), after refining the choanoflagellate GPCR dataset (3b) and including diverse metazoan, amoebozoan, stramenopile, alveolate, and chlorophyte GPCRs to the analysis. **(3b)** Because the HMMs used in our first round of
2230 screening were mostly based on seed alignments of metazoan sequences, we built choanoflagellate cluster-specific GPCR HMMs for each of the 18 choanoflagellate GPCR clusters previously identified (3a) to detect possible additional choanoflagellate GPCRs (Supplementary File 2). We also built individual HMMs for the GPCRs that did not fall into the 18 clusters. These HMMs were used to search the 23 choanoflagellate proteomes again. In
2235 parallel, we also selected choanoflagellate GPCR sequences from each GPCR cluster or single GPCR sequences in the case of GPCRs that did not group with other GPCRs, to use as BLAST queries against the 23 choanoflagellate proteomes. The resulting candidates were run through the filtration process in step (2), leaving 247 new GPCRs that were not predicted with the metazoan-biased HMMs. Added to the original 671, this gave a total of 918 GPCRs from the 23
2240 choanoflagellate species (Supplementary File 1). **(4)** Phylogenetic trees were inferred for various choanoflagellate GPCR families, including aGPCRs and Glutamate Receptors, using IQ-TREE and MrBayes.

2245

2250

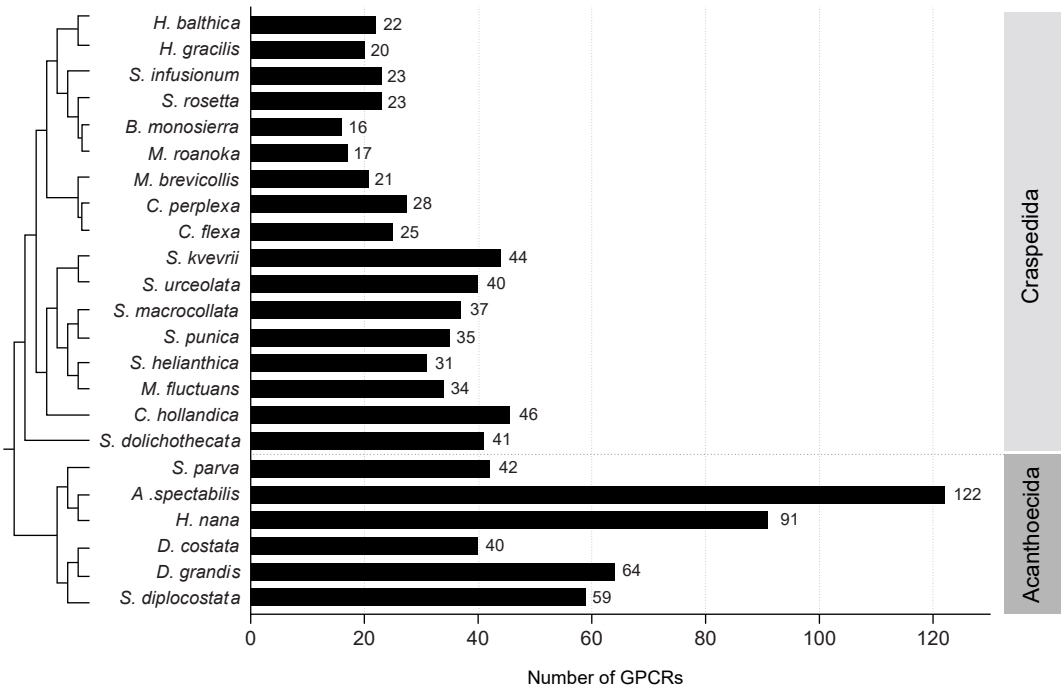

H. balthica
H. gracilis
S. infusionum
S. rosetta
B. monosierra
M. roanoka
M. brevicollis
C. perplexa
C. flexa
S. kvevrii
S. urceolata
S. macrocollata
S. punica
S. helianthica
M. fluctuans
C. hollandica
S. dolichothecata
S. parva
A .spectabilis
H. nana
D. costata
D. grandis
S. diplocostata
22
20
23
23
16
17
21
28
25
44
40
37
35
31
34
46
41
42
122
91
40
64
59
0
20
40
60
80
100
120
Number of GPCRs
Craspedida
Acanthoecida

Fig. S2

Figure S2: Total number of GPCRs across choanoflagellates varies by species and seems to correlate with phylogenetic affiliation. GPCR numbers tend to be higher in acanthoecids than in craspedids. Shown are GPCR numbers per species, mapped to a consensus phylogeny adapted from (Carr et al. 2017b; Ginés-Rivas and Carr 2025b). Choanoflagellate GPCR numbers range from 16 in *B. monosierra* to 122 in *A. spectabilis.* These are approximate numbers, as the transcriptomes may be incomplete and/or the predicted proteomes may include splice variants.

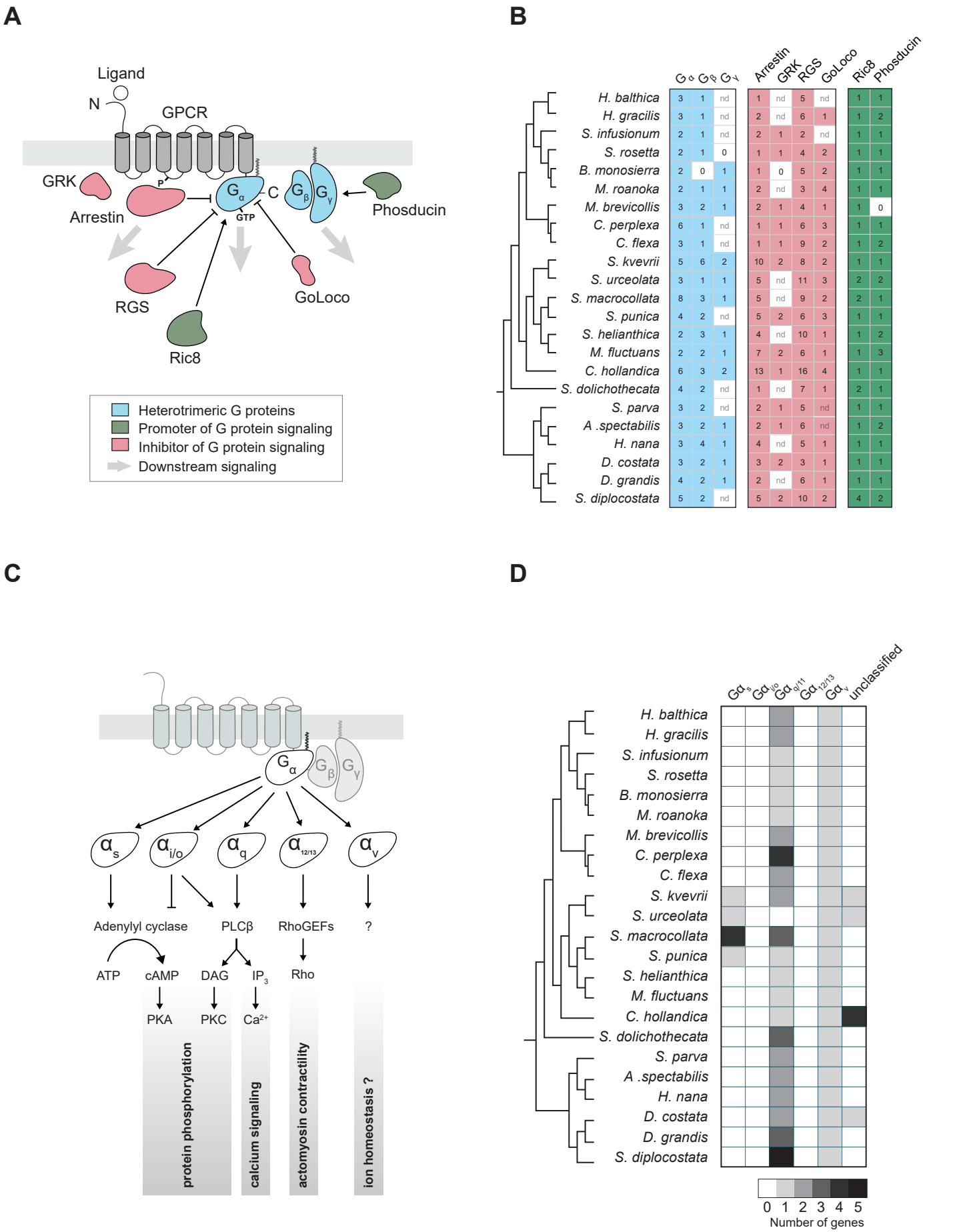

A
Ligand
N
GPCR
GRK
P
Arrestin
Gα
C
GTP
Gβ
Gγ
Phosducin
RGS
GoLoco
Ric8
Heterotrimeric G proteins
Promoter of G protein signaling
Inhibitor of G protein signaling
Downstream signaling
B
Gα Gβ Gγ Arrestin GRK RGS GoLoco Ric8 Phosducin
H. balthica 3 1 nd 1 nd 5 nd 1 1
H. gracilis 3 1 nd 2 nd 6 1 1 2
S. infusionum 2 1 nd 2 1 2 nd 1 1
S. rosetta 2 1 0 1 1 4 2 1 1
B. monosierra 2 0 1 1 0 5 2 1 1
M. roanoka 2 1 1 2 nd 3 4 1 1
M. brevicollis 3 2 1 2 1 4 1 1 0
C. perplexa 6 1 nd 1 1 6 3 1 1
C. flexa 3 1 nd 1 1 9 2 1 2
S. kvevrii 5 6 2 10 2 8 2 1 1
S. urceolata 3 1 1 5 nd 11 3 2 2
S. macrocollata 8 3 1 5 nd 9 2 2 1
S. punica 4 2 nd 5 2 6 3 1 1
S. helianthica 2 3 1 4 nd 10 1 1 2
M. fluctuans 2 2 1 7 2 6 1 1 3
C. hollandica 6 3 2 13 1 16 4 1 1
S. dolichothecata 4 2 nd 1 nd 7 1 2 1
S. parva 3 2 nd 2 1 5 nd 1 1
A .spectabilis 3 2 1 2 1 6 nd 1 2
H. nana 3 4 1 4 nd 5 1 1 1
D. costata 3 2 1 3 2 3 1 1 1
D. grandis 4 2 1 2 nd 6 1 1 1
S. diplocostata 5 2 nd 5 2 10 2 4 2
C
Gα
Gβ
Gγ
αs
αi/o
αq
α12/13
αv
Adenylyl cyclase
PLCβ
RhoGEFs
?
ATP
cAMP
DAG
IP3
Rho
PKA
PKC
Ca2+
protein phosphorylation
calcium signaling
actomyosin contractility
ion homeostasis ?
D
Gαs Gαi/o Gαq/11 Gα12/13 Gαv unclassified
H. balthica
H. gracilis
S. infusionum
S. rosetta
B. monosierra
M. roanoka
M. brevicollis
C. perplexa
C. flexa
S. kvevrii
S. urceolata
S. macrocollata
S. punica
S. helianthica
M. fluctuans
C. hollandica
S. dolichothecata
S. parva
A .spectabilis
H. nana
D. costata
D. grandis
S. diplocostata
0 1 2 3 4 5
Number of genes

Fig. S3

Figure S3: **Conservation of metazoan GPCR transducers in choanoflagellates. (A)** Schematic representation of the core GPCR signaling pathway and associated regulators. Upon binding to a ligand, GPCRs activate downstream heterotrimeric G proteins (blue) that dissociate, giving free Gα and Gβγ subunits that transduce the signal independently from each other (grey arrows). Additional regulators ensure the fine-tuning of GPCR signaling by either promoting G protein signaling (green) or inhibiting it (red) (see Supplementary text for more information). While heterotrimeric G proteins are the canonical signal transducers downstream of GPCR activation, Arrestins act as important signaling hubs (Gurevich and Gurevich 2019). **(B)** Phylogenetic distribution of heterotrimeric G proteins (Gα, Gβ, and Gγ; blue), negative regulators of G protein signaling (Arrestin, GRK, RGS, and GoLoco; red) and positive regulators of G protein signaling (Ric8 and Phosducin; green) within choanoflagellates. The number of proteins recovered in our analysis that correspond to each of these different categories is specified in the matrix for each choanoflagellate species. The GPCR signaling pathway is conserved in most choanoflagellate species. nd = not detected in lineages for which only transcriptome data are available. For species with both transcriptome and genome data (*S. rosetta*, *B. monosierra*, and *M. brevicollis*) (King et al. 2008; Fairclough et al. 2013; Hake et al. 2024), failure to detect a GPCR subfamily member is indicated with a “0”. All the choanoflagellate sequences identified in each of these categories are provided in Supplementary File 3. **(C)** Schematic representation of major pathways activated by different Gα subunits. Gα proteins can be subdivided into five main families ($G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$, $G\alpha_{12/13}$, and $G\alpha_v$) with different signaling properties (Marinissen and Gutkind 2001; Oka and Korsching 2009; Feng et al. 2022; Liu et al. 2024). $G\alpha_s$ proteins activate Adenylyl cyclase, resulting in an accumulation of intracellular cAMP and activation of protein kinase A (PKA). In contrast, the activated $G\alpha_{i/o}$ inhibits Adenylyl cyclase. $G\alpha_{i/o}$ proteins also stimulate phospholipase C- β (PLCβ) that generates diacylglycerol (DAG) and inositol trisphosphate (IP_3), eventually regulating Ca2+ signal and protein kinase C (PKC) activity; a function shared with $G\alpha_{q/11}$ subunits. $G\alpha_{12/13}$ controls actomyosin contractility through the activation of the Rho/Rho kinase signaling pathway. Little is known about signal transduction downstream of $G\alpha_v$ which could control ion homeostasis (Abu Obaid et al. 2024). **(D)** Phylogenetic distribution of the five Gα subunit families within choanoflagellates. While $G\alpha_{q/11}$ and $G\alpha_v$ are shared by all 23 choanoflagellate species, $G\alpha_{i/o}$ and $G\alpha_{12/13}$ were not detected in our dataset. We identified $G\alpha_s$ subunits in a monophyletic clade encompassing *S. kvevrii*, *S. urceolata*, *S. macrocollata*, and *S. punica*. Failure to detect $G\alpha_s$ in *S. helianthica* and *M. fluctuans* could be due to false negatives or a secondary loss in the ancestor or these two sister species. Gα proteins that were not assigned to any of these five families were counted as unclassified. The assignment of choanoflagellate Gα proteins to one of the five Gα subunit families was based on the maximum likelihood phylogenetic tree inferred by the G protein alpha subunit from both choanoflagellates and known Gα metazoan subunits (Supplementary Files 4 and 5).

