Title

Early eukaryotic origins and metazoan elaboration of MAPR family proteins

Authors

Elisabeth Hehenberger,[1,2] Michael Eitel,[3] Sofia A.V. Fortunato,[4] David J. Miller,[4] Patrick J. Keeling,[1] Michael A. Cahill[5,6,*]

[1] Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada.

[2] Current address: Marine Ecology Division, GEOMAR | Helmholtz Centre for Ocean Research Kiel, 24105 Kiel, Germany.

[3] Department of Earth and Environmental Sciences, Paleontology and Geobiology, Ludwig-Maximilians-Universität München, Munich, Germany.

[4] ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD, 4811, Australia.

[5] School of Biomedical Sciences, Charles Sturt University, Wagga Wagga, NSW 2678, Australia.

[6] ACRF Department of Cancer Biology and Therapeutics, The John Curtin School of Medical Research, Canberra, ACT 2601, Australia.

***Address Correspondence to:**

Dr. Michael Cahill, School of Biomedical Sciences, Charles Sturt University, Wagga Wagga, NSW 2678, Australia.

e-mail: mcahill@csu.edu.au

Tel: +61-2 69332729

ABSTRACT

Background: The membrane-associated progesterone receptor (MAPR) family consists of heme-binding proteins containing a cytochrome b_5 ($cytb_5$) domain characterized by the presence of a MAPR-specific interhelical insert region (MIHIR) between helices 3 and 4 of the canonical cytb5-domain fold. Animals possess three MAPR families (PGRMC-like, Neuferricin and Neudesin). Results: Here we show that all animal MAPR families were already present in the common ancestor of the Opisthokonta (comprising animals and fungi as well as related protistan taxa). All three MAPR genes acquired extensions C-terminal to the $cytb_5$ domain, either before or with the evolution of animals. The archetypical MAPR protein, progesterone receptor membrane component 1 (PGRMC1), contains phosphorylated tyrosines Y139 and Y180. The combination of Y139/Y180 appeared in the common ancestor of Cnidaria and bilaterally symmetrical animals, along with an early embryological organizer and synapsed neurons, and is strongly conserved in all bilateral animals. A predicted protein interaction motif in the PGRMC1 MIHIR is potentially regulated by Y139 phosphorylation. A multilayered model of animal MAPR function acquisition includes some pre-metazoan functions (e.g., heme binding and cytochrome P450 interactions) and some acquired animal-specific functions that involve regulation of strongly conserved protein interaction motifs acquired by early-branching animals. Conclusions: This study provides a conceptual framework for future studies, against which PGRMC1's multiple functions can perhaps be stratified and functionally dissected. In accompanying papers we show that mutational perturbation of PGRMC1 phosphorylation status of the Y180 motif is associated with dramatic changes cell pasticity assayed by protein abundances, cell morphology, mitochondrial function, genomic stability, and epigenetic status, with pathways analysis associating Y180 mutation with processes related to organizer function. These combined works reveal

previously unrecognized involvement of PGRMC1 in foundational animal processes of great relevance to disease.

KEY WORDS

Membrane-associated progesterone receptor, multicellularity, Phylogeny, organizer, opisthokont, protein evolution, tyrosine phosphorylation, Holozoa

BACKGROUND

Progesterone receptor membrane component 1 (PGRMC1) is the archetypal member of the membrane associated progesterone receptor (MAPR) family [1, 2]. Vertebrates including humans encode four MAPR proteins. PGRMC1 and the closely related PGRMC2 arose by gene duplication of an original 'PGRMC' gene during vertebrate evolution. We refer here to PGRMC1 and/or PGRMC2 for proteins from vertebrates that have inherited this gene duplication [3, 4], or otherwise to PGRMC. Other vertebrate MAPR proteins are neuron-derived neurotrophic factor, commonly known as neudesin (NENF), and neuferricin (NEUFC) [3-6].

PGRMC1 has a long list of seemingly disparate functions, ranging from involvement in steroid and heme synthesis, membrane trafficking, progesterone response in fertility and other situations, and conferral of progesterone-dependent anti-apoptosis [7]. PGRMC1 guides embroyonic axonal growth in nematodes and mammals [8, 9] (i.e. perhaps in all bilaterally symmetrical animals), is expressed in a variety of neurons of the central nervous system (CNS) [10-13], and is found in synpases where it affects membrane trafficking [14]. NENF is present in CNS regions of embryonic neural differentiation [15, 16]. In the CNS its expression pattern suggests an exclusive role in neurons, while *in vitro* NENF exhibits strong neurotrophic activity as a secreted protein [16]. Like all vertebrate MAPR proteins, NEUFC is implicated with cytochrome P450 reactions, steroidogenesis, and neurobiology [5, 17-19].

All MAPR proteins contain an insertion of an oligopeptide sequence between helices 3 and 4 of the canonical cytochrome b_5 (cytb$_5$) domain fold (as defined by human cytb$_5$). In place of a short loop at this position in classical cytb$_5$ domain proteins, MAPR proteins contain a MAPR interhelical insertion region (MIHIR) of variable length [1, 4]. The

MIHIR of PGRMC1 and PGRMC2, but not of NENF or NEUFC, contains a tyrosine (PGRMC1 Y139) that is strongly conserved in later-branching animals [4].

Interest in PGRMC1 phosphorylation was sparked when it was found to exist in different phosphorylated forms in breast cancers that were positive or negative for estrogen receptor expression [20]. Bioinformatics revealed the presence of two Src homology 2 (SH2) and one Src homology 3 (SH3) domain target motifs (short peptide sequences that would bind to a much larger SH2 or SH3 protein domain, respectively: hereafter SH3 and SH2 motifs), being a proline-rich SH3 motif centered on PGRMC1 P63, and SH2 motifs centred on Y139 and Y180 [2, 21]. Notably, the P63 SH3 motif and the Y180 SH2 motif were flanked by consensus casein kinase 2 (CK2) sites with phosphoacceptors at S57 and S181. CK2 is constitutively active in many cells, contributing to the order of 20% of the human phospho-proteome [22, 23]. This suggested a model where CK2 phosphorylation could sterically prevent interactions of kinases or other interaction partners with the motifs at P63 and Y180, thereby negatively regulating PGRMC1 function [2, 20]. Mutation of both of the CK2 sites, but not each individually, rendered MCF-7 breast cancer cells resistant to peroxide-induced cell death [20]. However, a recent knockout of CK2 activity in C2C12 mouse muscle cells revealed marginally higher phosphorylation of S181 (and Y180), clearly showing that a kinase other than CK2 can target S181 [24].

Phosphorylation of both residues at CK2 consensus sites and of the Y139 and Y180 SH2 motifs, as well as a variety of other residues, has been observed from high throughput proteomics studies [25]. Phylogenetic analysis revealed that PGRMC1 acquired signalling and phosphorylation motifs during animal evolution: e.g., the PGRMC1 SH3 motif is absent from PGRMC2, and was gained by terrestrial tetrapods. The adjacent S57 phosphorylation site appeared during primate evolution [4]. It was

previously incorrectly concluded that the ancestral metazoan appears to have possessed cognates of both the PGRMC1 Y139 and Y180 SH2 target motifs [4]. As we demonstrate here, this incorrect conclusion was due to mis-assignment of several early-branching metazoan MAPR proteins to the PGRMC family.

A CK2 consensus site adjacent to Y180 arose in the common ancestor of Bilateria [4], reflecting an embryological state before vertebrate body plan is determined. This is of particular interest to early animal evolution and embryology because, 1) ligands and receptors of the Wnt pathway evolved in early animals [26, 27], and PGRMC1 regulates this pathway in mammalian pluripotent stem cells [28]; 2) Both nematode MAPR proteins are expressed through early nematode embryogenesis from the oocyte stage until the induction of germline segregation and neural differentiation [29]; 3) PGRMC1 is implicated in essential progesterone (P4)/progestin responses in male [30-32] and female [21, 33, 34] germline and reproductive cells; 4) PGRMC-like proteins direct ventral embroyonic neural axon migration conserved between nematodes and mammals [8, 9]; 5) PGRMC1 Y180 phosphorylation was observed only in synaptic fractions of mouse neurons [35]; and 6) PGRMC1 is involved in synaptic membrane trafficking [14], implicating a role of PGRMC1 phosphorylation in the synaptic signaling that serves a key organismic coordination role in all animals with a nervous system.

It remains unclear which function(s) may be regulated by PGRMC1 phosphorylation. PGRMC1 is a multifunctional protein [2, 7]. Heme-binding and cytochrome P450 (cyP450) modulation are properties attested from protist MAPR proteins [6, 36-38]. We reasoned that the plethora of PGRMC1 functions should be separable into ancient eukaryotic roles (such as cyP450 interaction) and newly acquired metazoan roles, such as hypothesized tyrosine phosphorylation-mediated signaling in animals. In the present

study we examine the nature of MAPR diversity in early-branching animals as well as in unicellular lineages that represent the closest relatives of animals, with particular interest in the origins of PGRMC1 functional SH2 motifs in opisthokonts.

The Opisthokonta are a eukaryotic supergroup forming two lineages, the Holozoa and the Holomycota [39]. While the Holomycota include fungi and their relatives (such as nucleariids), the Holozoa consists of animals together with their closely related unicellular sister lineages (choanoflagellates, filastereans, pluriformeans and ichthyosporeans) (Figure 1A). Tyrosine kinases and SH2 domains (which bind to phosphorylated tyrosine residues) evolved in those Holozoan unicellular animal relatives, as well as many other proteins normally associated with animals [27, 40-42], including new transcription factors, cell surface adhesion molecules, transposons, and extracellular matrix [43].

A previous pilot study [4] of the phylogenetic distribution of animal PGRMC1 suffered from taxonomic bias against early-diverging animals and their unicellular relatives due to insufficient taxon sampling, especially in critical transitions of early animal evolution, and from poor discrimination between MAPR family members in early-branching organisms. The guiding motivation of the present study was to identify the members of the MAPR protein families present in early-branching animals and their protistan relatives, aiming to understand the changes in MAPR proteins, evolutionarily conserved regions of importance, and particularly to better define the acquisition of PGRMC1 functions.