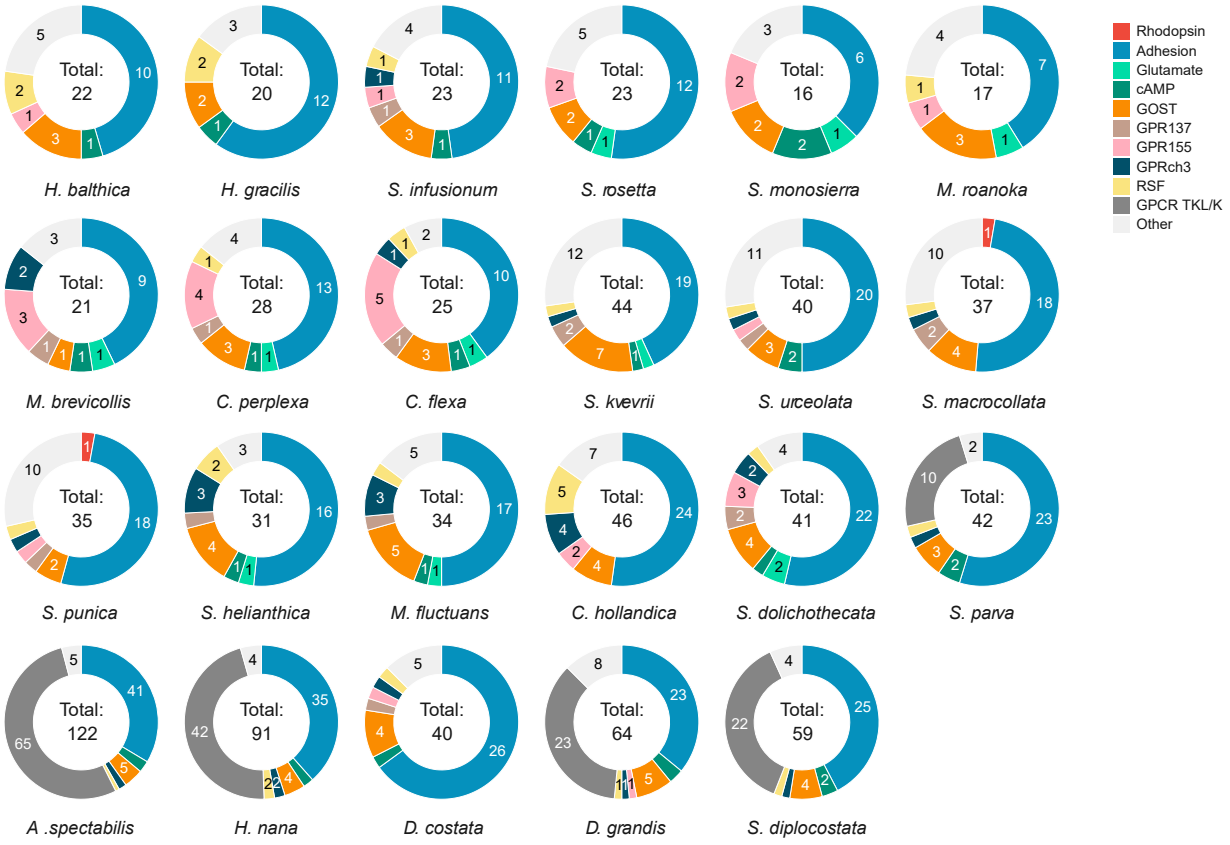

Rhodopsin
Adhesion
Glutamate
cAMP
GOST
GPR137
GPR155
GPRch3
RSF
GPCR TKL/K
Other
Total: 22
H. balthica
Total: 20
H. gracilis
Total: 23
S. infusionum
Total: 23
S. rosetta
Total: 16
S. monosierra
Total: 17
M. roanoka
Total: 21
M. brevicollis
Total: 28
C. perplexa
Total: 25
C. flexa
Total: 44
S. kvevrii
Total: 40
S. urceolata
Total: 37
S. macrocollata
Total: 35
S. punica
Total: 31
S. helianthica
Total: 34
M. fluctuans
Total: 46
C. hollandica
Total: 41
S. dolichothecata
Total: 42
S. parva
Total: 122
A .spectabilis
Total: 91
H. nana
Total: 40
D. costata
Total: 64
D. grandis
Total: 59
S. diplocostata

Fig. S4

Figure S4: aGPCRs dominate the GPCRome of most choanoflagellates analyzed. Donut charts show the number and proportion of genes in the main GPCR families – Rhodopsin, Adhesion, Glutamate, cAMP, GOST (including TMEM87, GPR107/108, TMEM145, GPR180, and Hi-GOLD), GPR137, GPR155, GPRch3, RSF, GPCR TKL/K and Other GPCRs – for each of the 23 choanoflagellate species analyzed. Adhesion receptors represent, on average, 47% of the GPCRs encoded in these choanoflagellate species. The presence of a large repertoire of GPCR TKL/Ks in acanthoecids contributes, in part, to the larger size of their GPCRomes. The total number of GPCRs detected is shown in the center.

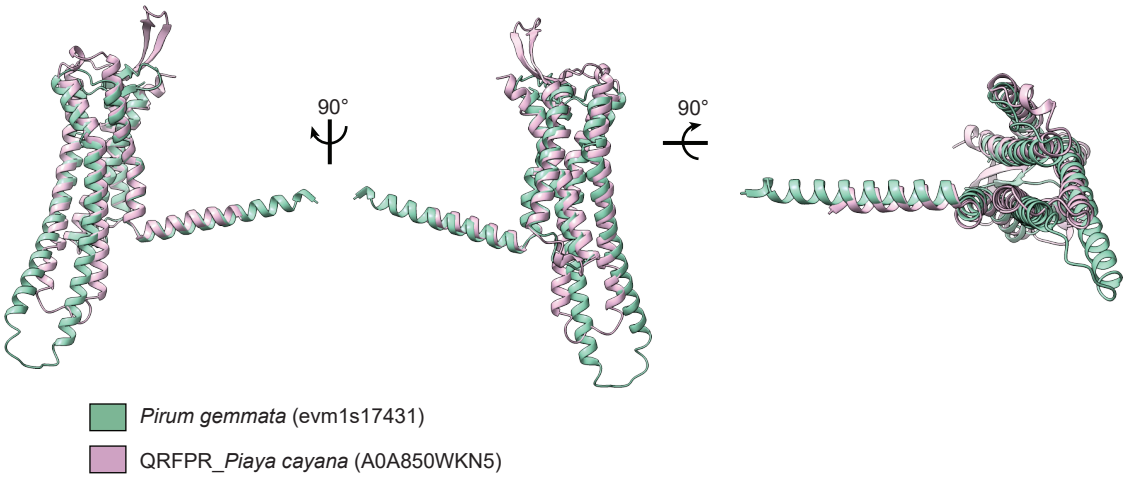

90°
90°
Pirum gemmata (evm1s17431)
QRFPR_*Piaya cayana* (A0A850WKN5)

Fig. S5

Figure S5: Ichthyosporean Rhodopsin shares structural similarities with metazoan peptide receptor Rhodopsin. Predicted structural homology of model *Pirum_gemmata*_evm1s17431 (green) with Foldseek top hit (E-value: $4.92e^{-9}$) Pyroglutamylated RFamide peptide receptor (QRFPR)_Piaya_cayana (AF-A0A850WKN5; pink). A cytoplasmic helix 8 (H8), found in most metazoan Rhodopsins, is also predicted in the ichthyosporean rhodopsin. Low confidence regions (>70 pLDDT) were removed for clarity in both models. View of the superimposed models from the plane of the membrane (left and center) and the top (right).

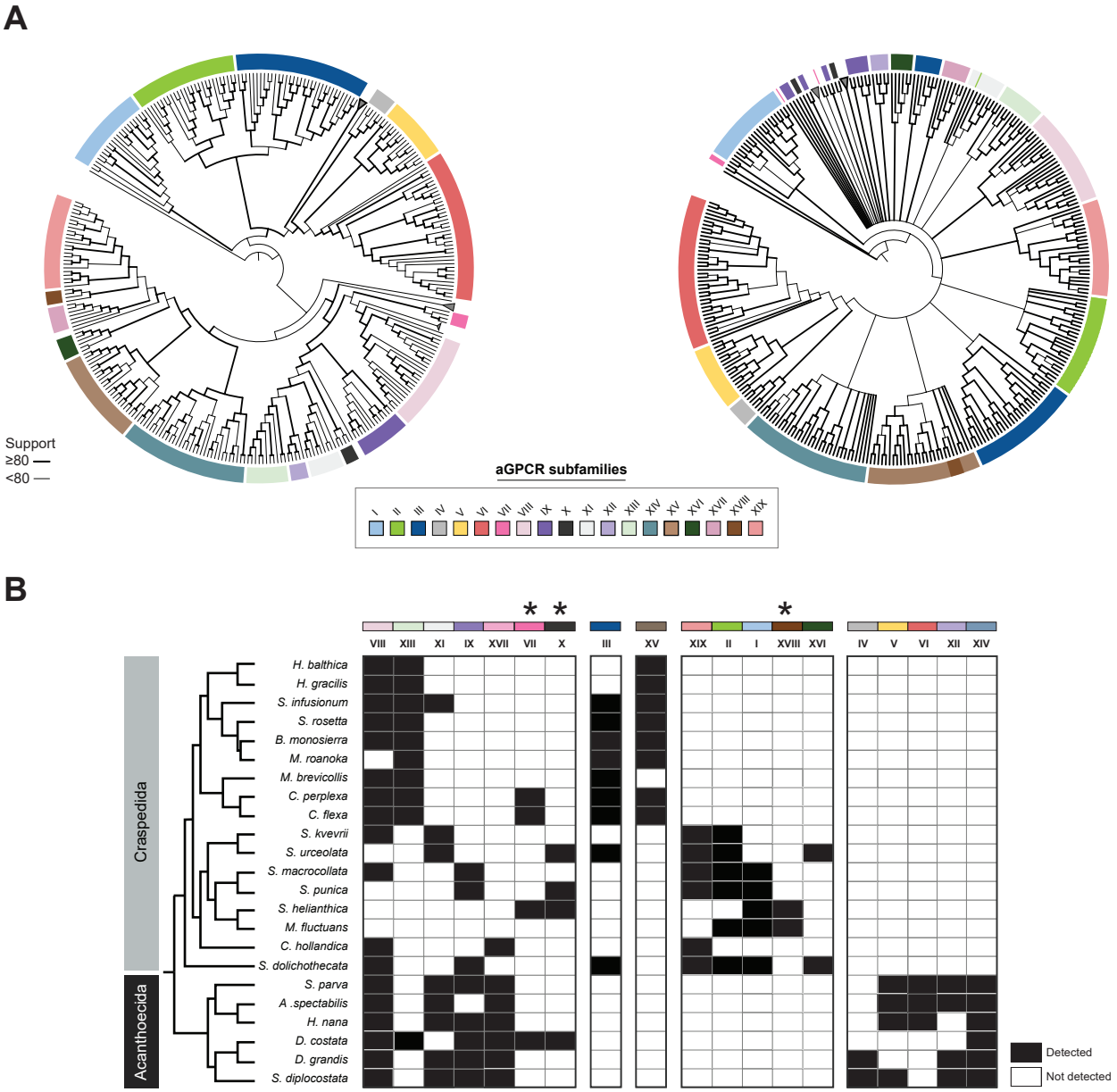

A
Support
≥80
<80
aGPCR subfamilies
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
XVII
XVIII
XIX
B
VIII
XIII
XI
IX
XVII
VII
X
III
XV
XIX
II
I
XVIII
XVI
IV
V
VI
XII
XIV
Craspedida
Acanthoecida
H. balthica
H. gracilis
S. infusionum
S. rosetta
B. monosierra
M. roanoka
M. brevicollis
C. perplexa
C. flexa
S. kvevrii
S. urceolata
S. macrocollata
S. punica
S. helianthica
M. fluctuans
C. hollandica
S. dolichothecata
S. parva
A .spectabilis
H. nana
D. costata
D. grandis
S. diplocostata
Detected
Not detected

Fig. S6

Figure S6: Diversity of aGPCR subfamilies in choanoflagellates. (A) Phylogenetic analyses of the 7TMs of choanoflagellate aGPCRs reveal the presence of at least 19 subfamilies, labeled I-XIX. Maximum-likelihood inference (left tree) and Bayesian inference (right tree) include full 7TM sequences of 229 choanoflagellate aGPCRs along with a 7TM sequence of a ciliate aGPCR (Stentor_coeruleus_OMJ80129.1_SteCoe_19670), used as an outgroup to root the trees. The 7TM sequences were aligned with MAFFT V7.463 using the E-INS-I algorithm, and phylogenies were built using either IQ-TREE or MrBayes v3.2.7a (see Methods). The two phylogenies cross-verified the grouping of most aGPCRs into the same subfamilies (depicted with the same color code in the two trees; see legend). The evolutionary relatedness between most aGPCR subfamilies could not be determined with the Bayesian approach (polytomy). The width of branches indicates scales with UFboot support (left tree) or Bootstrap support (right tree) for the ancestral node. Branch lengths do not scale with evolutionary distance in this rendering. Clades poorly supported (< 70% Bootstrap support) or composed of sequences from a single choanoflagellate species were collapsed in this rendering (grey triangles). All the sequences used to build the trees are listed in Supplementary File 15. The fully annotated version of the two phylogenies, including bootstrap values, branch lengths, and all species names, are found in Supplementary File 16. **(B)** aGPCR subfamilies differ in their phylogenetic distribution across the choanoflagellate diversity. Subfamilies VIII, XIII, XI, IX, XVII, VII, and X are detected in diverse craspedid and acanthoecid choanoflagellates – therefore they were probably present in stem choanoflagellates. Notably, the subfamily VIII is the most widespread of all choanoflagellate aGPCR subfamilies and is conserved in filastereans and metazoans (Ansel et al. 2024). In contrast, subfamilies III, XV, XIX, II, I, XVIII, and XVI are restricted to craspedids. While subfamily III is broadly conserved across the diversity of craspedids, subfamilies XV, XIX, II, I, XVIII, XVI are confined to distinct subsets of craspedids. Finally, subfamilies IV, V, VI, XII, and XIV were only observed in acanthoecids and, therefore, might be acanthoecid-specific. Asterisks (*) indicate aGPCR subfamilies that were recovered in only one of the two phylogenetic reconstructions (see (**A**)).

A

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX
Laminin_G	0	7	4	0	0	0	0	13	0	0	0	0	82	0	0	33	0	0	4
HRM	0	41	23	0	0	0	0	0	8	25	0	0	0	0	9	17	43	75	8
EGF	0	0	12	0	0	0	0	0	0	0	11	0	82	0	4	17	0	0	0
LRR	5	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0	0	0	13
FN3	5	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
Furin-like	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Calx_beta	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	0	0	0	0
Lectin_C	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Cadherin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	8
EGF_CA	0	0	0	0	0	0	0	0	0	0	0	0	82	0	0	17	0	0	0
Somatomedin_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sushi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F5_F8_C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
VWF (A/D)	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0
TSP-1	0	0	0	0	0	0	0	0	0	0	0	0	82	0	13	33	0	0	0
PAN	0	0	4	0	0	0	0	25	0	0	0	0	0	3	4	0	0	0	0
Ephrin_rec-like	10	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0
Collagen	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kringle	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
Gal_lectin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIG	5	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0
TNFR_c6	14	0	8	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	17
HYR	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0
Cohesin	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	25	0
PT	0	0	0	0	19	32	0	0	15	0	22	0	0	28	0	0	0	0	0
TLDc	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
Disintegrin	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CIMR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

80
60
40
20
0
% of domain containing genes per aGPCR subfamily

B

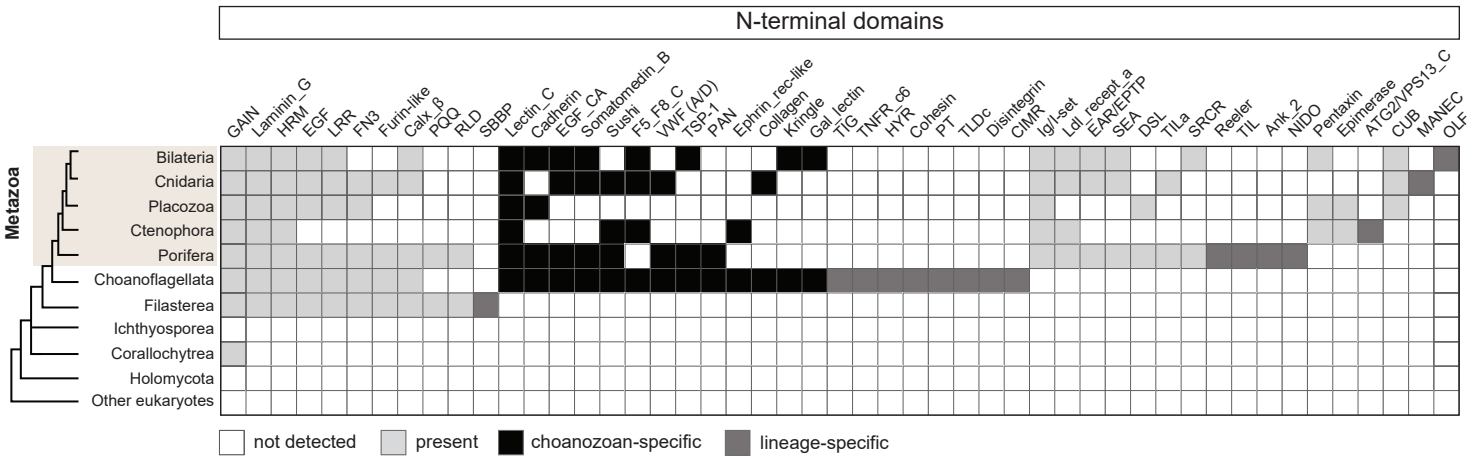


Fig. S7

Figure S7: Diversity and evolution of aGPCR extracellular protein domains in choanoflagellates and other opisthokonts. (A) Subfamilies of choanoflagellate aGPCRs exhibit distinct N-terminal protein domain signatures. The matrix summarizes the systematic identification of the protein domains (rows) present in the extracellular region of the members of the 19 choanoflagellate aGPCRs subfamilies previously established (columns) (**Fig. S6**). Percentages represent the number of domain-containing genes per aGPCR subfamily. For example, 52% of aGPCRs from the subfamily I contain a Hyalin Repeat domain (HYR) in their N-termini while mostly absent or underrepresented in the other aGPCR subfamilies. The GAIN domain, present in all aGPCRs, is not shown in this matrix. All the aGPCR NTFs analyzed are found in Supplementary File 19. **(B)** Phylogenetic distribution of protein domains in the N-termini of aGPCRs in opisthokonts. While non-holozoan aGPCRs do not possess conserved extracellular protein domains, most holozoan lineages, with the possible exception of ichthyosporeans, encode additional domains in their N-termini. Clear recruitment of ECDs is observed in the NTF of filasterean, choanoflagellate, and metazoan aGPCRs, with domains being conserved either in most of these lineages (light grey), in a subset of these lineages (e.g black for choanozoan-specific domains), or being lineage-specific (dark grey).

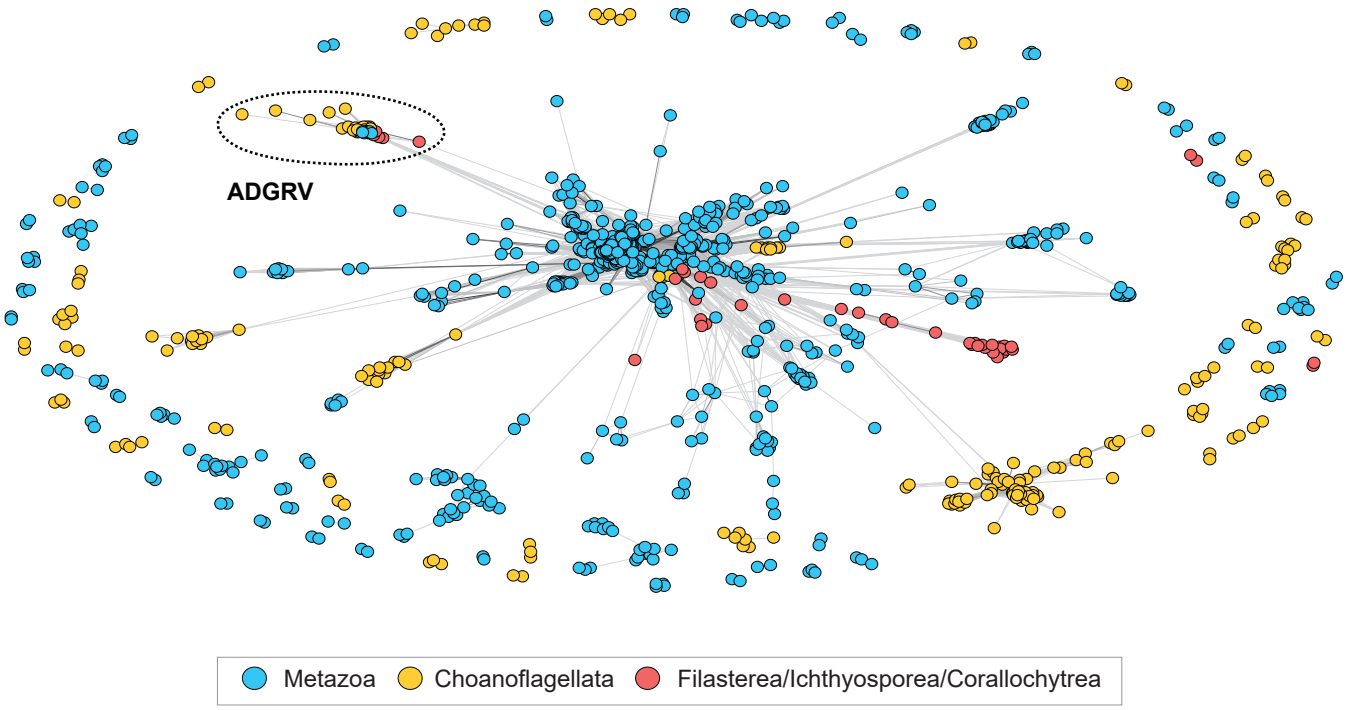
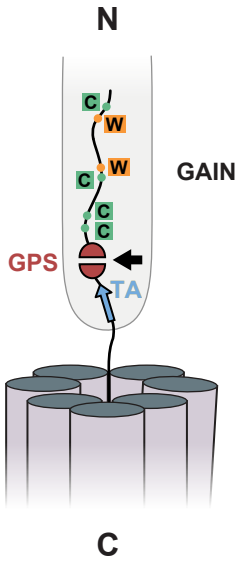

ADGRV
Metazoa
Choanoflagellata
Filasterea/Ichthyosporea/Corallochytrea

Fig. S8

Figure S8: Independent diversification of most aGPCR NTFs in metazoans and CRMs. Pairwise similarity-based clustering analysis of the N-terminal fragment (NTF) of 1074 metazoan (blue), 337 choanoflagellate (orange), and 81 filasterean/ichthyosporean aGPCRs. The NTF of most metazoan aGPCRs cluster separately from choanoflagellate or other holozoan NTFs. A noticeable exception is observed for metazoan NTFs from the ADGRV family, which group with choanoflagellate and filasterean NTF sequences (delineated with a dotted oval). The choanoflagellate NTFs that cluster with metazoan ADGRV NTFs belong to the aGPCR subfamily VIII, which we found to be orthologous to metazoan ADGRV based on their 7TM domain (**Fig. 5A**). Edges correspond to BLAST connections of P value <1e-20. All the NTF sequences used in the clustering analysis are available in Supplementary File 19.