RESULTS

The origin of animal MAPR protein families predates Opisthokonta

The relationships of the major opisthokont groups investigated in our analysis are depicted schematically in Figure 1A based upon published phylogenies [44, 45], acknowledging that the tree of Opisthokonta is not resolved (see the legend of Figure 1A for details). Since metazoan evolution has been associated with the evolution of many new membrane proteins relative to intracellular enzymes [46], we were interested in the evolution of individual MAPR protein families, particularly in early-branching animals. We obtained diverse opisthokont MAPR proteins across a wide selection of taxa and investigated their phylogenetic relationships. The unrooted tree topology resulting from our MAPR analysis indicated the presence of three well-supported major branches of opisthokont MAPR proteins, representing PGRMC, NENF and NEUFC proteins [6], with members of each family also present in fungi (Figure 1B and Figure S1). Not all species sampled possessed examples of each gene family (e.g., both yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* contain only one MAPR protein: Dap1, a PGRMC family protein). However, this result clearly reveals that the three MAPR families were already distinct in the common ancestor of opisthokonts. Below we describe in detail the characteristics and the evolution of each MAPR protein family.

The consensus Logo plots of the MAPR domains for each individual family from Figure 1B, as well as for all MAPR sequences from all families combined are shown in Figure 2. Human PGRMC1 sequence is shown above the plots for orientation. The location of the MAPR domain in each of the human MAPR family proteins is schematized in Figure 3. Changes to each protein family, particularly along the transition from unicellular to multicellular holozoans as well as non-bilaterian animals to bilatieria, are discussed below. These include changes to the C-termini of each protein family (Figure 3).

The PGRMC family

Notably, PGRMC proteins feature prominent F106, P109 and P112 (with reference to human PGRMC1) around the heme-binding pocket of the MAPR domain (Figure 2), suggesting unique ligand properties of this family. Choanoflagellates, the closest unicellular sister group to animals, possess an apomorphic PGRMC-specific insertion between PGRMC1 G83 and V84 (Figure 2).

N-terminus: There was no observable systematic change of PGRMC protein size N-terminally to the MAPR domain between early- and late-diverging opisthokonts (not shown). PGRMC1 residues 47-49 encode a putative RGD protein interaction motif which has been present at least since the emergence of placental mammals (Figure 4).

MAPR domain: In the transition from unicellular holozoans to cnidarians the PGRMC MAPR domain acquired a frequently represented N-terminal KKR consensus corresponding to PGRMC1 69-71. The two placozoan sequences, cnidarians and bilaterians all featured increasing frequency of K96 and K137, which are ubiquitinated in mammals. K96 is also acetylated (Phosphosite [47], action?id=5744). The frequency of tyrosine at the position of SH2 motif Y139 increases markedly in cnidarian and bilaterian animals (Figure 3, Figure S2).

C-terminal extension to the MAPR domain: Relative to unicellular Holozoa, Porifera and Ctenophora, a C-terminal extension beyond the PGRMC MAPR domain is present in placozoans, cnidarians and bilateral animals, consistent with closer affiliation of placozoans to later-branching metazoans than to Ctenophora and Porifera [48, 49]. A conspicuous conserved feature corresponds to the terminal 192-195 RKND motif of PGRMC1, which was already recognisable in both placozoan sequences available, and which featured strongly in cnidarian and bilaterian proteins (Figure S2).

The most prominent characteristic of the bilaterian PGRMC C-terminus corresponds to the PGRMC1 TxYSxDDE motif surrounding Y180, where phosphorylation of T and S is postulated to sterically impede Y phosphorylation and/or access of SH2 domain proteins

to phosphorylated Y180. Although this has not been formally proven, no doubly phosphorylated peptides have been reported in the Phosphosite database [47], suggesting mutually exclusive rather than cooperative phosphorylation at these sites. Cnidaria lack T178 but commonly possess an acidic stretch C-proximally to the cognate of Y180, also commonly including proximal C-terminal S and T potential sites of phosphorylation, resembling the CK2 consensus site of S181 (Figure S2). We propose that a novel functional antecedent to the PGRMC1 Y180 motif evolved in the common cnidarian/bilaterian ancestor.

K193 at the PGRMC1 C-terminus is a consensus SUMOylation site [4], which predates divergence of cnidarians and bilaterians (Figure S2). SUMOylated proteins are frequently nuclear [50], perhaps hinting that the occasional nuclear localisation of PGRMC1 reflects an evolutionarily novel function acquired by early-diverging animals that is conserved in bilateral animals.

The PGRMC1 Y139 and Y180 combination has been conserved since Cnidaria

Although synapses are present in both Cnidaria and Ctenophora [51], the groups are thought to have independently evolved neurons. Cnidarian and Bilaterian synapses are thought to have evolved from a common ancestor [49]. In our study, Cnidarians were the earliest-diverging animals to acquire the combination of the cognates of PGRMC1 Y139 and Y180, and these are strongly conserved across the Bilateria indicating synapomorphic evolutionary appearance in the common ancestor of those groups. This combination corresponds closely to the presence of ubiquitinated K96 and K137 and consensus SUMOylation site K193 mentioned above. Sabbir has recently demonstrated the inducible presence of PGRMC1 phosphorylation as well as SUMOylation and ubiquitination that affected PGRMC1 stability as well as Tcf/LEF transcriptional activation [52]. By employing sequences of earlier-branching animals (sponges,

ctenophores and placozoans) and 88 cnidarian PGRMC sequences we can conclude with high certainty that Y180 arose in the Cnidaria/Bilateria common ancestor not shared with earlier-diverging animals. The position of Y139 was commonly a W in many unicellular holozoans, with occasional examples of substitution for Y in Choanozoa, Ctenophora and Placozoa, however the Y139/Y180 combination provided new animal-specific functionality to PGRMC1 in Cnidaria/Bilateria because it is strongly conserved (Figure S2). Y139 is the less strongly conserved of these two residues, suggesting its function is not as critical to animal biology.

Animal PGRMC MIHIR regions are predicted to form a coiled-coil protein interaction region

In order to explore possible function of the SH2 motif Y139-containing PGRMC1 MIHIR region we performed low stringency BLAST using human PGRMC1 MIHIR sequence as the search string. Standard BLAST parameters returned only PGRMC1 and PGRMC2 with 100% and 63% similarity and BLAST scores of 82 and 53. However, low stringency BLAST also recognised Neudesin as the next top hit with 48% similarity and a score of 28, followed by a long list of proteins including multiple myosins with BLAST scores above 25. The region of best alignment (PGRMC1 133-164) was then BLASTed against animal species, revealing similar myosin-like motifs from sponge, insects and chordates (Figure 5a). This region was found to have partial predicted coiled-coil character in both PGRMC1 and PGRMC2 (Figure 5b). α-Helical coiled coils are protein-folding and -interaction motifs in which two or more α-helical chains interact to form bundles, typically involving amino acids that exhibit heptad repeats to align on the protein-interactions side of each α-helix. As such, they are associated with protein-protein interactions [53, 54]. The corresponding motif of Myosin 10 was in the coiled-coil region of the protein (Figure 5c). For both PGRMC and myosin motifs,

predicted coiled-coil probability was higher in the N-terminal portion of the motif, and reduced in the C-terminal residues.

We analyzed the probability for coiled-coil formation of this MIHIR motif among selected MAPR proteins by summing the single digit coiled-coil probability scores for each residue in the homologous motif (Figure S3). The motif showed higher propensity for coiled-coil in PGRMC and NENF, with neglibile levels in NEUFC species. However, coiled-coil probability was not a consistently conserved feature of PGRMC or NENF proteins across species (Figure S3). These data are suggestive of protein-protein interactions occurring through the MIHIR region, possibly via coiled coil interactions in vertebrate PGRMC proteins. Lack of coiled-coil formation does not argue against mediation of protein interactions by the MIHIR. There is no requirement for coiled-coil formation to enable functional protein interactions, as long as the respective interaction surfaces co-evolved compatibly in any given species.

NENF family

N-terminus: The NENF MAPR domain is quite proximal to its transmembrane domain/signal peptide, with no evident systematic patterns observed from early- to late-diverging opisthokonts between transmembrane helix and MAPR domain (Supplemental Information File 2).

MAPR domain: In our phylogenetic reconstruction, Holomycota plus Ichthyosporea formed a well-supported clade separate from the remaining Holozoa. Therefore we have denoted all sequences from this clade as “NENF-like”. The major differences between NENF-like and NENF MAPR proteins are color-coded in Figure S4. With regards to accepted rooted phylogeny of the groups concerned (Figure 1A), NENF-like proteins appear to represent an ancestral/pleisiomorphic state for the NENF supergroup. Therefore, the animal-like NENF MAPR domain status represents a synapomorphy that

appeared after the common ancestor of icthyosporeans and metazoans (absent from Icthyosporea, present in Choanozoa, Figure S4) from where it was inherited by animals.

C-terminus: Cnidaria/Bilateria NENF proteins have acquired a C-terminal extension relative to Porifera, Choanozoa, and early-branching metazoans (Figure 3, Figure S4). The ProteinPredict server predicts a protein interaction region between residues 145-150 (Figure S5). There is no sequence similaritiy between the C-terminal extensions of PGRMC and NENF proteins, indicating probable pronounced functional divergence and specialization of these proteins during early metazoan evolution.

NEUFC family

MAPR domain: There are several conspicuous developments during the evolution of NEUFC. Ctenophora, Cnidaria and bilateria acquired a common histidine at H72 (human NEUFC numbering) in the vicinity of the heme binding pocket, suggesting altered ligand-binding. Animals but not single-celled Holozoa have a greater probability of aspartate at D103, and relative to choanozoa, animals possess a two-residue extension at the C-terminus of the MAPR domain including highly conserved G135 (Figure S6). This region corresponds to a site of predicted protein intetraction (Figure S7).

C-terminus: Like all MAPR families, the NEUFC family acquired a C-terminal extension during opisthokont evolution. Unlike PGRMC and NENF families, the NEUFC C-terminus is elongated already in choanoflagellates (Figure 3), relative to earlier-branching holozoans and also fungi. Elements of the evolutionarily newly acquired NEUFC C-terminus are strongly conserved between choanoflagellates and all animals surveyed (Figure S8), implying that this region plays a necessary role in the organismal biology of later-diverging holozoans. It is the largest conserved C-terminus of the animal MAPR family.