A


N
C
W
W
C
C
C
GAIN
GPS
TA
C

B

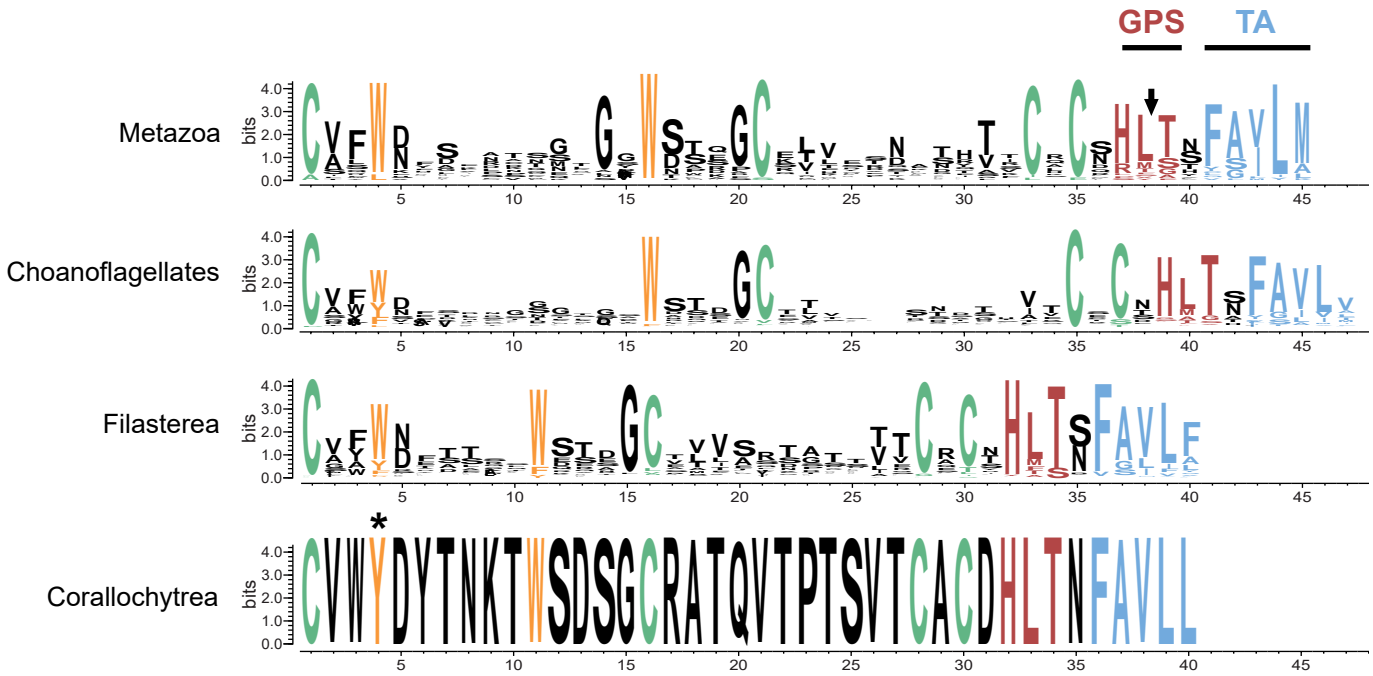
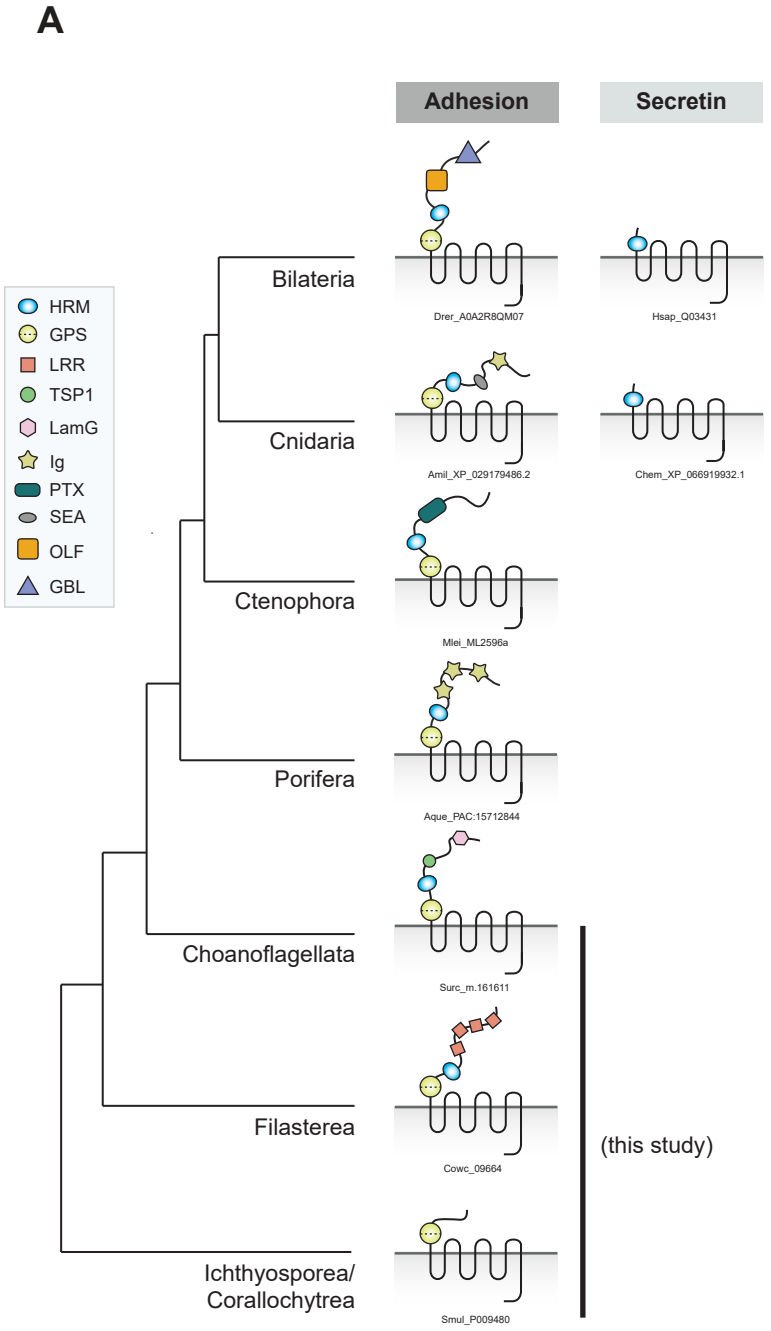
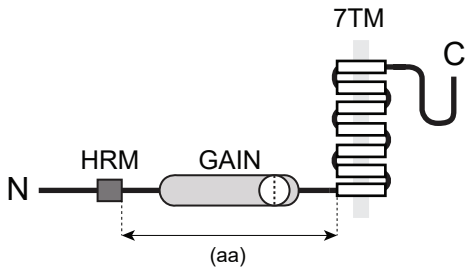
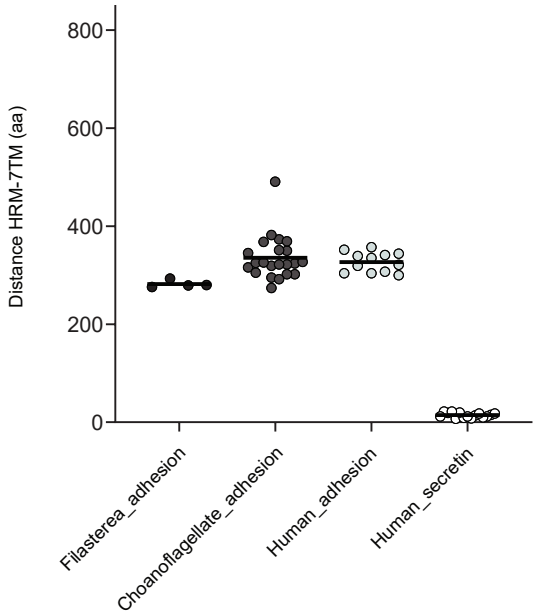

GPS
TA
Metazoa
Choanoflagellates
Filasterea
Corallochytrea
bits
4.0
3.0
2.0
1.0
0.0
5
10
15
20
25
30
35
40
45
*

Fig. S9

Figure S9: Conservation of the GAIN domain in metazoans, choanoflagellates, filastereans, and corallochytreans. (A) Schematic of the 7TM-proximal GAIN domain, which extends into the extramembrane milieu. Highlighted are key tryptophans (orange Ws) and cysteines (green Cs) flanking the proteolytic GPS motif (red, cleavage site indicated with an arrow) and the hydrophobic tethered agonist peptide (TA, blue). **(B)** Sequence logos of the metazoan, choanoflagellate, filasterean, and corallochytrean C-terminal GAIN domains. The two canonical cysteines and tryptophans, important for the proper folding of the GAIN domain (Araç et al. 2012), are conserved in choanoflagellates, filastereans, and corallochytreans. Similarly, the autoproteolytic GPS motif and most of the TA consensus sequence, both required for the activation of aGPCRs, are conserved in all these clades (Prömel et al. 2012b; Barros-Álvarez et al. 2022). An alignment of the complete repertoire of aGPCRs (30 sequences) from *Mus musculus* was used to generate the metazoan logo; 301 aGPCR sequences were used for the choanoflagellate logo; and 89 aGPCR sequences were used for the filasterean logo. Because we recovered only one aGPCR sequence with a GAIN domain in corallochytreans, their logo has no real statistical value and is provided as a qualitative point of comparison. All the sequences used to create the logos are found in Supplementary File 20.


A
Adhesion
Secretin
HRM
GPS
LRR
TSP1
LamG
Ig
PTX
SEA
OLF
GBL
Bilateria
Cnidaria
Ctenophora
Porifera
Choanoflagellata
Filasterea
Ichthyosporea/
Corallochytrea
Drer_A0A2R8QM07
Hsap_Q03431
Amil_XP_029179486.2
Chem_XP_066919932.1
Mlei_ML2596a
Aque_PAC:15712844
Surc_m.161611
Cowc_09664
Smul_P009480
(this study)


B
7TM
C
N
HRM
GAIN
(aa)


C
Distance HRM-7TM (aa)
800
600
400
200
0
Filasterea_adhesion
Choanoflagellate_adhesion
Human_adhesion
Human_secretin

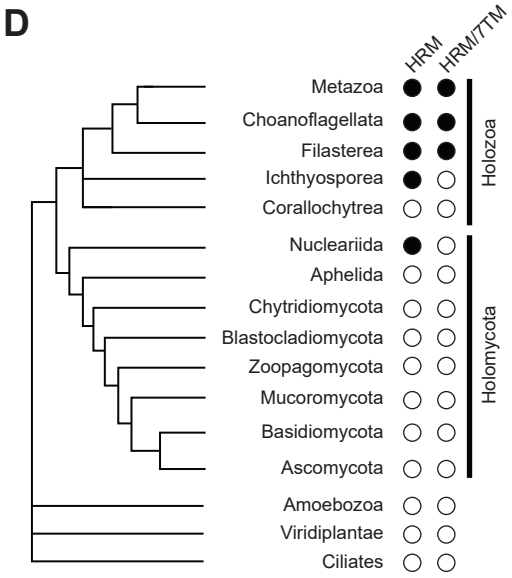

D
HRM
HRM/7TM
Metazoa
Choanoflagellata
Filasterea
Ichthyosporea
Corallochytrea
Holozoa
Nucleariida
Aphelida
Chytridiomycota
Blastocladiomycota
Zoopagomycota
Mucoromycota
Basidiomycota
Ascomycota
Holomycota
Amoebozoa
Viridiplantae
Ciliates

Fig. S10

Figure S10: Pre-metazoan origin of the HRM domain and HRM-GAIN-7TM module. (A) Evolution of HRM-containing aGPCRs and Secretin GPCRs in Holozoa. Some aGPCRs and all Secretin GPCRs possess a Hormone Receptor Motif (HRM) domain (blue oval) in their extracellular region; the HRM domain is always found in combination with a GAIN domain (yellow circle) in aGPCRs while the GAIN domain is absent from the Secretin receptors. Secretin GPCRs likely evolved from aGPCRs (Nordström et al. 2009b; Scholz et al. 2019) and are only found in Cnidarians (this study) and in Bilaterians (Cardoso et al. 2024) (right). In contrast, HRM-containing aGPCRs are more ancient and were detected in metazoans, choanoflagellates, and filastereans (left). **(B)** Representative illustration of the extracellular region of HRM-containing aGPCRs. The GAIN domain (light grey) sits on the top of the 7TM. The HRM domain (dark grey rectangle) is always positioned N-terminal to the GAIN domain. The distance separating the C-terminal end of the HRM domain from the start of the 7TM (HRM-7TM distance) is depicted with a double-headed arrow and is assessed in panel (C). Diverse additional protein domains (ECDs), distal to the HRM/GAIN/7TM module, are often found in HRM-containing aGPCRs and are not depicted here. **(C)** Conserved HRM-7TM distance in filasterean, choanoflagellate, and human HRM-containing aGPCRs. We measured an average HRM-7TM distance of 282 aa, 335 aa, and 326 aa in HRM-containing filasterean, choanoflagellate, and human aGPCRs, respectively. In contrast, human Secretin GPCRs show an average HRM-7TM distance of 14 aa due to these receptors' absence of the GAIN domain. Four filasterean, 22 choanoflagellate, 12 human HRM-containing aGPCRs, and 15 human Secretin GPCRs were assessed in this analysis. **(D)** Phylogenetic distribution of HRM and HRM/7TM module across eukaryotes. HRM is found in Holozoa and in the nucleariid *Fonticula alba*, one of the closest relatives of Fungi (Galindo et al. 2019). In contrast, the association of an HRM domain with 7TM (HRM/7TM) is only observed in holozoans – metazoans, choanoflagellates, and filastereans. The five HRM-containing proteins identified in *Fonticula alba* are provided in Supplementary File 21.

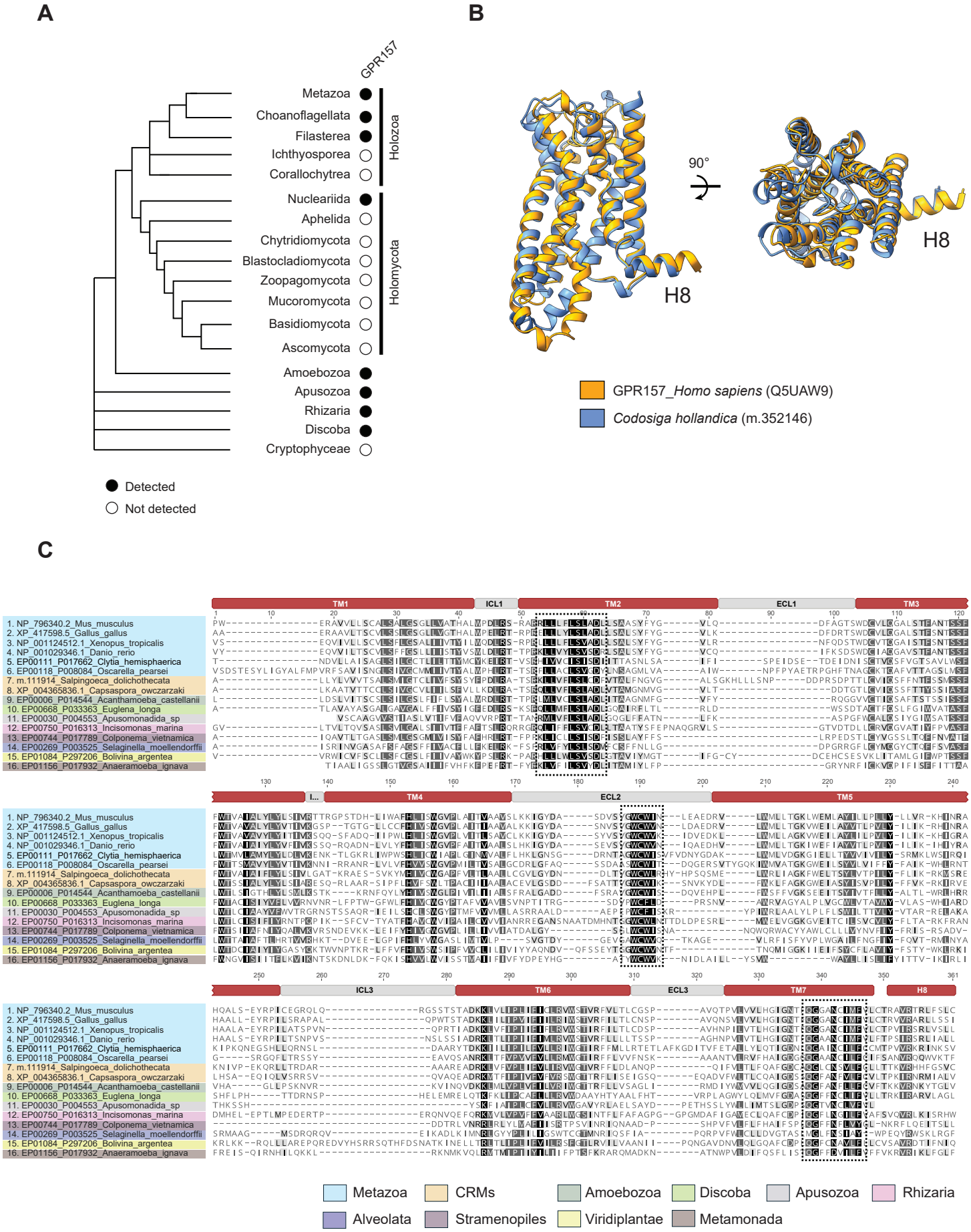

A
GPR157
Metazoa
Choanoflagellata
Filasterea
Ichthyosporea
Corallochytrea
Holozoa
Nucleariida
Aphelida
Chytridiomycota
Blastocladiomycota
Zoopagomycota
Mucoromycota
Basidiomycota
Ascomycota
Holomycota
Amoebozoa
Apusozoa
Rhizaria
Discoba
Cryptophyceae
Detected
Not detected
B
90°
H8
H8
GPR157_Homo sapiens (Q5UAW9)
Codosiga hollandica (m.352146)
C
TM1
ICL1
TM2
ECL1
TM3
TM4
ECL2
TM5
ICL3
TM6
ECL3
TM7
H8
1. NP_796340.2_Mus_musculus
2. XP_417598.5_Gallus_gallus
3. NP_001124512.1_Xenopus_tropicalis
4. NP_001029346.1_Danio_rerio
5. EP00111_P017662_Clytia_hemisphaerica
6. EP00118_P008084_Oscarella_pearsei
7. m.111914_Salpingoeca_dolichothecata
8. XP_004365836.1_Capsaspora_owczarzaki
9. EP00006_P014544_Acanthamoeba_castellanii
10. EP00668_P033363_Euglena_longa
11. EP00030_P004553_Apusomonadida_sp
12. EP00750_P016313_Incisomonas_marina
13. EP00744_P017789_Colponema_vietnamica
14. EP00269_P003525_Selaginella_moellendorffii
15. EP01084_P297206_Bolivina_argentea
16. EP01156_P017932_Anaeramoeba_ignava
Metazoa
CRMs
Amoebozoa
Discoba
Apusozoa
Rhizaria
Alveolata
Stramenopiles
Viridiplantae
Metamonada

Fig. S11

Figure S11: GPR157 is an ancient GPCR family conserved in eukaryotes. (A) Phylogenetic distribution of GPR157 across eukaryotes. GPR157 likely forms an ancient GPCR family in eukaryotes with homologs detected in Holozoa, Holomycota (only present in nucleariids), Amoebozoa, Apusozoa, Rhizaria, Discoba, and Cryptophyceae. Homologs were searched by BLASTing murine GPR157 (Q8C206) sequence against the entire dataset (993 species) of the EUKPROT v3 BLAST server (E-value: $1e^{-5}$) (https://evocellbio.com/eukprot/; (Richter et al. 2022)). We defined *bona fide* GPR157 hits as those with at least 70% query coverage and at least 30% sequence identity. All EUKPROT blast hit sequences are listed in Supplementary File 22. **(B)** Predicted structural homology of the 7TM region of *H.sapiens*_GPR157_Q5UAW9 (orange) and *C.hollandica*_m.352146 (blue) receptors. An additional helix 8 (H8) is predicted in both metazoan and choanoflagellate GPR157. All structural models shown here have a confidence score >70 pLDDT. View of the superimposed models from the plane of the membrane (left) and the top (right). **(C)** Multi-alignment of the 7TM/H8 domain of GPR157 from various eukaryotes reveals motif conservation, including a “LLxxLSL/V” motif in transmembrane helix 2 (TM2) and a "WCWI/V” motif in the extracellular loop 2 (ECL2) (dotted boxes).

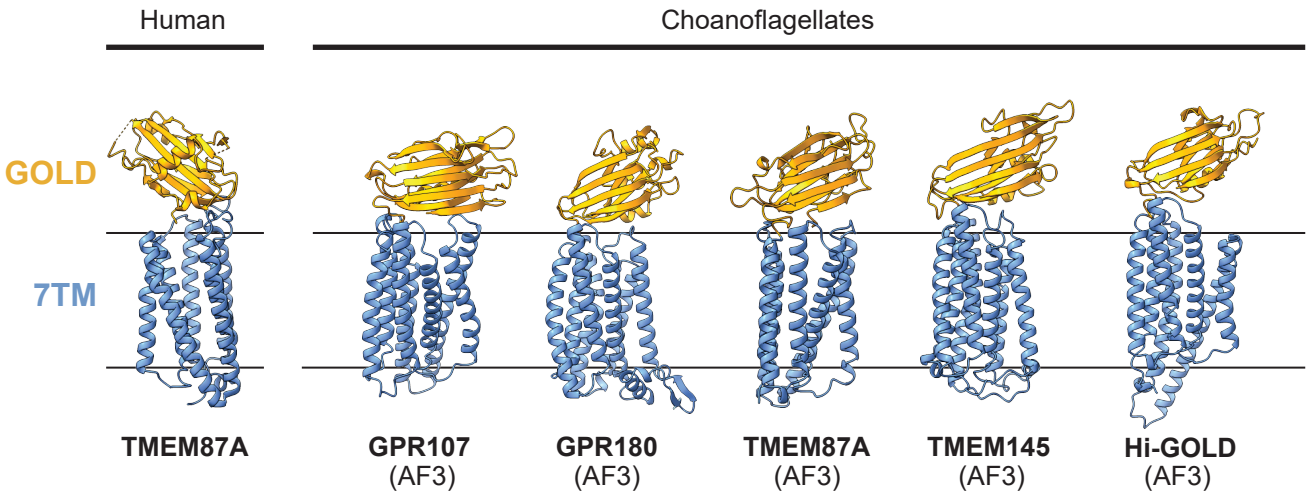
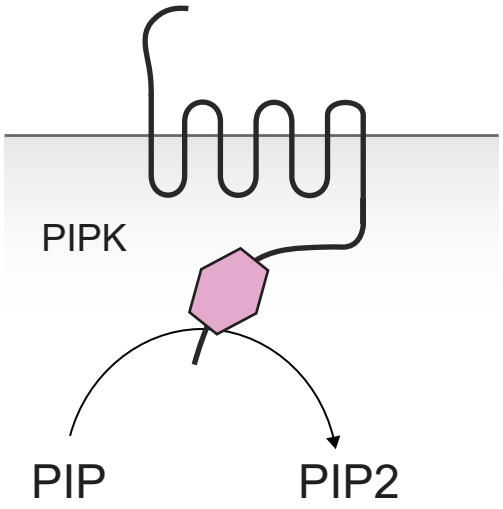

Human
Choanoflagellates
GOLD
7TM
TMEM87A
GPR107
(AF3)
GPR180
(AF3)
TMEM87A
(AF3)
TMEM145
(AF3)
Hi-GOLD
(AF3)

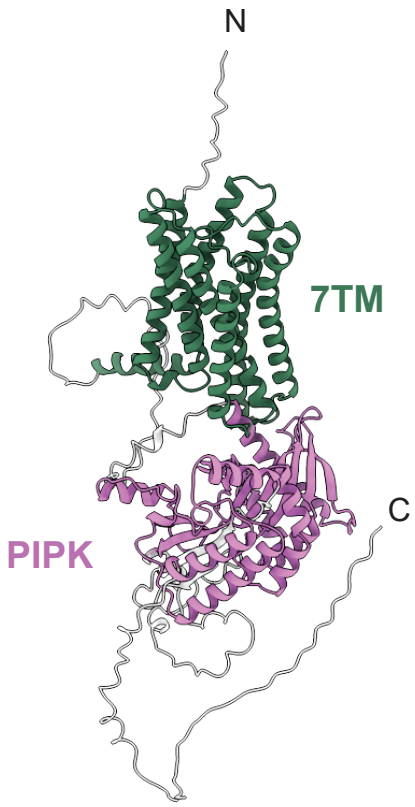
Fig. S12

Figure S12: Predicted structural similarities among metazoan and choanoflagellate GOST family receptors. Experimentally determined human TMEM87A structure (PDB:8CTJ; (Hoel et al. 2022)) and AlphaFold3 (AF3) predicted structures of representative proteins from the five GOST subfamilies detected in choanoflagellates (GPR107/108, GPR180, TMEM87A, TMEM145, and the newly identified Hi-GOLD) shown as cartoons, viewed from the plane of the membrane. Low confidence (<70 pLDDT) regions of predicted structures have been removed. Both metazoan and choanoflagellate GOST GPCRs exhibit a GOLD domain (orange) on top of a 7TM domain (blue).