Various sites of protein interaction were predicted in the C-terminus by ProteinPredict, as well as a predicted solvent-exposed helix from approximately residues 150-170 (Figure S7A-B). That helical region exhibited a high probability of coiled-coil formation in some but not all NEUFC species sampled (Figure S7C-E). The overall conservation of those residues seems to be more associated with charged residues rather than heptad hydrophobic residues required for coiled-coils. Choanozoans exhibit what appears to be a classical evolutionary intermediary stage between the lower holozoan and fungal state on the one hand, and that of animals on the other (Figure S8). In summary it is highly likely that the NEUFC C-terminus is involved in protein interactions through solvent-accessible residues, however further studies will be required to determine the nature of such interactions, and shed light on the function of NEUFC.

DISCUSSION

A major finding of this study is that all three animal MAPR families had already diverged in the last common ancestor of the opisthokonts. We detected NENF genes in Choanozoa, unlike Ren at al. [3], demonstrating that all three gene families were present in the common ancestor of Choanozoa and animals. Indeed, Ren et al. concluded that the sequence repertoire for animal MAPR genes likely arose in an ancestral animal sequence, which our results refute. While unicellular opisthokonts with all three genes are rare, one such lineage must have proliferated from the ancestral Opisthokont to give rise to choanoflagellates and animals. Loss of genes in particular lineages is a common observance in opisthokont evolution. For instance, although a common ancestor of the divergent yeasts *S. cerevisiae* and *S. pombe* must have possessed PGRMC, NENF and NEUFC family genes, because the common ancestor of all fungi must have (all three MAPR sub-families are found in the Holomycota (Figure 1B)), both extant organisms possess only a single PGRMC-related Dap1 gene (this study).

Strikingly, the extension of the C-termini of all three MAPR families appeared either before the choanoflagellate divergence or at some point in early animal evolution. For NEUFC the origins of this extension occurred prior to the emergence of the choanoflagellates, the sister group to animals. The PGRMC C-terminus gained an extension before the divergence of Placozoa and Cnidaria, while for NENF this occurred with the Cnidaria/Bilateria common ancestor. There is no sequence similarity between these C-terminal extensions, so this phenomenon represents a further functional divergence between the three MAPR proteins of early animals, associated with the transition to multicellularity and increased organismal complexity. Because little is known about the functions of MAPR proteins, or functional differences between family members [7], cellular functions cannot be ascribed to these novel animal-specific MAPR regions.

PGRMC1/2 tyrosine phosphorylation motifs

Another major finding of this study is that the combination of SH2 motif phosphoacceptor residues Y139 and Y180 first appeared in the common ancestor of Cnidaria and Bilateria, being absent from Porifera, Ctenophora and Placozoa (Figure S2). That ancestor was among the first animals to possess neurons with Bilaterian-like synapses [49]. PGRMC1 is Y-phosphorylated in synapses [35], and affects synaptic function [14]. The common Cnidaria/Bilateria ancestor, which was probably bilaterally symmetrical, evolved an organizer capable of inducing differentiation of surrounding cells to define embryological body axes and tissue identities that involved gastrulation at an animal pole. That set in motion an orchestrated set of events involving the induced expression of conserved transcription factors such as *brachyury*, *goosecoid* and *foxa* [55]. This is reflected by organizers recognized, for instance, in Cnidaria [56, 57], and Bilateria including Planarian

flat worms [58], arthropods [59], spiralian protostomes [60], and of course the deuterostome/chordate Spemann-Mangold organizer [61, 62].

All of the Bilateria *sensu stricto* posses the Y180 motif with adjacent T178 and S181 (Figure S2), all of which residues can be phosphorylated in mammals [25]. Because of the evolutionary appearance of this motif at the same time as the rules governing vertebrate embryological cell-type and tissue differentiation became established, we predict that inappropriate alterations in the phosphorylation status of Y180 in PGRMC1 (or PGRMC2) could impose profound effects on human cells, and therefore could be of potentially monumental clinical importance. Tyrosine phosphorylation is typically caused by induced signal transduction pathways. The signal systems surrounding the regulation of Y180 are likely to feature prominently in human disease.

Cnidarians such as *Hydra* possess a Wnt-dependent head organizer which drives axis specification through a protein gradient [63-65]. The Spemann-Mangold organizer also specifies vertebrate dorso-ventral axis via Wnt-signaling [66]. PGRMC1 is involved in the maintenance of human embryonic stem cell pluripotency via regulation of the Wnt pathway [28]. Based upon the strongly conserved coincident presence of the Y139 and Y180 motifs shared between Bilateria and Cnidaria we hypothesize that PGRMC1 phosphorylation might also play an important role in the cnidarian organizer. Furthermore, PGRMC1 may ancestrally be involved in the transition from the single morphotype protist state to the development of the collective of multiple states of differentiated morphology that is characteristic of the clonal metazoan condition, where the ability to phosphorylate Y139/Y180 could have been associated with the evolutionary origin of organizer function and tissue differentiation. PGRMC1 is known to be SUMOYlated which affects TCF/LEF-driven transcription [52, 67]. TCF/LEF is

probably the exclusive conduit for Wnt signalling [68]. Taken together, PGRMC1 is strongly implicated in the evolution of animal organizer function with profound potential to influence animal cell differentiation status and its plasticity (e.g. cancer) [7, 52, 69].

In accompanying papers we show that mutational manipulation of the phosphorylation status of the Y180 motif dramatically affects cultured cancer cell morphology, PI3K/Akt signaling activity, mitochondrial form and function, glycolytic/energy metabolism, and ability to form tumours in MIA PaCa-2 pancreatic cancer cells. This is reflected in altered cell metabolites, genomic stability, and dramatic changes in genomic CpG methylation [70, 71], which are all consistent with PGRMC1 invovelment in organizer biology.

PGRMC and NENF involvement in neurology

In addition to the combination of PGRMC1 Y139/Y180 discussed above, the C-terminal extension of NENF was also acquired in the Cnidaria/Bilateria common ancestor, coincident with the appearance of neurons with synapses. NENF neurotrophism [16] and PGRMC1 axonal guidance [8, 9] are at least superficially similar functions.

For reasons such as these we were expecting to find that animals inherited a single MAPR gene from their unicellular ancestors, which diverged into three families during metazoan evolution, with e.g. ancestral neural-related functions having undergone functional specialization following gene duplication. While this manuscript was in preparation Ren at al. [3] concluded that choanoflagellates and the first metazoans contained only PGRMC and NEUFC genes. The findings that all three MAPR families were already distinct in the ancestral opisthokont strongly negates both hypotheses. However, the observation remains that PGRMC1 and NENF both acquired new conserved functional features at the emergence of Cnidaria/Bilateria common ancestor, correlating with the evolution of

nerve synapses, despite having been inherited into animals as separate genes from a unicellular ancestor. There seems to have been some feature of MAPR biology which was important in the ancestors of animals and which favoured the subsequent evolution of a cnidarian/bilaterian grade of body organization that included synapses. We hypothesize that at least PGRMC1 Y180 was crucial in the latter.

In this light, it is interesting that in NENF knockout mice Novais et al. [72] observed decreased hippocampal levels of the steroids progesterone and allopregnenolone (APα), and increased levels of 3-oxo-5-alpha-steroid 4-dehydrogenases 2 and 3 (which convert progesterone to APα) and gamma-aminobutyric acid A (GABA) receptor delta (GABRD), an important neuromodulator which is positively allosterically regulated by APα [73]. From yeast (phylogenetically disparate *S. pombe* and *S. cerevisiae*) to mammals PGRMC1-like proteins interact with lanosterol-14-demethylase (CYP51A1) [5, 36, 74], the first enzyme in sterol metabolism leading to animal cholesterol. This situation is consistent with a model where PGRMC proteins in unicellular opisthokonts may have been involved in sterol synthesis pathways for the production of 3-ketosterones such as progesterone, whereas NENF may have regulated the conversion of 3-ketosterones to other biologically active steroids, both of which were required for animals to evolve. i.e. where a PGRMC/NENF gene divergence led to new steroid-associated functions, albeit that we are unaware of 3-keto-steroids being attested in earlier diverging opisthokonts. This hypothesis would unite both these MAPR protein functions with an ancestral role in steroid biology. Consistently with a possible ancestral steroidogenic role of MAPR proteins, NEUFC attenuation led to reduced levels of CYP51A1 in HeLa cells, the same enzyme with which PGRMC1 interacts [5, 75], reinforcing a proposed ancestral association between MAPR proteins and sterol biology. The hypothesis predicts that MAPR gene divergence in the opisthokont lineage leading to animals involved functional

diversification towards new pathways of sterol production, among other outcomes including the multiple attested functions of PGRMC1 (reviewed elsewhere [7]).

Is the MIHIR a protein-interaction motif?

We identified a motif with probable coiled-coil characteristics in the MIHIR sequence of both PGRMC1 and PGRMC2, as well as similar sequences in some myosins. The motif was predicted with high probability to form a short coiled-coil at its N-terminus, and lower probability at its C-terminus. Of particular interest was the observation that Y139 formed part of the hydrophobic heptad repeat required for interaction of adjacent helices in the coiled-coil. In lower Holozoans Y139 was commonly a tryptophan, another large bulky hydrophobic residue. The putative involvement of Y139 in a coiled-coil interaction presents immediate connotations when the residue is phosphorylated. Not only would it be unable to interact with coiled-coil interaction partners, but it could then interact with a new set of SH2 domain-containing proteins in the tyrosine-phosphorylated form. This suggests the acquisition of a regulatory switch in MIHIR functionality by PGRMC genes in the Cnidaria/Bilateria common ancestor.

It should be noted that this region of PGRMC1 in the 4X8Y crystal structure exhibited a high B or scatter factor, indicating relatively poor mapping of electron density to amino acid sequence [76], as indicative of a region of low structural stability in the crystal. We therefore conclude that the MIHIR region should be able to rapidly sample many conformations in solution, which is compatible with availability to form protein-protein interactions.

In myosins, the analogous motif is found within the rod-like coiled-coil region, and represents an area where the probability of coiled coil interaction is diminished, as shown

for myosin 10 in Figure 5b. Such coiled-coil regions with weaker stability, often including disruption of the requisite heptad repeat coiled-coil motif, have been proposed to form sites of binding to potential target interacting proteins [77]. A phylogenetic survey of some MAPR proteins from opisthokonts (Figure 5f-g) showed that high predicted probability of coiled-coil formation was conserved in animals (except the sea anemone *Aiptasia pallida*), as well as in *Monosiga brevicollis* from the choanoflagellate sister group of animals. The degree of coiled-coil propensity was weaker in earlier-branching single celled opisthokonts, being minimaly present in the yeast *S. pombe* (Figure 5f-g). We predict that the MAPR MIHIR sequence, a highly conserved early eukaryotic innovation, enables protein interactions not shared with other cytb$_5$ proteins, which may or may not involve coiled-coil formation. Association with the actin/myosin cytoskeleton through the MIHIR motif may be related to the membrane-trafficking functions of PGRMC1.