A


PIPK
PIP
PIP2

B


N
7TM
C
PIPK

C

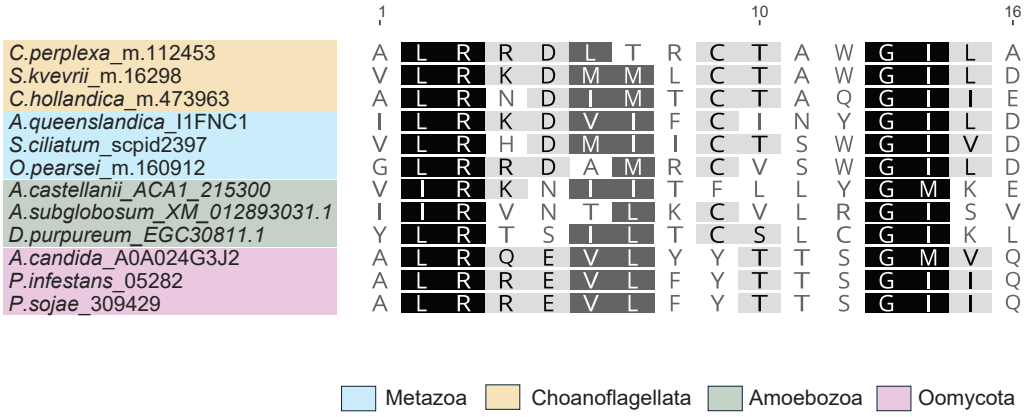

1
10
16
C.perplexa_m.112453 A L R R D L T R C T A W G I L A
S.kvevrii_m.16298 V L R K D M M L C T A W G I L D
C.hollandica_m.473963 A L R N D I M T C T A Q G I I E
A.queenslandica_I1FNC1 I L R K D V I F C I N Y G I L D
S.ciliatum_scpid2397 V L R H D M I I C T S W G I V D
O.pearsei_m.160912 G L R R D A M R C V S W G I L D
A.castellanii_ACA1_215300 V I R K N I I T F L L Y G M K E
A.subglobosum_XM_012893031.1 I I R V N T L K C V L R G I S V
D.purpureum_EGC30811.1 Y L R T S I L T C S L C G I K L
A.candida_A0A024G3J2 A L R Q E V L Y Y T T S G M V Q
P.infestans_05282 A L R R E V L F Y T T S G I I Q
P.sojae_309429 A L R R E V L F Y T T S G I I Q
Metazoa
Choanoflagellata
Amoebozoa
Oomycota

Fig. S13

Figure S13: Features of choanoflagellate GPCR PIPKs. (A) A schematic depiction of a GPCR PIPK. GPCR PIPKs present a canonical 7TM domain combined with a phosphatidylinositol phosphate kinase (PIPK; pink hexagon) at the C-terminus (Van Den Hoogen et al. 2018; van den Hoogen and Govers 2018). While experimental evidence is lacking, the PIPK domain of GPCR PIPK receptors is likely to signal via the production of phospholipid-based second messengers (e.g PIP2). **(B)** Alphafold-predicted structure of a choanoflagellate GPCR PIPK (*S.urceolata*_m_147488) with the 7TM and PIPK domains depicted in green and pink respectively. Regions with a low prediction score (<70 pLDDT) are depicted in white. **(C)** Multiple sequence alignment showing the conservation in choanoflagellates of the diagnostic GPCR PIPK motif "LR$(x)_9$GI" in the linker region separating the 7TM domain from the PIPK domain.

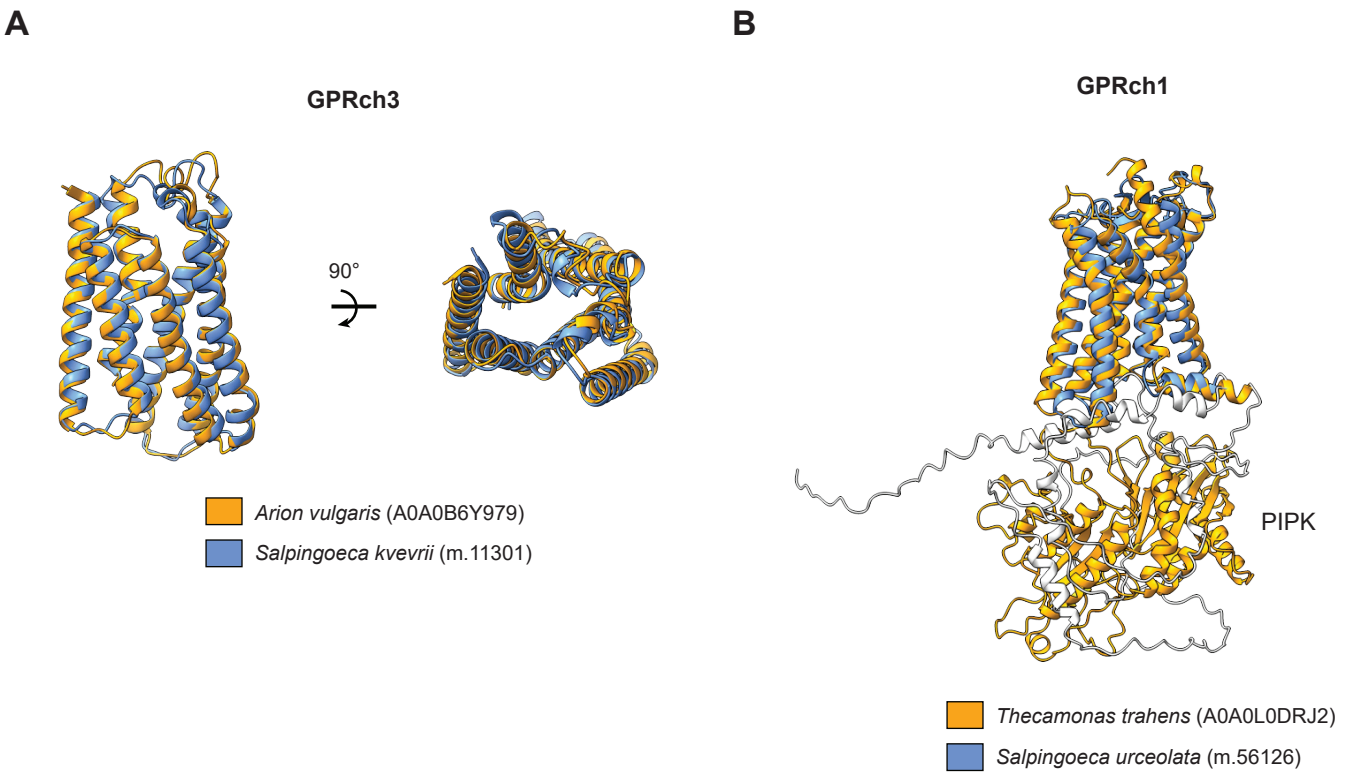

A
GPRch3
90°
Arion vulgaris (A0A0B6Y979)
Salpingoeca kvevrii (m.11301)
B
GPRch1
PIPK
Thecamonas trahens (A0A0L0DRJ2)
Salpingoeca urceolata (m.56126)

Fig. S14

Figure S14: Structural features of GPRch3 and GPRch1 GPCRs. (A) Predicted structural homology of the 7TM domain from choanoflagellate GPRch3 GPCR *S.kvevrii*_m.11301 (orange) and its top structural metazoan hit *A.vulgaris*_A0A0B6Y979 (blue) (E-value: $4.64e^{-13}$). View of the superimposed models from the plane of the membrane (left) and the top (right). All structural models shown here have a confidence score >70 pLDDT. **(B)** Superimposed structures of full-length choanoflagellate GPRch1 GPCR *S.urceolata*_m.56126 (blue) with its top structural hit *T.trahens*_A0A0L0DRJ2 (orange) (E-value: $5.92e^{-10}$). While ciliates, along with other non-choanozoan eukaryotes that encode GPRch1 GPCRs, possess an additional PIPK domain in the C-terminal region of the GPCR, no GPRch1 GPCRs appeared to have this additional cytosolic enzymatic domain in choanoflagellates. Regions with a low prediction score (<70 pLDDT) are colored in white.

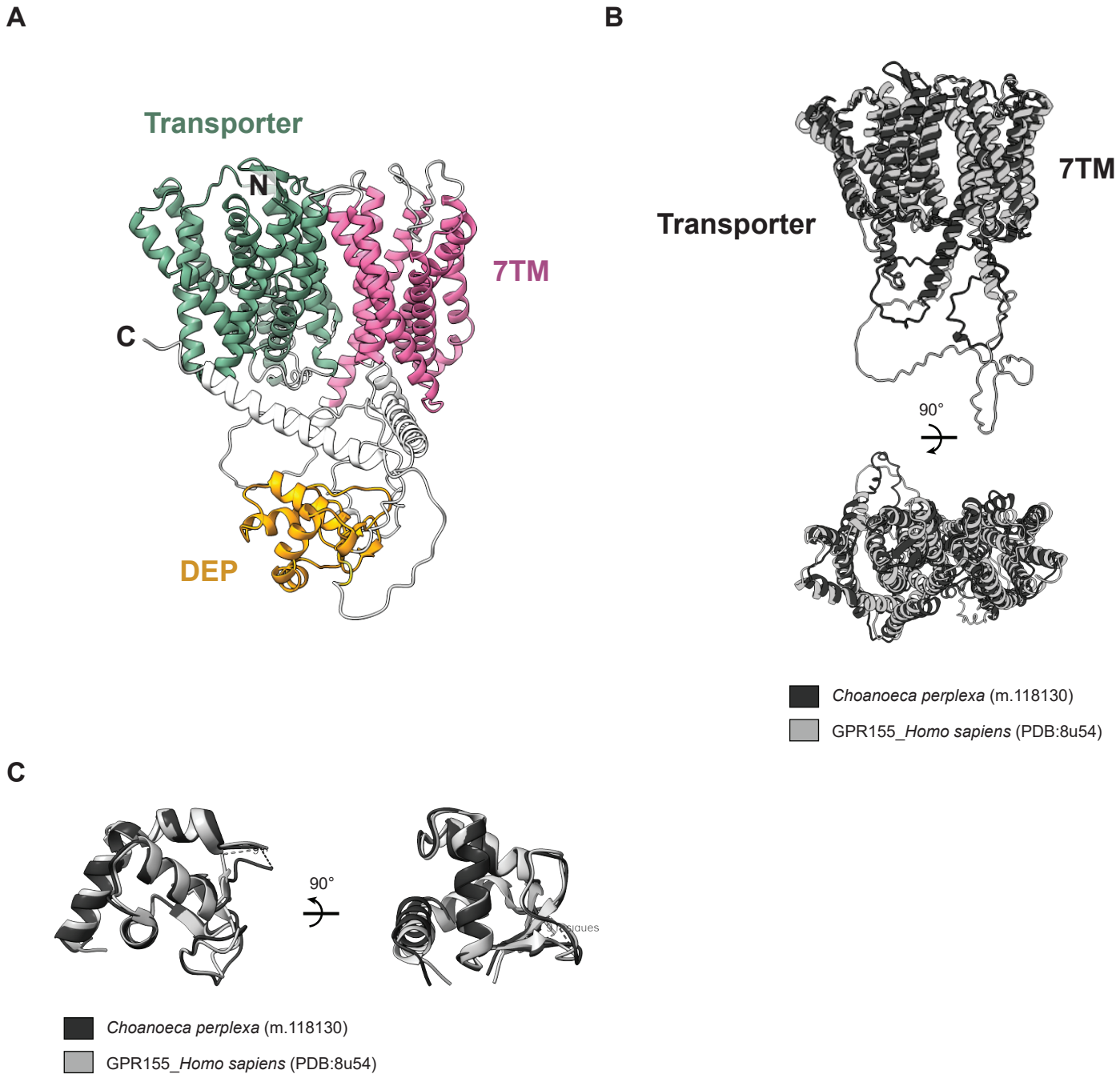
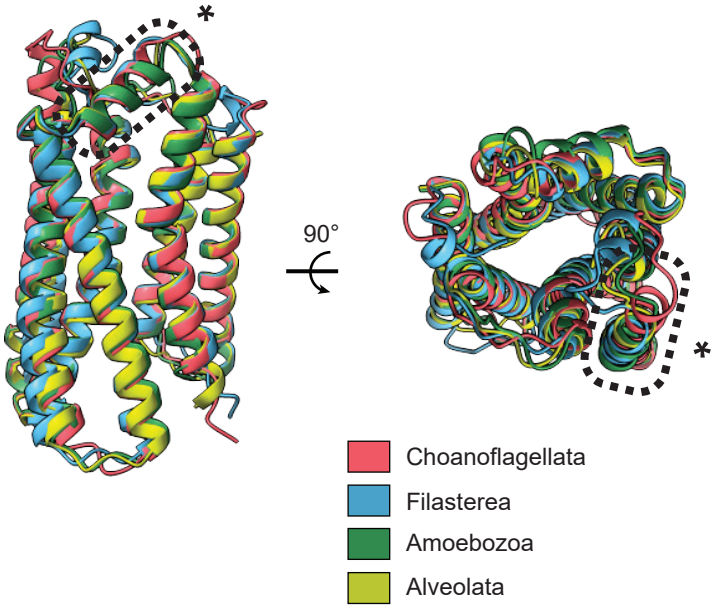

A
Transporter
N
7TM
C
DEP
B
7TM
Transporter
90°
Choanoeca perplexa (m.118130)
GPR155_Homo sapiens (PDB:8u54)
C
90°
Choanoeca perplexa (m.118130)
GPR155_Homo sapiens (PDB:8u54)

Fig. S15

Figure S15: Conservation of metazoan GPR155 in choanoflagellates. (A) Predicted structure of a full-length choanoflagellate GPR155 receptor (*C. perplexa*_m.118130). Similar to metazoan GPR155 (Shin et al. 2022; Bayly-Jones et al. 2024)), choanoflagellate GPR155 possesses a 7TM domain (pink) fused to a transporter domain (green), and an additional Dishevelled, EGL-10 and Pleckstrin (DEP) domain (orange) in the C-terminal region. Regions with a low prediction score (<70 pLDDT) are depicted in white. **(B)** Structural homology of the Transporter/7TM module from the alphafold-predicted choanoflagellate GPR155 structure (*C. perplexa*_m.118130 ; dark grey) and the experimentally solved human GPR155 structure (PDB: 8u54 ; light grey). View of the superimposed models from the plane of the membrane (top) and the top (bottom). **(C)** Superimposed structures of the DEP domain from choanoflagellate (*C. perplexa*_m.118130 ; dark grey) and human (PDB: 8u54 ; light grey) GPR155 GPCRs. The choanoflagellate DEP model has a confidence score >70 pLDDT.