A PGRMC1 RGD integrin-interaction motif?

We observe a potential vertebrate RGD motif in what is conventionally considered the cytoplasmic region of PGRMC1 of some vertebrates, being absent in the turtles, birds and marsupials studied (Figure 4). This assignment is tenuous because RGD motifs are predicted to be involved with extracellular integrin interactions [78]. Integrins are important in many processes, including the regulation of synaptic plasticity and memory [79]. The RGD motif would need to be extracellular to be functional. However, the conventionally cytoplasmic C-terminal region of PGRMC1 is extracellular in several attested situations, including synapses, where PGRMC1 is involved in a mechanism that affects synaptic plasticity [14], in pluripotent stem cells [80], where PGRMC1 is associated with the maintenance of pluripotency involving the Wnt/beta-catenin pathway [28], and in cancer cells, where PGRMC1 may be secreted by the exosome pathway [81].

PGRMC1 is documented as an exosomal protein in the Exocarta database (http://www.exocarta.org/) [82], however it remains unclear whether it may be translated with an alternative membrane topology as a transmembrane protein [7, 83]. We surmise that the likelihood of a conserved functional RGD integrin-interaction motif in vertebrates would be much greater if the protein exhibits alternative translational topology as a transmembrane protein, rather than as a secreted protein. In view of the association of PGRMC1 with fertility [34, 84, 85], it is conveivable that this motif is involved in the evolution of post-fertilization vertebrate embryology, perhaps involving the relationship between amniotic sac and eggshell, or other major features related to differences in vertebrate embryology and oocyte/egg biology between these groups.

Positively charged residues cytoplasmically proximal to transmembrane signal peptide helices are critical in determining the orientation of transmembrane proteins [86]. There is potential to alter membrane topology of PGRMC1 by post-translational modification of K44 or R47 during translation. It is known that R47 can be methylated (Phosphosite action?id=5744). This may be relevant to the observation of extracellular PGRMC1 C-terminus in the synaptic extracellular space of neurons [14], potentially via regulated non-conventional orientiation during translation. However we stress that in the absence of validated integrin interaction data, the presence of a functional RGD motif in vertebrate PGRMC1 must be considered cautiously.

Conclusions

We initiated this study to try to provide a systematic platform to understand the reported multifunctionality of PGRMC1. Surprisingly, we discovered that members of all three animal MAPR families had originated prior to the origin of opisthokonts. Although many protist species have lost one or more MAPR genes, the common ancestor of each major

opisthokont branch leading to animals must have possessed all three MAPR families, evidencing a degree of conservation which indicates that each family performs separate unicellular functions that were essential in the evolution of the holozoan and holomycotan subfamilies of Opisthokonts. We are unaware of a major eukaryotic group which does not possess at least one MAPR gene, or is not thought to have ancestrally done so. Some parasitic groups are thought to have secondarily lost MAPR genes [83].

We propose that the attested PGRMC1 membrane-trafficking function may be related to the eukaryotic acquisition of the MAPR-defining MIHIR region. In a separate study (submitted) we show that the MIHIR is a eukaryotic development, and propose that the ancestral MAPR protein was one of the pivotal proteins involved in the development of the first truly eukaryotic cell, with roles in the evolution of steroid biology.

Mitochondria require cholesterol to increase membrane potential across the inner membrane, and many of PGRMC1's functions involve steroid biology [83]. Interaction of the eukaryotic MIHIR with the actin cytoskeleton may have been involved in the origin of membrane trafficking that was necessary to target cytoplasmically-synthesized steroids to the proto-mitochondrion before the endosymbiont could lose genes to the nucleus. Additional or parallel to such ancient roles, we hypothezise that the early evolutionary diaspora of eukaryotic diversity involved functional specialisation of at least three opisthokont MAPR genes. The ancestor of Opisthokonta already contained the tripartite MAPR repertoire, although multiple unicellular species have either variously lost MAPR genes or we were not able to find them. The transition from pre-choanozoan to animal multicellularity involved enlargement of the C-terminus in all three MAPR families, commencing with the common ancestors of Bilateria and Placozoa (PGRMC), Cnidaria (NENF), and Choanozoa (NEUFC), respectively (Figure 3). The PGRMC gene acquired the combination of Y139 and Y180 at the stage of evolution of the common ancestor of

Cnidaria and Bilateria, concommittantly with the appearance of synapsed nerves and prior to the evolution of Bilateria *sensu stricto* and the associated new mechanisms for embryological body pattern formation and tissue differentiation. Our study suggests a stratified acquisition of PGRMC1 functions, and points towards potentially dramatic effects of cell differentiation status if this ancient axis of PGRMC1 tyrosine phosphorylation is perturbed in disease processes. Results presented in accompanying papers strongly support this hypothesis [70, 71]. These combined findings signpost the direction for productive future studies, which was our aim.

MATERIALS AND METHODS

Identification of MAPR proteins

The human sequences of PGRMC1, NENF and NEUFC were used in an initial BLASTP search against the NCBI non-redundant protein sequence database (using default NCBI BLASTp settings), retaining the best hit per organism. Additional sequences were identified by using the same queries in BLASTp searches (evalue 1e-25) against a custom database as described [44], and further expanded using newly available unicellular opisthokont datasets (*Parvularia atlantis* and *Chromospahera perkinsii*, available from http://multicellgenome.com) as well as additional fungal datasets (downloaded from https://genome.jgi.doe.gov/mycocosm/home). The dataset was extended for the placozoan *Hoilungia hongkongensis*, three classes of Porifera, Ctenophora, Cnidaria and also Choanoflagellata by BLAST searches against a set of non-bilaterian proteomes that have been previously established [87, 88]. BLASTP searches were performed with specifying an evalue of 1e-10 and otherwise default settings using the human NENF, NEUFC and PGRMC1 as well as the initially identified placozoan *Trichoplax adhaerens* [89] NEUFC and PGRMC protein sequences. Specimens of the calcareous sponge *Pericharax orientalis* were collected from Dunk Island Mission beach in 2016 (under the

authorization CMES59 provided by James Cook University). MAPR sequences of *Pericharax* were retrieved from a draft assembled transcriptome (Adamski et al. in prep). BLASTP searches using mammalian sequences as query allowed the identification of the *P. orientalis* homologs. The MAPR identity of all proteins used was verified by sequence alignment and confirmation of the presence of a MIHIR.

Phylogenetic reconstruction

All MAPR sequences described above were initially aligned with MAFFT v. 7.212 L-INS-i [90]. Ambiguously aligned positions were trimmed off with trimAL v. 1.2 [91] using a gap threshold of 20% and a tree was calculated using FastTree v. 2.1.7 with default options [92]. The resulting phylogeny and the underlying alignment were manually inspected and obvious contaminations, duplicates and paralogs were removed. The cleaned, unaligned sequences were then subjected to filtering with PREQUAL [93] using the default options to remove non-homologous residues introduced by poor-quality sequences, followed by alignment with MAFFT G-INS-i using the VSM option (--unalignlevel 0.6) [94] to control over-alignment. The alignments were subjected to Divvier [95] using the -divvygap option to improve homology inference before removing ambiguously aligned sites with trimAl v. 1.2 (-gt 0.01). We then extracted the region from position 46 to 139 (94 amino acid residues (aa), relative to human PGRMC1), that is conserved in all taxa of our alignment, from the trimmed alignment and performed the tree reconstruction based on this central conserved region only. Final trees were calculated with IQ-TREE v. 1.6.5 [96], using the -mset option to restrict model selection (to DAYHOFF, DCMUT, JTT, WAG, VT, BLOSUM62, LG, PMB, JTTDCMUT; model selected: LG+R5) for ModelFinder [97], while branch support was assessed with 1000 ultrafast bootstrap replicates [98]. We also prepared a conservative phylogeny without performing PREQUAL filtering prior to sequence alignment or Divvier analysis after

alignment, followed by stringent trimming with trimAl v. 1.2 (-gt 0.8), resulting in an alignment of 126 aa length. The tree was calculated using IQ-TREE and ModelFinder as described above (model selected: LG+R7), branch support was assessed with 1000 ultrafast bootstrap replicates.

With both approaches, we observed a similar topology, particularly the conspicuous split within the NENF clade: while cnidaria, sponges and choanoflagellates formed a well-supported holozoan clade with the metazoan NENF representatives, the holozoan lineage of ichthyosporeans grouped with high support with the holomycota (fungi and nucleariids).

The alignment of selected sequences chosen to provide indicative phylogentic representation of land vertebrates in Figure 4 was made with the Computational Biology Research Consortium (CBRC) MAFFT platform (https://mafft.cbrc.jp/) using the L-INS-i strategy [99].

Other database platform queries

The same alignment used for tree reconstruction in Figure 1 has been used to generate logo plots to identify conserved regions in groups of interest. The untrimmed 3-protein-alignment was split into three separate alignments for each protein (PGRMC, NENF and NEUFC), followd by removal of gap-only columns in each alignment. For the alignments in Figures S1-S5, a choanoflagellate-specific insert indicated in Figure 2 has been removed. The species corresponding to phylogenetic groups labelled in Figures S2, S4, S6 and S8 were entered to Weblogo separately. Logo plot representation of consensus sequences were generated with the WebLogo platform (http://weblogo.berkeley.edu/) [100], as described in respective figure legends.

Low stringency protein sequence BLAST was performed using human PGRMC1 search string CLDKEALKDEYDDLSDLTAAQQETLSDWESQFTFKYHH with NCBI BLASTp (https://blast.ncbi.nlm.nih.gov), employing word size 3, expect threshold 1000, organism restricted to Homo sapiens (taxid:9606), 500 maximum target sequences, and with all other parameters set to default values. Subsequent BLAST queries targeted the organisms from Figure 5A. Coiled-coil prediction [101] was performed by the PRABI (Pôle Rhône-Alpes de Bioinformatique) server (https://npsa-prabi.ibcp.fr/). Sites of observed phosphorylation were obtained from Phosphosite (www.phosphosite.org) [102] or UniProt (www.uniprot.org). Protein interaction sites were predicted with the PredictProtein server (www.predictprotein.org) [103]. RGD protein interaction motif detection was by the ISIS (interaction sites identified from sequence) [104] function of ProteinPredict.

LIST OF ABBREVIATIONS

CK2: casein kinase 2

CNS: central nervous system

cyP450: cytochrome P450

cytb$_5$: cytochrome b$_5$

MAPR: membrane-associated progesterone receptor

MIHIR: membrane-associated progesterone receptor-specific interhelical insert region

NENF: neudesin

NEUFC: Neuferricin

P4: progesterone

PGRMC: progesterone receptor membrane component 2

PGRMC1: progesterone receptor membrane component 1

SH2: Src homology 2

SH3: Src homology 3

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The raw tree files in newick, colored trees with taxon information in pdf format, and underlying trimmed alignments corresponding to both phylogenetic reconstructions have been deposited to figshare repository doi: 10.6084/m9.figshare.9162164.

Competing interests

M.A.C. is scientific advisor to and minor shareholder of Cognition Therapeutics, a company developing sigma-2 receptor ligands against Alzheimer's disease. This work was performed independently of and without input from the company. The authors declare that they have no other competing interests.

Funding

M.A.C. has received no direct Australian competitive grant support since relocation to the country in 2008. The present results have been compiled largely due to the generosity of collaborating authors. This work was supported by Charles Sturt University (CSU) School of Biomedical Sciences (SBMS) Compact grant A541-900-xxx-40513, SBMS support A534-900-xxx-41066, and CSU Competitive grant A102-900-xxx-40002, all to MAC. Open access publication fees were jointly supported by CSU's Faculty of Science, SBMS, and Research Office. M.E. acknowledges financial support through the LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative (granted to Gert Wörheide, LMU Munich) and the European

Union's Horizon 2020 Marie Skłodowska-Curie Innovative Training Network IGNITE (grant number 764840).

Authors' contributions

M.A.C. conceived the project, performed WebLogo studies, and wrote the first draft of the manuscript. E.H. performed MAPR sequence identification, sequence alignment and curation and phylogenetic reconstruction with advice from P.J.K. M.E. retrieved additional choanoflagellate, Porifera, Ctenophora, Placozoa and Cnidaria protein sequences. S.A.V.F. sequenced the Pericharax orientalis genome, under supervision from D.J.M. All authors read and provided critical comment to the manuscript.

Acknowledgements

Not applicable.

REFERENCES

1. Mifsud W, Bateman A: **Membrane-bound progesterone receptors contain a cytochrome b5-like ligand-binding domain**. *Genome biology* 2002, **3**(12):RESEARCH0068.
2. Cahill MA: **Progesterone receptor membrane component 1: an integrative review**. *The Journal of steroid biochemistry and molecular biology* 2007, **105**(1-5):16-36.
3. Ren J, Chung-Davidson YW, Jia L, Li W: **Genomic sequence analyses of classical and non-classical lamprey progesterone receptor genes and the inference of homologous gene evolution in metazoans**. *BMC evolutionary biology* 2019, **19**(1):136.
4. Cahill MA: **The evolutionary appearance of signaling motifs in PGRMC1**. *Bioscience trends* 2017, **11**(2):179-192.
5. Ryu CS, Klein K, Zanger UM: **Membrane Associated Progesterone Receptors: Promiscuous Proteins with Pleiotropic Functions - Focus on Interactions with Cytochromes P450**. *Frontiers in pharmacology* 2017, **8**:159.
6. Kimura I, Nakayama Y, Konishi M, Terasawa K, Ohta M, Itoh N, Fujimoto M: **Functions of MAPR (membrane-associated progesterone receptor) family members as heme/steroid-binding proteins**. *Current protein & peptide science* 2012, **13**(7):687-696.
7. Cahill MA, Jazayeri JA, Catalano SM, Toyokuni S, Kovacevic Z, Richardson DR: **The emerging role of progesterone receptor membrane component 1 (PGRMC1) in cancer biology**. *Biochimica et biophysica acta* 2016, **1866**(2):339-349.
8. Runko E, Kaprielian Z: **Expression of Vema in the developing mouse spinal cord and optic chiasm**. *The Journal of comparative neurology* 2002, **451**(3):289-299.
9. Runko E, Kaprielian Z: **Caenorhabditis elegans VEM-1, a novel membrane protein, regulates the guidance of ventral nerve cord-associated axons**. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2004, **24**(41):9015-9026.
10. Zeng C, Garg N, Mach RH: **The PGRMC1 Protein Level Correlates with the Binding Activity of a Sigma-2 Fluorescent Probe (SW120) in Rat Brain Cells**. *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging* 2016, **18**(2):172-179.
11. Bali N, Arimoto JM, Iwata N, Lin SW, Zhao L, Brinton RD, Morgan TE, Finch CE: **Differential responses of progesterone receptor membrane component-1 (Pgrmc1) and the classical progesterone receptor (Pgr) to 17beta-estradiol and progesterone in hippocampal subregions that support synaptic remodeling and neurogenesis**. *Endocrinology* 2012, **153**(2):759-769.
12. Bali N, Arimoto JM, Morgan TE, Finch CE: **Progesterone antagonism of neurite outgrowth depends on microglial activation via Pgrmc1/S2R**. *Endocrinology* 2013, **154**(7):2468-2480.
13. Olbrich L, Wessel L, Balakrishnan-Renuka A, Boing M, Brand-Saberi B, Theiss C: **Rapid impact of progesterone on the neuronal growth cone**. *Endocrinology* 2013, **154**(10):3784-3795.
14. Izzo NJ, Xu J, Zeng C, Kirk MJ, Mozzoni K, Silky C, Rehak C, Yurko R, Look G, Rishton G *et al*: **Alzheimer's therapeutics targeting amyloid beta 1-42 oligomers II: Sigma-2/PGRMC1 receptors mediate Abeta 42 oligomer binding and synaptotoxicity**. *PLoS One* 2014, **9**(11):e111899.

15. Kimura I, Konishi M, Miyake A, Fujimoto M, Itoh N: **Neudesin, a secreted factor, promotes neural cell proliferation and neuronal differentiation in mouse neural precursor cells**. *Journal of neuroscience research* 2006, **83**(8):1415-1424.
16. Kimura I, Yoshioka M, Konishi M, Miyake A, Itoh N: **Neudesin, a novel secreted protein with a unique primary structure and neurotrophic activity**. *Journal of neuroscience research* 2005, **79**(3):287-294.
17. Hasegawa S, Kasubuchi M, Terasawa K, Kimura I: **Perspectives On Membrane-associated Progesterone Receptors As Prospective Therapeutic Targets**. *Current drug targets* 2016, **17**(10):1189-1197.
18. Petersen SL, Intlekofer KA, Moura-Conlon PJ, Brewer DN, Del Pino Sans J, Lopez JA: **Nonclassical progesterone signalling molecules in the nervous system**. *Journal of neuroendocrinology* 2013, **25**(11):991-1001.
19. Petersen SL, Intlekofer KA, Moura-Conlon PJ, Brewer DN, Del Pino Sans J, Lopez JA: **Novel progesterone receptors: neural localization and possible functions**. *Frontiers in neuroscience* 2013, **7**:164.
20. Neubauer H, Clare SE, Wozny W, Schwall GP, Poznanovic S, Stegmann W, Vogel U, Sotlar K, Wallwiener D, Kurek R *et al*: **Breast cancer proteomics reveals correlation between estrogen receptor status and differential phosphorylation of PGRMC1**. *Breast Cancer Research* 2008, **10**(5):R85.
21. Peluso JJ: **Multiplicity of progesterone's actions and receptors in the mammalian ovary**. *Biol Reprod* 2006, **75**(1):2-8.
22. Meggio F, Pinna LA: **One-thousand-and-one substrates of protein kinase CK2?** *FASEB J* 2003, **17**(3):349-368.
23. Salvi M, Sarno S, Cesaro L, Nakamura H, Pinna LA: **Extraordinary pleiotropy of protein kinase CK2 revealed by weblogo phosphoproteome analysis**. *Biochimica et biophysica acta* 2009, **1793**(5):847-859.
24. Franchin C, Borgo C, Cesaro L, Zaramella S, Vilardell J, Salvi M, Arrigoni G, Pinna LA: **Re-evaluation of protein kinase CK2 pleiotropy: new insights provided by a phosphoproteomics analysis of CK2 knockout cells**. *Cellular and molecular life sciences : CMLS* 2018, **75**(11):2011-2026.
25. Cahill MA, Jazayeri JA, Kovacevic Z, Richardson DR: **PGRMC1 regulation by phosphorylation: potential new insights in controlling biological activity**. *Oncotarget* 2016, **7**(32):50822-50827.
26. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U *et al*: **The Amphimedon queenslandica genome and the evolution of animal complexity**. *Nature* 2010, **466**(7307):720-726.
27. Suga H, Chen Z, de Mendoza A, Sebe-Pedros A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sanchez-Pons N *et al*: **The Capsaspora genome reveals a complex unicellular prehistory of animals**. *Nature communications* 2013, **4**:2325.
28. Kim JY, Kim SY, Choi HS, Kim MK, Lee HM, Jang YJ, Ryu CJ: **Progesterone Receptor Membrane Component 1 suppresses the p53 and Wnt/beta-catenin pathways to promote human pluripotent stem cell self-renewal**. *Scientific reports* 2018, **8**(1):3048.
29. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I: **Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer**. *Nature* 2015, **519**(7542):219-222.
30. Losel R, Breiter S, Seyfert M, Wehling M, Falkenstein E: **Classic and non-classic progesterone receptors are both expressed in human spermatozoa**. *Hormone*

and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme 2005, **37**(1):10-14.