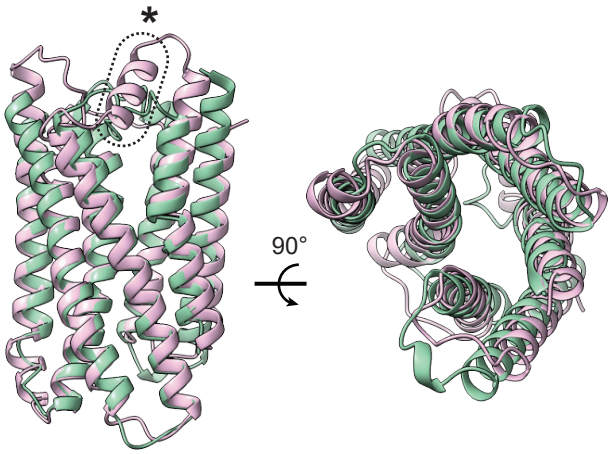
A



B

C.perplexa_m.31142	N	V	L	Q	N	E	L	N	E	L	Q	K	Q	D	-	-	-	-	D	L	T
S.macrocollata_m.202162	N	L	V	Q	D	Y	M	N	D	L	Q	R	K	P	-	-	-	-	S	M	I
S.punica_m.92936	N	R	V	Q	T	F	L	N	E	L	Q	E	R	S	-	-	-	-	D	M	M
S.helianthica_m.34715	N	L	L	Q	S	M	M	N	D	L	Q	D	R	D	-	-	-	-	P	L	T
M.fluctuans_m.80795	N	I	L	Q	N	K	M	N	D	L	Q	N	S	P	-	-	-	-	D	L	T
S.dolichothecata_m.104504	N	Y	V	Q	K	W	L	N	E	L	Q	G	S	-	-	-	-	-	H	L	K
S.parva_m.62654	N	V	L	Q	N	G	M	N	A	L	Q	T	T	-	-	-	-	-	D	T	T
A.spectabilis_m.458266	N	A	L	Q	S	G	M	N	D	L	Q	T	R	-	-	-	-	-	D	A	F
H.nana_m.129869	N	V	L	Q	S	G	M	N	D	L	Q	T	K	-	-	-	-	-	D	A	M
D.grandis_m.160170	N	P	L	Q	A	L	M	N	R	L	E	V	T	-	-	-	-	-	D	L	E
S.diplocostata_m.693985	N	P	L	Q	S	R	M	N	H	L	Q	S	Q	A	-	-	-	-	N	P	Y
C.owczarzaki_XP_004364167.1	N	I	A	Q	Q	Q	M	N	D	M	I	I	D	-	-	-	-	-	S	N	N
P.chileana_Opistho-2@20593	N	V	I	Q	N	Y	L	N	K	L	N	S	D	D	A	R	H	S	N	E	I
V.vermiformis_DN7877	N	E	L	Q	D	L	M	S	D	W	L	D	D	-	-	-	-	-	S	P	A
M.gubernata_DN949	N	P	V	Q	E	R	M	S	N	W	I	S	R	-	-	-	-	-	D	H	E
M.erythrocladioides_ CAMPEP_0198318540	N	P	L	Q	Q	Q	L	S	A	W	A	Q	S	N	E	E	-	R	F	L	N

C



Tobamovirus multiplication protein 1 (*Zea mays*)

Stephanoeca diplocostata p.026032 m.693985

Fig. S16

Figure S16: Structural features of the newly established RSF GPCR family. (A) Predicted structural homology of the RSF 7TM domain from *S. diplocostata*_m.693985 (red), *C. owczarzaki*_P006019 (blue), *M.gubernata*_dn949 (yellow), and *V. vermiformis*_dn7877 (green). A conserved short helix between TM6 and TM7 is predicted in all the structures depicted (dotted box). View of the superimposed models from the plane of the membrane (left) and the top (right). All structural models shown here have a confidence score >70 pLDDT. **(B)** A multi-alignment of sequences corresponding to the short extra helix depicted in (A), revealed a NxLQxxMN" motif conserved across eukaryotes. **(C)** Mild structural similarities are shared between predicted *S. diplocostata*_m693985 (pink) and top structural Foldseek hit *Z. mays*_TOM1_B7ZYE1 (green) (E-value:1.66e^{-5}). The short extra helix (dotted box) is not observed in TOM1 GPCRs. View of the superimposed models from the plane of the membrane (left) and the top (right).

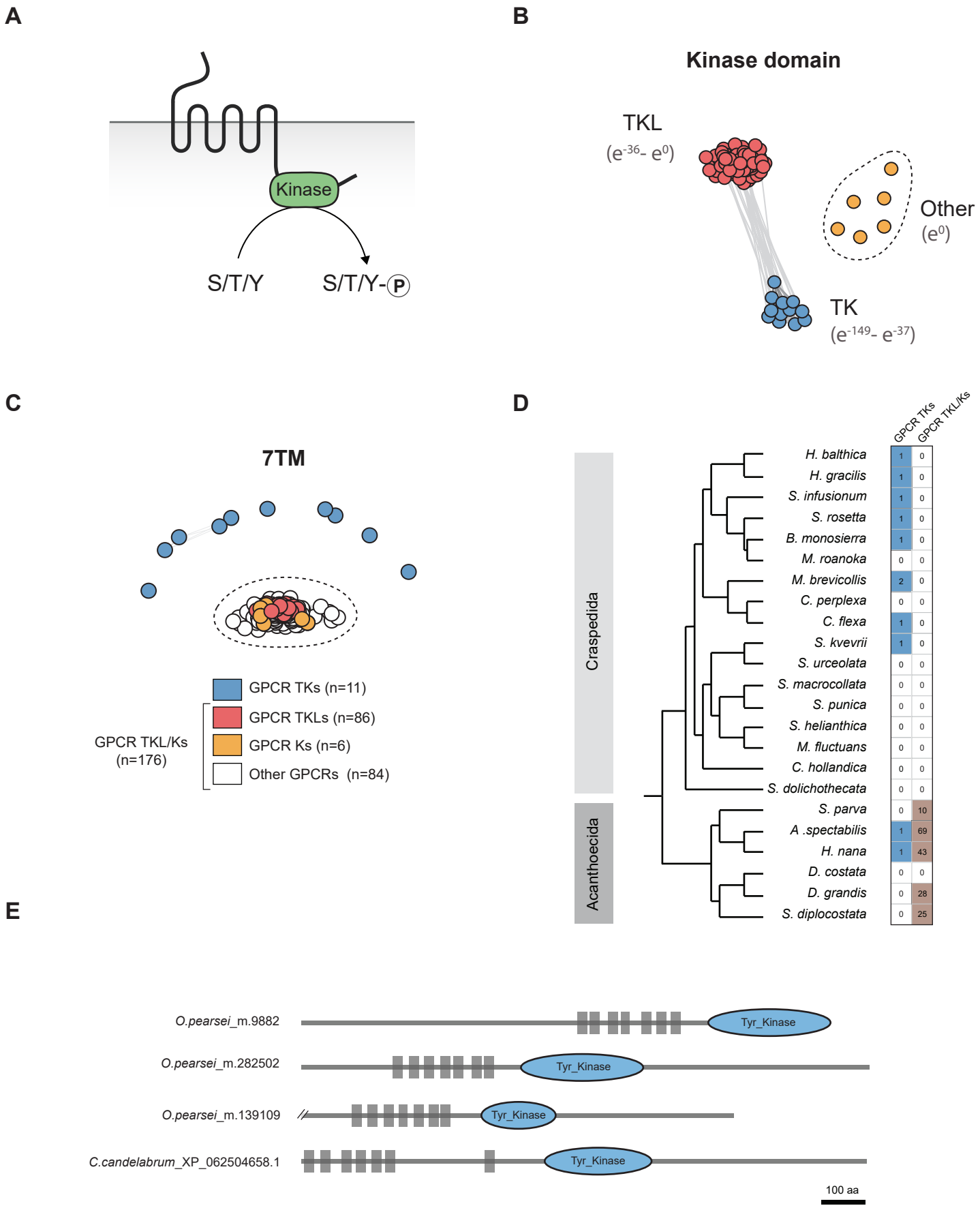

A
Kinase
S/T/Y
S/T/Y-P
B
Kinase domain
TKL
(e-36- e0)
Other
(e0)
TK
(e-149- e-37)
C
7TM
GPCR TKs (n=11)
GPCR TKLs (n=86)
GPCR Ks (n=6)
Other GPCRs (n=84)
GPCR TKL/Ks
(n=176)
D
GPCR TKs
GPCR TKL/Ks
Craspedida
Acanthoecida
H. balthica 1 0
H. gracilis 1 0
S. infusionum 1 0
S. rosetta 1 0
B. monosierra 1 0
M. roanoka 0 0
M. brevicollis 2 0
C. perplexa 0 0
C. flexa 1 0
S. kvevrii 1 0
S. urceolata 0 0
S. macrocollata 0 0
S. punica 0 0
S. helianthica 0 0
M. fluctuans 0 0
C. hollandica 0 0
S. dolichothecata 0 0
S. parva 0 10
A .spectabilis 1 69
H. nana 1 43
D. costata 0 0
D. grandis 0 28
S. diplocostata 0 25
E
O.pearsei_m.9882
Tyr_Kinase
O.pearsei_m.282502
Tyr_Kinase
O.pearsei_m.139109
Tyr_Kinase
C.candelabrum_XP_062504658.1
Tyr_Kinase
100 aa

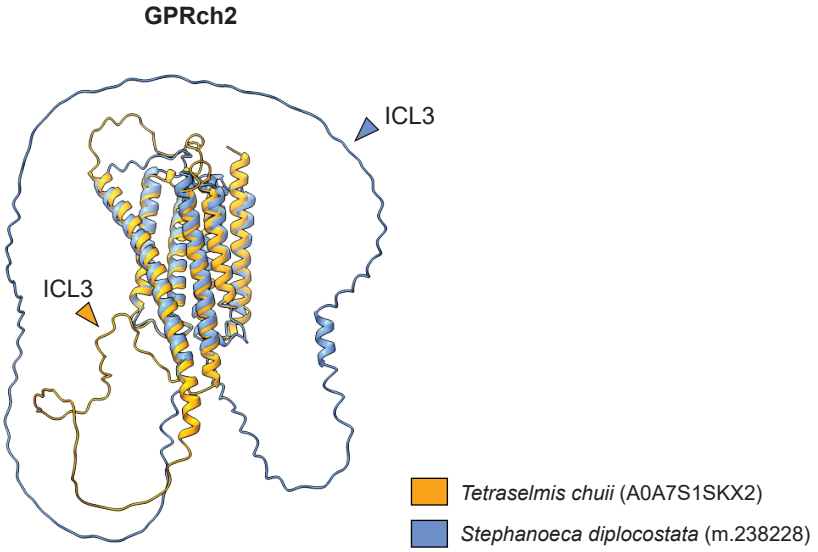
Fig. S17

2870 **Figure S17: Features and evolution of choanoflagellate GPCR TK/TKL/Ks**. **(A)** A schematic depiction of a GPCR Kinase. GPCR Kinases present a canonical 7TM domain combined with a kinase domain (green oval) in their C-terminus (Van Den Hoogen et al. 2018; van den Hoogen and Govers 2018). The kinase domain possibly phosphorylates the GPCR itself or other proteins (van den Hoogen and Govers 2018). **(B)** All-against-all pairwise comparison of the
2875 catalytic domain of choanoflagellate GPCR Kinases reveals distinct kinase domain populations with 11 tyrosine kinase (TK) domains grouping together (blue) and sharing sequence similarities with a second cluster encompassing 86 tyrosine kinase-like (TKL); 6 additional domains (orange) do not share sequence similarities with any other kinase domains in this dataset (Other) (Supplementary Files 24 and 25). Each kinase domain (circle) was blasted against a
2880 local version of KinBase, the curated protein kinase dataset from www.kinase.com (Manning et al. 2002; Bradham et al. 2006; Goldberg et al. 2006)), for identification (Supplementary File 26). The E-values depicted summarize the lowest and highest e-values associated with top blast hits obtained for members of each cluster. While choanoflagellate GPCR TKs systematically recovered metazoan tyrosine kinases as their top BLAST hits, choanoflagellate GPCR TKL
2885 preferentially matched with amoebozoan tyrosine kinase-like proteins. No BLAST hits were recovered for the 6 isolated kinase domains. Connecting lines correspond to pairwise BLAST scores of p-value $<1e^{-20}$. **(C)** Distinct families of GPCR Kinases are found in choanoflagellates. All-against-all pairwise comparison of the 7TM domain of choanoflagellate GPCR Kinases revealed that they cluster into different families, with the 86 GPCR TKL receptors (red)
2890 previously identified in (B) grouping with the 6 GPCR K (orange), and 84 additional choanoflagellate GPCRs that do not possess a C-terminal kinase domain (white), forming a cluster of 176 GPCRs in total (dotted line; Fig. 1). On the opposite, the 11 GPCR TKs (blue) previously identified in (B) grouped independently from the GPCR TKL/Ks and were either single GPCRs or grouped with other GPCR TKs. Connecting lines correspond to pairwise
2895 BLAST scores of p-value $<1e^{-6}$. The 7TM sequences analyzed and the output of the clustering analysis are found in Supplementary Files 27 and 28. **(D)** Phylogenetic distribution of GPCR TKs (blue) and GPCR TKL/Ks (brown) within the choanoflagellate phylogeny. While GPCR TKs are found in a large range of choanoflagellate species, including both craspedids and acanthoecids, GPCR TKL/Ks appear to be restricted to acanthoecids (Fig. 2). **(E)** Schematic
2900 representation of selected sponge GPCR TKs recovered in our analysis by blasting the kinase domain of choanoflagellate GPCR kinases against the EUKPROT database (Richter et al. 2022).GPCRs are depicted at scale. Scale bar = 100 aa.