31. Correia JN, Conner SJ, Kirkman-Brown JC: **Non-genomic steroid actions in human spermatozoa. "Persistent tickling from a laden environment"**. *Seminars in reproductive medicine* 2007, **25**(3):208-219.
32. Thomas P: **Characteristics of membrane progestin receptor alpha (mPRalpha) and progesterone membrane receptor component 1 (PGMRC1) and their roles in mediating rapid progestin actions**. *Frontiers in neuroendocrinology* 2008, **29**(2):292-312.
33. Thomas P: **Role of G-protein-coupled estrogen receptor (GPER/GPR30) in maintenance of meiotic arrest in fish oocytes**. *J Steroid Biochem Mol Biol* 2017, **167**:153-161.
34. Peluso JJ: **Non-genomic actions of progesterone in the normal and neoplastic mammalian ovary**. *Seminars in reproductive medicine* 2007, **25**(3):198-207.
35. Munton RP, Tweedie-Cullen R, Livingstone-Zatchej M, Weinandy F, Waidelich M, Longo D, Gehrig P, Potthast F, Rutishauser D, Gerrits B *et al*: **Qualitative and quantitative analyses of protein phosphorylation in naive and stimulated mouse synaptosomal preparations**. *Mol Cell Proteomics* 2007, **6**(2):283-293.
36. Hughes AL, Powell DW, Bard M, Eckstein J, Barbuch R, Link AJ, Espenshade PJ: **Dap1/PGRMC1 binds and regulates cytochrome P450 enzymes**. *Cell Metab* 2007, **5**(2):143-149.
37. Hand RA, Craven RJ: **Hpr6.6 protein mediates cell death from oxidative damage in MCF-7 human breast cancer cells**. *J Cell Biochem* 2003, **90**(3):534-547.
38. Mallory JC, Crudden G, Johnson BL, Mo C, Pierson CA, Bard M, Craven RJ: **Dap1p, a heme-binding protein that regulates the cytochrome P450 protein Erg11p/Cyp51p in Saccharomyces cerevisiae**. *Molecular and cellular biology* 2005, **25**(5):1669-1679.
39. Cavalier-Smith T: **The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa**. *International journal of systematic and evolutionary microbiology* 2002, **52**(Pt 2):297-354.
40. Suga H, Torruella G, Burger G, Brown MW, Ruiz-Trillo I: **Earliest Holozoan expansion of phosphotyrosine signaling**. *Molecular biology and evolution* 2014, **31**(3):517-528.
41. Hunter T: **The genesis of tyrosine phosphorylation**. *Cold Spring Harbor perspectives in biology* 2014, **6**(5):a020644.
42. Tong K, Wang Y, Su Z: **Phosphotyrosine signalling and the origin of animal multicellularity**. *Proceedings Biological sciences* 2017, **284**(1860).
43. Grau-Bove X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, Ruiz-Trillo I: **Dynamics of genomic innovation in the unicellular ancestry of animals**. *eLife* 2017, **6**.
44. Hehenberger E, Tikhonenkov DV, Kolisko M, Del Campo J, Esaulov AS, Mylnikov AP, Keeling PJ: **Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals**. *Current biology : CB* 2017, **27**(13):2043-2050 e2046.
45. Borowiec ML, Lee EK, Chiu JC, Plachetzki DC: **Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa**. *BMC genomics* 2015, **16**:987.
46. Attwood MM, Krishnan A, Almen MS, Schioth HB: **Highly diversified expansions shaped the evolution of membrane bound proteins in metazoans**. *Scientific reports* 2017, **7**(1):12387.

47. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B: **PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation**. *Proteomics* 2004, **4**(6):1551-1561.
48. Brunet T, King N: **The Origin of Animal Multicellularity and Cell Differentiation**. *Developmental cell* 2017, **43**(2):124-140.
49. Moroz LL, Kohn AB: **Independent origins of neurons and synapses: insights from ctenophores**. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2016, **371**(1685):20150041.
50. Ovaa H, Vertegaal ACO: **Probing ubiquitin and SUMO conjugation and deconjugation**. *Biochemical Society transactions* 2018, **46**(2):423-436.
51. Ovsepian SV: **The birth of the synapse**. *Brain structure & function* 2017, **222**(8):3369-3374.
52. Sabbir MG: **Progesterone induced Warburg effect in HEK293 cells is associated with post-translational modifications and proteasomal degradation of progesterone receptor membrane component 1**. *J Steroid Biochem Mol Biol* 2019, **191**:105376.
53. Woolfson DN: **Coiled-Coil Design: Updated and Upgraded**. *Sub-cellular biochemistry* 2017, **82**:35-61.
54. Woolfson DN, Bartlett GJ, Bruning M, Thomson AR: **New currency for old rope: from coiled-coil assemblies to alpha-helical barrels**. *Current opinion in structural biology* 2012, **22**(4):432-441.
55. Genikhovich G, Technau U: **On the evolution of bilaterality**. *Development (Cambridge, England)* 2017, **144**(19):3392-3404.
56. Hayward DC, Grasso LC, Saint R, Miller DJ, Ball EE: **The organizer in evolution-gastrulation and organizer gene expression highlight the importance of Brachyury during development of the coral, Acropora millepora**. *Developmental biology* 2015, **399**(2):337-347.
57. Kraus Y, Aman A, Technau U, Genikhovich G: **Pre-bilaterian origin of the blastoporal axial organizer**. *Nat Commun* 2016, **7**:11694.
58. Sureda-Gomez M, Adell T: **Planarian organizers**. *Seminars in cell & developmental biology* 2019, **87**:95-104.
59. Oda H, Iwasaki-Yokozawa S, Usui T, Akiyama-Oda Y: **Experimental duplication of bilaterian body axes in spider embryos: Holm's organizer and self-regulation of embryonic fields**. *Development genes and evolution* 2019.
60. Henry JQ, Lyons DC, Perry KJ, Osborne CC: **Establishment and activity of the D quadrant organizer in the marine gastropod Crepidula fornicata**. *Developmental biology* 2017, **431**(2):282-296.
61. Nielsen C, Brunet T, Arendt D: **Evolution of the bilaterian mouth and anus**. *Nature ecology & evolution* 2018, **2**(9):1358-1376.
62. Lapraz F, Haillot E, Lepage T: **A deuterostome origin of the Spemann organiser suggested by Nodal and ADMPs functions in Echinoderms**. *Nature communications* 2015, **6**:8434.
63. Hobmayer B, Rentzsch F, Kuhn K, Happel CM, von Laue CC, Snyder P, Rothbacher U, Holstein TW: **WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra**. *Nature* 2000, **407**(6801):186-189.
64. Lengfeld T, Watanabe H, Simakov O, Lindgens D, Gee L, Law L, Schmidt HA, Ozbek S, Bode H, Holstein TW: **Multiple Wnts are involved in Hydra organizer formation and regeneration**. *Developmental biology* 2009, **330**(1):186-199.
65. Bode HR: **The head organizer in Hydra**. *The International journal of developmental biology* 2012, **56**(6-8):473-478.

66. Reid CD, Zhang Y, Sheets MD, Kessler DS: **Transcriptional integration of Wnt and Nodal pathways in establishment of the Spemann organizer**. *Developmental biology* 2012, **368**(2):231-241.
67. Peluso JJ, Lodde V, Liu X: **Progesterone regulation of progesterone receptor membrane component 1 (PGRMC1) sumoylation and transcriptional activity in spontaneously immortalized granulosa cells**. *Endocrinology* 2012, **153**(8):3929-3939.
68. Schuijers J, Mokry M, Hatzis P, Cuppen E, Clevers H: **Wnt-induced transcriptional activation is exclusively mediated by TCF/LEF**. *The EMBO journal* 2014, **33**(2):146-156.
69. Shih CC, Chou HC, Chen YJ, Kuo WH, Chan CH, Lin YC, Liao EC, Chang SJ, Chan HL: **Role of PGRMC1 in cell physiology of cervical cancer**. *Life sciences* 2019.
70. Thejer BM, Adhikary PP, Kaur A, Teakel SL, Van Oosterum A, Seth I, Pajic M, Hannan KM, Pavy M, Poh P *et al*: **PGRMC1 phosphorylation status and cell plasticity 1: glucose metabolism, mitochondria, and mouse xenograft tumorigenesis**. **Accompanying Paper-Citation* ####### 2019-I, #######: bioRxiv preprint (not peer-reviewed): https://doi.org/10.1101/737718.
71. Thejer BM, Adhikary PP, Teakel SL, Fang J, Weston PA, Gurusinghe S, Anwer AG, Gosnell M, Jazayeri JA, Ludescher M *et al*: **PGRMC1 phosphorylation status and cell plasticity 2: genomic integrity and CpG methylation**. **Accompanying Paper-Citation* ####### 2019-II, #######: bioRxiv preprint (not peer-reviewed): https://doi.org/10.1101/737783.
72. Novais A, Silva A, Ferreira AC, Falcao AM, Sousa N, Palha JA, Marques F, Sousa JC: **Adult Hippocampal Neurogenesis Modulation by the Membrane-Associated Progesterone Receptor Family Member Neudesin**. *Frontiers in cellular neuroscience* 2018, **12**:463.
73. Guennoun R, Labombarda F, Gonzalez Deniselle MC, Liere P, De Nicola AF, Schumacher M: **Progesterone and allopregnanolone in the central nervous system: response to injury and implication for neuroprotection**. *J Steroid Biochem Mol Biol* 2015, **146**:48-61.
74. Hand RA, Jia N, Bard M, Craven RJ: **Saccharomyces cerevisiae Dap1p, a novel DNA damage response protein related to the mammalian membrane-associated progesterone receptor**. *Eukaryotic cell* 2003, **2**(2):306-317.
75. Bruce A, Rybak AP: **CYB5D2 requires heme-binding to regulate HeLa cell growth and confer survival from chemotherapeutic agents**. *PLoS One* 2014, **9**(1):e86435.
76. Kabe Y, Nakane T, Koike I, Yamamoto T, Sugiura Y, Harada E, Sugase K, Shimamura T, Ohmura M, Muraoka K *et al*: **Haem-dependent dimerization of PGRMC1/Sigma-2 receptor facilitates cancer proliferation and chemoresistance**. *Nature communications* 2016, **7**:11030.
77. Sunitha MS, Nair AG, Charya A, Jadhav K, Mukhopadhyay S, Sowdhamini R: **Structural attributes for the recognition of weak and anomalous regions in coiled-coils of myosins and other motor proteins**. *BMC research notes* 2012, **5**:530.
78. Nieberler M, Reuning U, Reichart F, Notni J, Wester HJ, Schwaiger M, Weinmuller M, Rader A, Steiger K, Kessler H: **Exploring the Role of RGD-Recognizing Integrins in Cancer**. *Cancers* 2017, **9**(9).
79. Park YK, Goda Y: **Integrins in synapse regulation**. *Nature reviews Neuroscience* 2016, **17**(12):745-756.