2905

2910

A


GPRch2
ICL3
ICL3
Tetraselmis chuii (A0A7S1SKX2)
Stephanoeca diplocostata (m.238228)

B

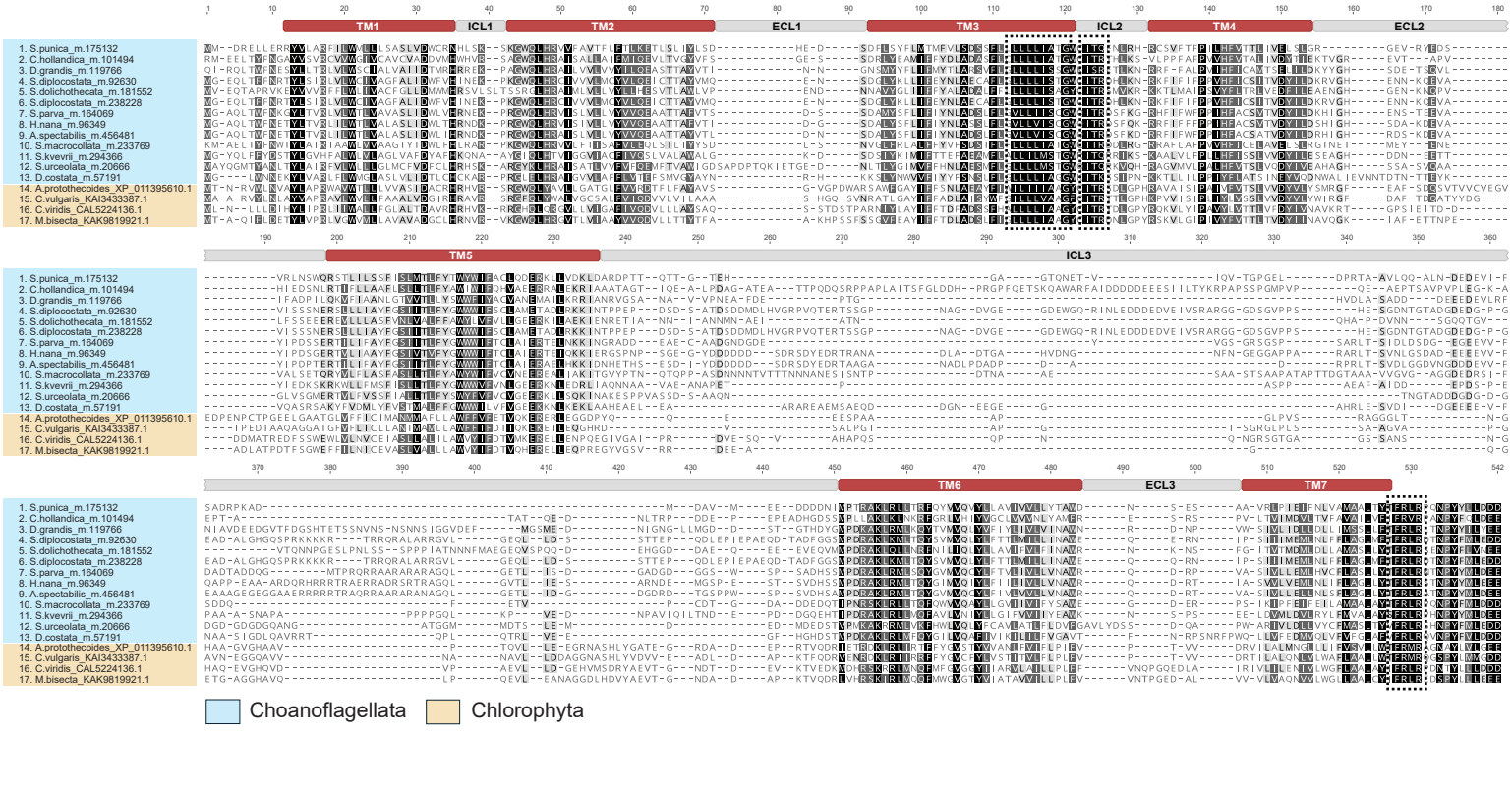
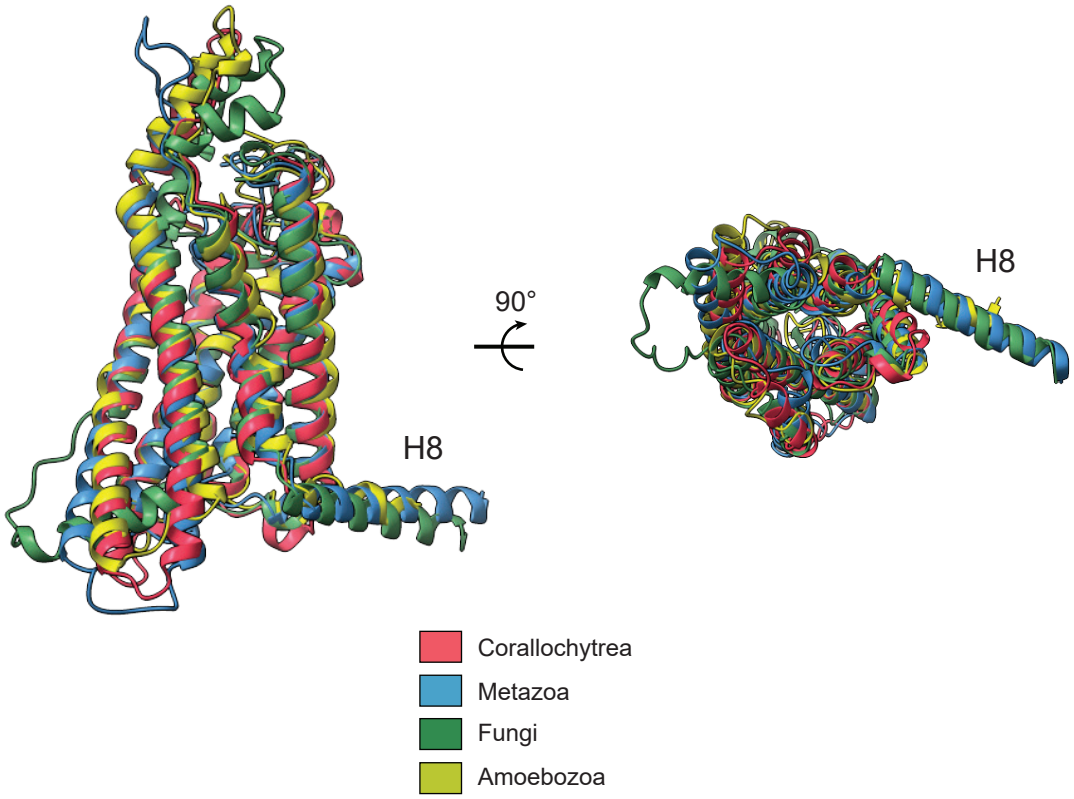

TM1
ICL1
TM2
ECL1
TM3
ICL2
TM4
ECL2
TM5
ICL3
TM6
ECL3
TM7
1. S.punica_m.175132
2. C.hollandica_m.101494
3. D.grandis_m.119766
4. S.diplocostata_m.92630
5. S.dolichothecata_m.181552
6. S.diplocostata_m.238228
7. S.parva_m.164069
8. H.nana_m.96349
9. A.spectabilis_m.456481
10. S.macrocollata_m.233769
11. S.kvevrii_m.294366
12. S.urceolata_m.20666
13. D.costata_m.57191
14. A.protothecoides_XP_011395610.1
15. C.vulgaris_KAI3433387.1
16. C.viridis_CAL5224136.1
17. M.bisecta_KAK9819921.1
Choanoflagellata
Chlorophyta

Fig. S18

Figure S18: Structural features of GPRch2 GPCRs. (A) Predicted structural homology of the 7TM domain from choanoflagellate GPRch2 GPCR *S.diplocostata*_m.238228 (blue) and its top structural chlorophyte hit *T.chuii*_A0A7S1SKX2 (orange) (E-value: $6.83e^{-12}$). A large intracellular loop 3 (ICL3) between TM5 and TM6 (indicated with an arrowhead) is observed in choanoflagellate and, to a lesser extent, in chlorophyte GPRch2 GPCRs. ICL3 is unstructured with a low confidence score found in both models (<70 pLDDT). **(B)** Multiple sequence alignment of the 7TM region of GPRch2 GPCRs from choanoflagellates and chlorophytes revealed a conserved “LLLLIS/AxG” motif in TM3, an “ITR” motif in ICL2, and a “FRLRxxNPY” motif in the C-terminal region (dotted boxes).

A


H8
90°
H8
Corallochytrea
Metazoa
Fungi
Amoebozoa

B

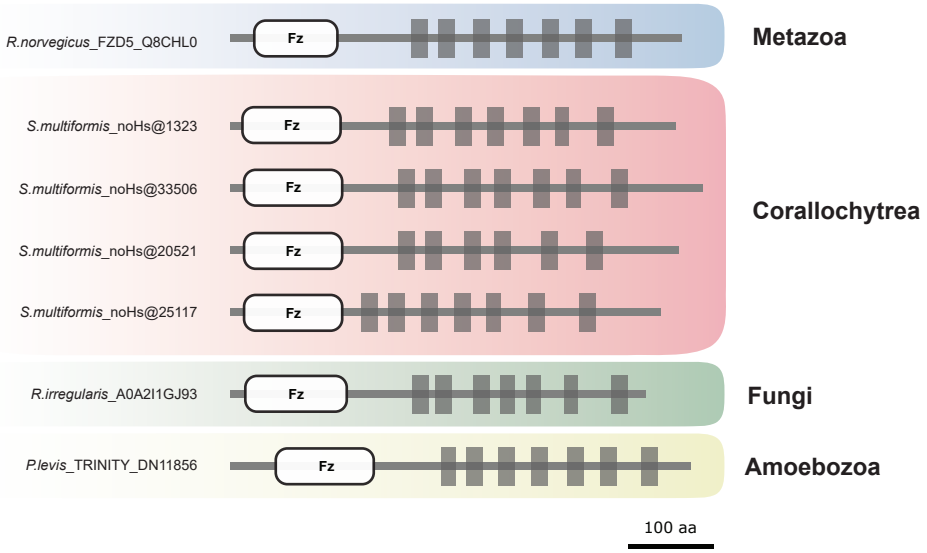

R.norvegicus_FZD5_Q8CHL0
Fz
Metazoa
S.multiformis_noHs@1323
S.multiformis_noHs@33506
S.multiformis_noHs@20521
S.multiformis_noHs@25117
Corallochytrea
R.irregularis_A0A2I1GJ93
Fungi
P.levis_TRINITY_DN11856
Amoebozoa
100 aa

Fig. S19

Figure S19: Metazoan Frizzled/Smoothened homologs detected in corallochytreans, fungi, and amoebozoans. (A) Predicted structural similarities of the 7TM region of *S.multiformis*_p007117_colp12_trinity150504 (red), *R.norvegicus*_Q8CHL0 (blue), *R.irregularis*_a0a2i1gj93 (green), and *P.levis*_p001581_dn11856 (yellow) Frizzled/Smoothened receptors. An additional helix 8 (H8) C-terminal to the 7TM domain is also depicted. All predicted regions shown here have a confidence score >70 pLDDT. View of the superimposed models from the plane of the membrane (left) and the top (right). **(B)** Schematic representation of Frizzled/Smoothened GPCRs found in metazoans, corallochytreans, fungi, and amoebozoans. The structural homology extends beyond their 7TMs as a canonical Frizzled domain (Fz) is detected in the N-termini of all these receptors. The Frizzled/Smoothened homologs detected in *S.multiformis* are listed in Supplementary File 29. GPCRs are depicted at scale. Scale bar = 100 aa.