80. Kim JY, Kim SY, Choi HS, An S, Ryu CJ: **Epitope mapping of anti-PGRMC1 antibodies reveals the non-conventional membrane topology of PGRMC1 on the cell surface**. *Scientific reports* 2019, **9**(1):653.
81. Mir SU, Ahmed IS, Arnold S, Craven RJ: **Elevated progesterone receptor membrane component 1/sigma-2 receptor levels in lung tumors and plasma from lung cancer patients**. *Int J Cancer* 2012, **131**(2):E1-9.
82. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, Samuel M, Pathan M, Jois M, Chilamkurti N *et al*: **ExoCarta: A Web-Based Compendium of Exosomal Cargo**. *Journal of molecular biology* 2016, **428**(4):688-692.
83. Cahill MA, Medlock AE: **Thoughts on interactions between PGRMC1 and diverse attested and potential hydrophobic ligands**. *The Journal of steroid biochemistry and molecular biology* 2017, **171**:11-33.
84. Engmann L, Losel R, Wehling M, Peluso JJ: **Progesterone regulation of human granulosa/luteal cell viability by an RU486-independent mechanism**. *J Clin Endocrinol Metab* 2006, **91**(12):4962-4968.
85. Peluso JJ: **Progesterone signaling mediated through progesterone receptor membrane component-1 in ovarian cells with special emphasis on ovarian cancer**. *Steroids* 2011, **76**(9):903-909.
86. Hartmann E, Rapoport TA, Lodish HF: **Predicting the orientation of eukaryotic membrane-spanning proteins**. *Proc Natl Acad Sci U S A* 1989, **86**(15):5786-5790.
87. Eitel M, Francis WR, Varoqueaux F, Daraspe J, Osigus HJ, Krebs S, Vargas S, Blum H, Williams GA, Schierwater B *et al*: **Comparative genomics and the nature of placozoan species**. *PLoS biology* 2018, **16**(7):e2005359.
88. Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Queinnec E, Ereskovsky A *et al*: **A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals**. *Current biology : CB* 2017, **27**(7):958-967.
89. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML *et al*: **The Trichoplax genome and the nature of placozoans**. *Nature* 2008, **454**(7207):955-960.
90. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Molecular biology and evolution* 2013, **30**(4):772-780.
91. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses**. *Bioinformatics (Oxford, England)* 2009, **25**(15):1972-1973.
92. Price MN, Dehal PS, Arkin AP: **FastTree 2--approximately maximum-likelihood trees for large alignments**. *PLoS One* 2010, **5**(3):e9490.
93. Whelan S, Irisarri I, Burki F: **PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences**. *Bioinformatics (Oxford, England)* 2018, **34**(22):3929-3930.
94. Katoh K, Standley DM: **A simple method to control over-alignment in the MAFFT multiple sequence alignment program**. *Bioinformatics (Oxford, England)* 2016, **32**(13):1933-1942.
95. Ali RH, Bogusz M, Whelan S: **Identifying clusters of high confidence homologies in multiple sequence alignments**. *Molecular biology and evolution* 2019.
96. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ: **IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies**. *Molecular biology and evolution* 2015, **32**(1):268-274.

97. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS: **ModelFinder: fast model selection for accurate phylogenetic estimates**. *Nature methods* 2017, **14**(6):587-589.
98. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS: **UFBoot2: Improving the Ultrafast Bootstrap Approximation**. *Molecular biology and evolution* 2018, **35**(2):518-522.
99. Katoh K, Rozewicki J, Yamada KD: **MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization**. *Briefings in bioinformatics* 2017.
100. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator**. *Genome research* 2004, **14**(6):1188-1190.
101. Lupas A, Van Dyke M, Stock J: **Predicting coiled coils from protein sequences**. *Science (New York, NY)* 1991, **252**(5009):1162-1164.
102. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E: **PhosphoSitePlus, 2014: mutations, PTMs and recalibrations**. *Nucleic acids research* 2015, **43**(Database issue):D512-520.
103. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Honigschmid P, Schafferhans A, Roos M, Bernhofer M *et al*: **PredictProtein--an open resource for online prediction of protein structural and functional features**. *Nucleic acids research* 2014, **42**(Web Server issue):W337-343.
104. Ofran Y, Rost B: **ISIS: interaction sites identified from sequence**. *Bioinformatics (Oxford, England)* 2007, **23**(2):e13-16.
105. Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, Lartillot N, Worheide G, Pisani D: **Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals**. *Current biology : CB* 2017, **27**(24):3864-3870 e3864.
106. Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM: **Ctenophore relationships and their placement as the sister group to all other animals**. *Nature ecology & evolution* 2017, **1**(11):1737-1746.
107. Torruella G, de Mendoza A, Grau-Bove X, Anto M, Chaplin MA, del Campo J, Eme L, Perez-Cordon G, Whipps CM, Nichols KM *et al*: **Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi**. *Current biology : CB* 2015, **25**(18):2404-2410.
108. Larsson A: **AliView: a fast and lightweight alignment viewer and editor for large datasets**. *Bioinformatics (Oxford, England)* 2014, **30**(22):3276-3278.

Figures

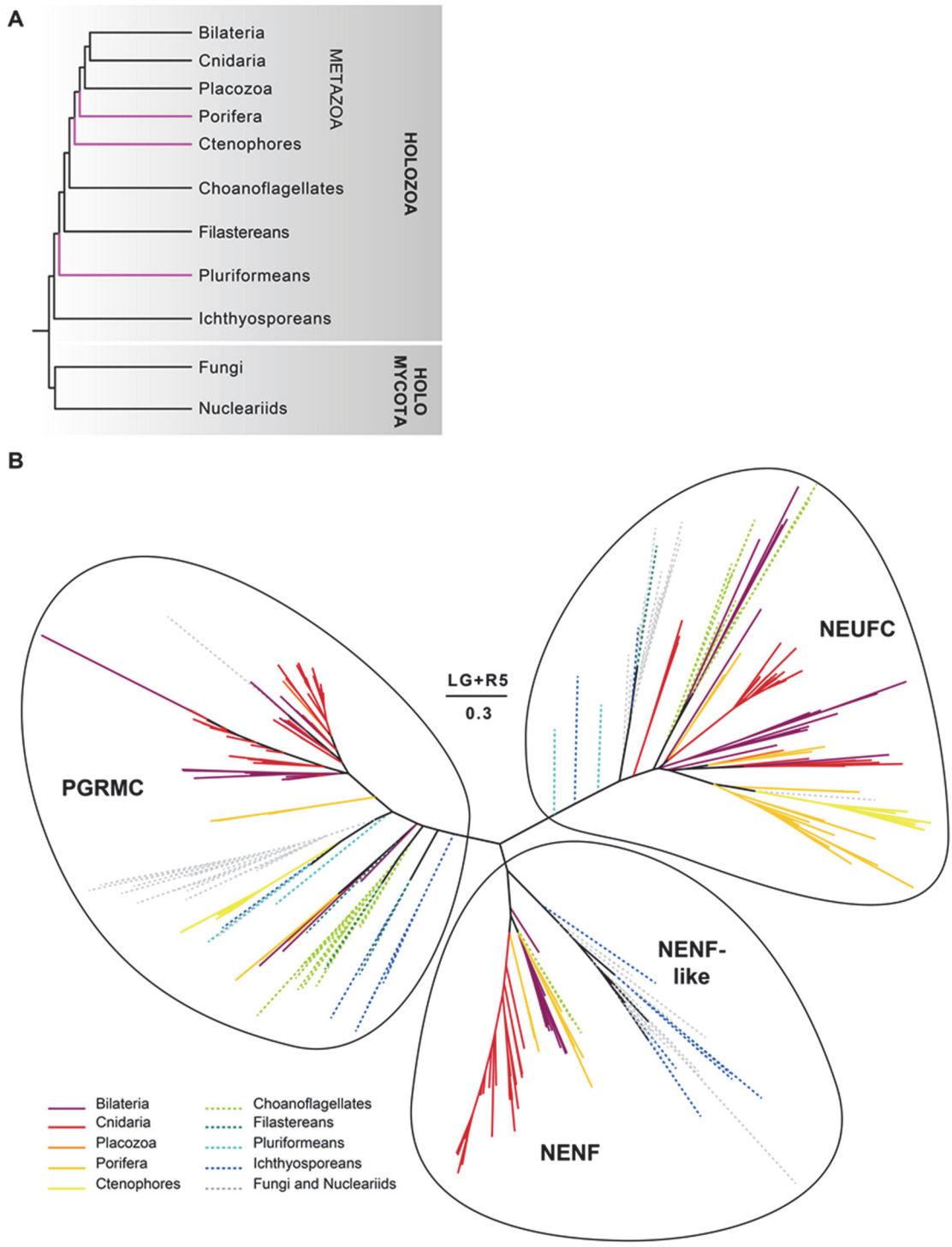


Figure 1. Phylogenetic reconstruction of MAPR proteins in opisthokonts. (A) Schematic tree of opisthokont lineages analyzed in this work, with contentious branches colored in magenta. The branch topology of Ctenophores and Porifera divergence is subject to strong ongoing debate, with both Porifera [88, 105] and Ctenophores [106] argued as forming sister groups to all other animals. Similarly, the Pluriformea

(*Corallochytrium*) were found to branch together with Ichthyosporea as sister to all other holozoans in alternative tree reconstructions [107]. (B) Phylogeny of 3 types of MAPR proteins in opisthokonts: progesterone receptor membrane component (PGRMC), neudesin (NENF and NENF-like) and neuferricin (NEUFC). Solid lines represent metazoan lineages, dashed lines represent non-metazoan lineages. Different lineages are indicated by color in the key. The scale bar and the number beneath it indicate the estimated number of substitutions per site, above the scale bar the model for tree reconstruction is indicated. For bootstrap support see Supplemental Figure 1, for a phylogeny containing also taxon information, see figshare repository doi: 10.6084/m9.figshare.9162164.

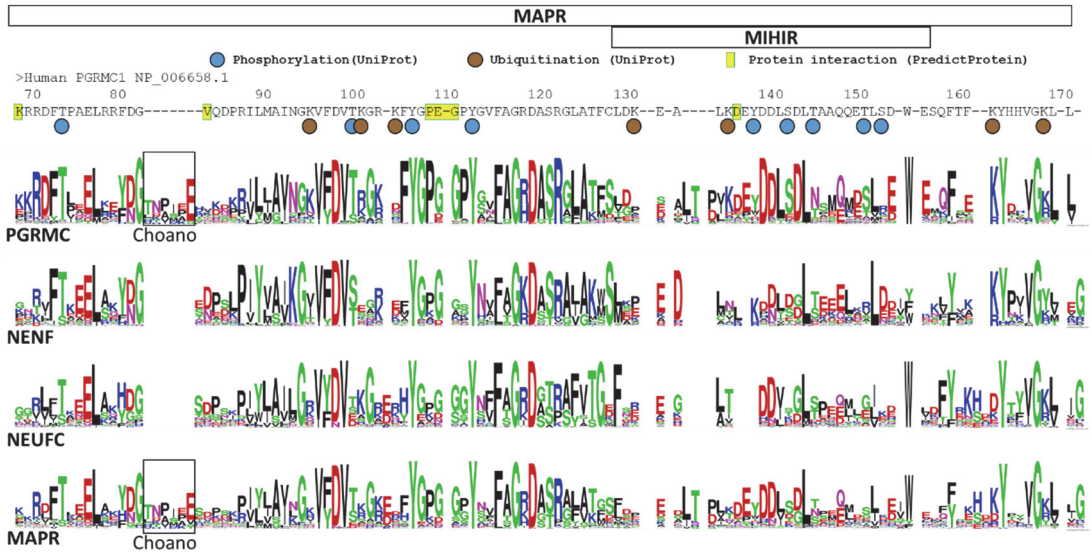


Figure 2. Consensus MAPR and subfamily Logo plots.

Logo plots are presented for all members of the PGRMC, NENF, and NEUFC families from Figure 1. The consensus plot of all MAPR sequences in the lowest row highlights the overall MAPR domain sequence identity. Apomorphic sequence insertions from some individual sequences were deleted to facilitate presentation. The box represents an apomorphic insertion in choanoflagellates that is absent from PGRMC proteins of other species. The MAPR domain of human PGRMC1 is presented above the Logo plots for reference. Documented sites of phosphorylation, ubiquitination (UniProt) and predicted sites of interaction (ProteinPredict) are indicated for PGRMC1.

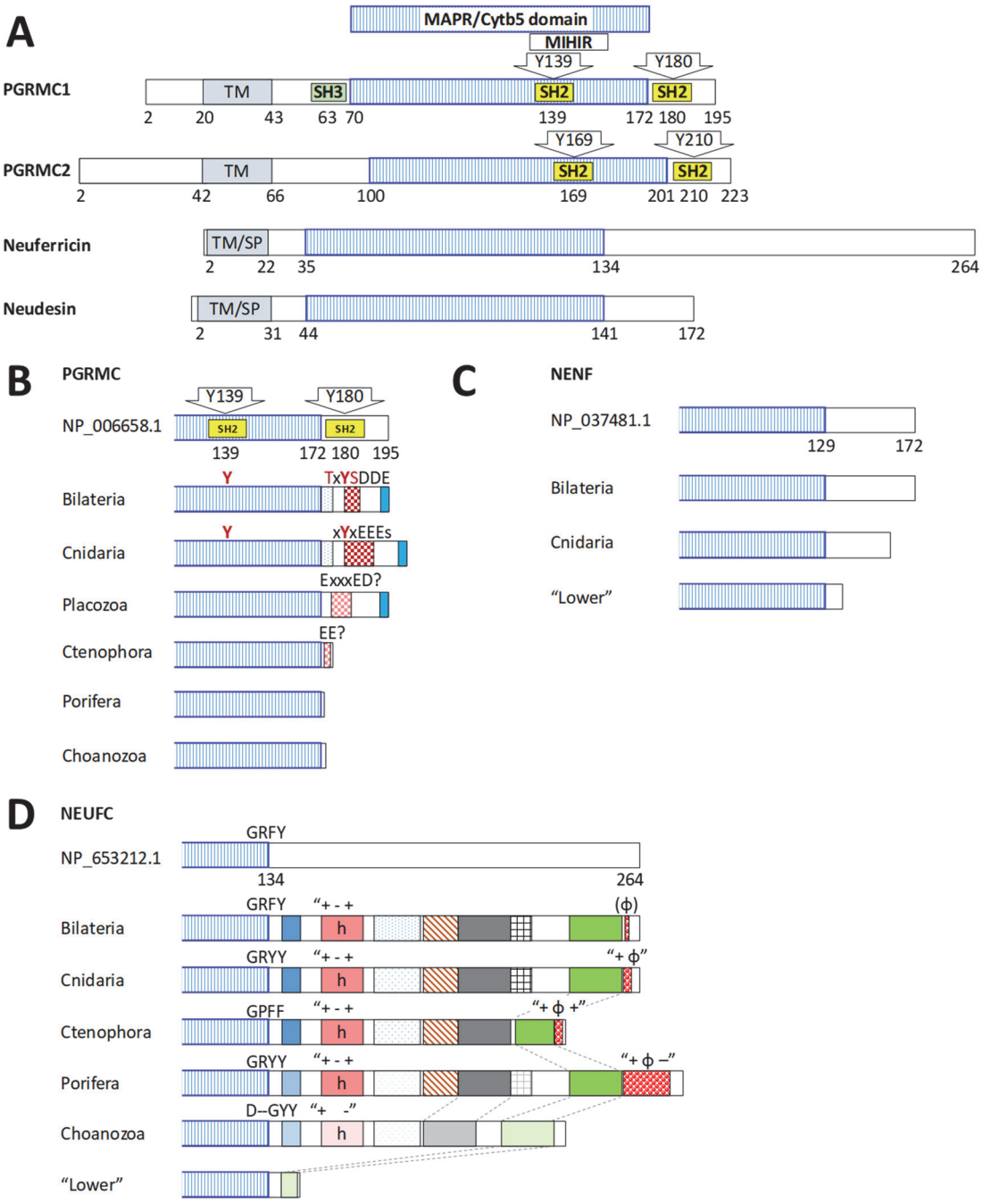


Figure 3. Evolution of MAPR C-termini in the evolution of animals.

(A) Schematic depiction of the four human MAPR genes. Numbering refers to human proteins with accession numbers provided in subsequent panels. TM: transmebrane peptide of PGRMC1/2; TM/SP: Transmembrane/signal peptide of Neuferricin and Neudesin [3].

(B) Schematic depiction of the evolution of the PGRMC C-terminus in the evolution from Choanozoa to Bilateria. Human PGRMC1 from A is at the top for orientation. The

cognate positions of Y139 and Y180 are shown to have appeared in Cnidaria but are absent from earlier-diverging animals. Boxed regions show regions of amino acid similarity without identifiable known domains. These are not implied to possess specific functions. The PGRMC1 Y180 motif consisting of T178, Y180, S181 and adjacent negative D/E region appears to have evolved in a stepwise pattern. Logo plots and further details of the schematic diagrams can be found in Figure S2.

(C) The C-terminus of NENF proteins expanded during the evolution of Cnidaria and Bilateria. Logo plots and further details of the schematic diagrams can be found in Figure S4. “Lower” refers to earlier-branching groups.

(D) The C-terminus of NEUFC was expanded in the evolution from earler-diverging opisthokonts to Holozoa (Choanozoa and Metazoa). Variously shaded boxes represent regions of presumed amino acid similarity by descent. No function is ascribed to any particular region. Consensus changes to the GFRY motif at the C-terminus of the human Neuferricin (NP_653212.1) are shown for each group. The boxes labelled h with “+ -“ above represents regions of positive and negative charge that is predicted to be surface-exposed. Another region contains positively (+) or negatively (-) charged and/or aliphatic (ϕ) residues (see Figure S7B-E). Logo plots and further details of the schematic diagrams can be found in Figure S8. “Lower” refers to earlier-branching groups.

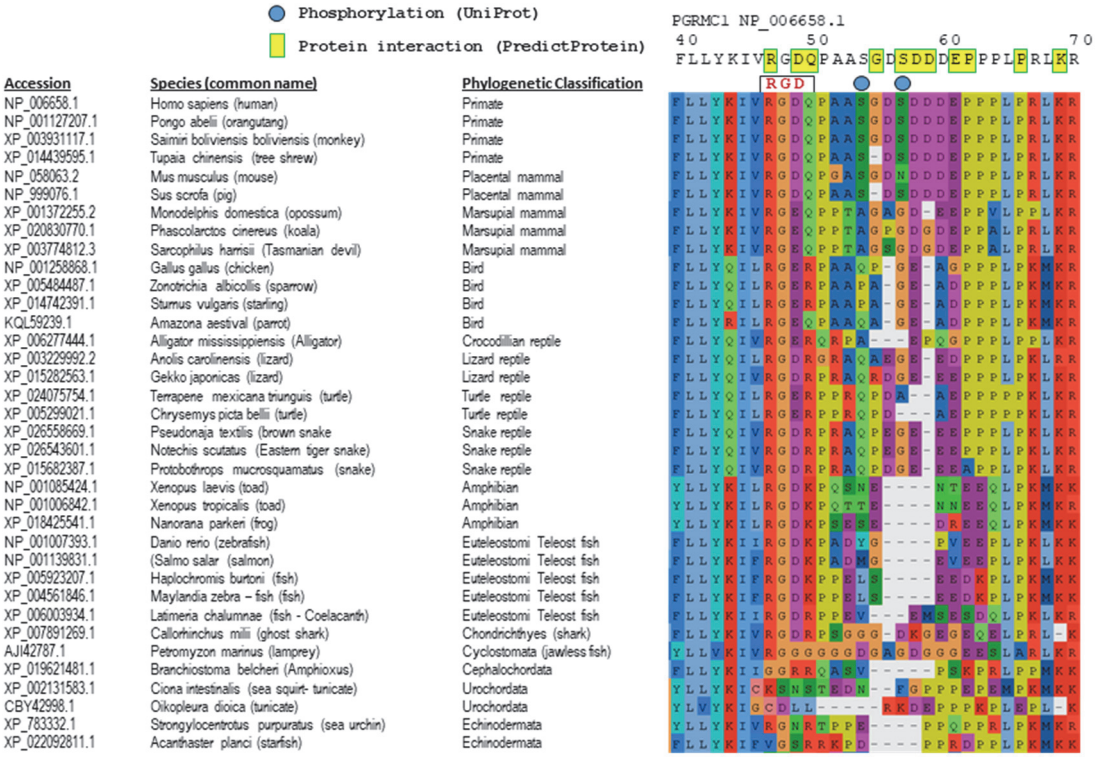


Figure 4. Alignment of PGRMC1 39-74 region of selected chordates.

This region spans part of the transmembrane helix (left) to the start of the MAPR domain (right). The indicated metazoan PGRMC sequences were aligned using MAFFT L-INS-i. The graphical presentation of the alignment was made using AliView [108].

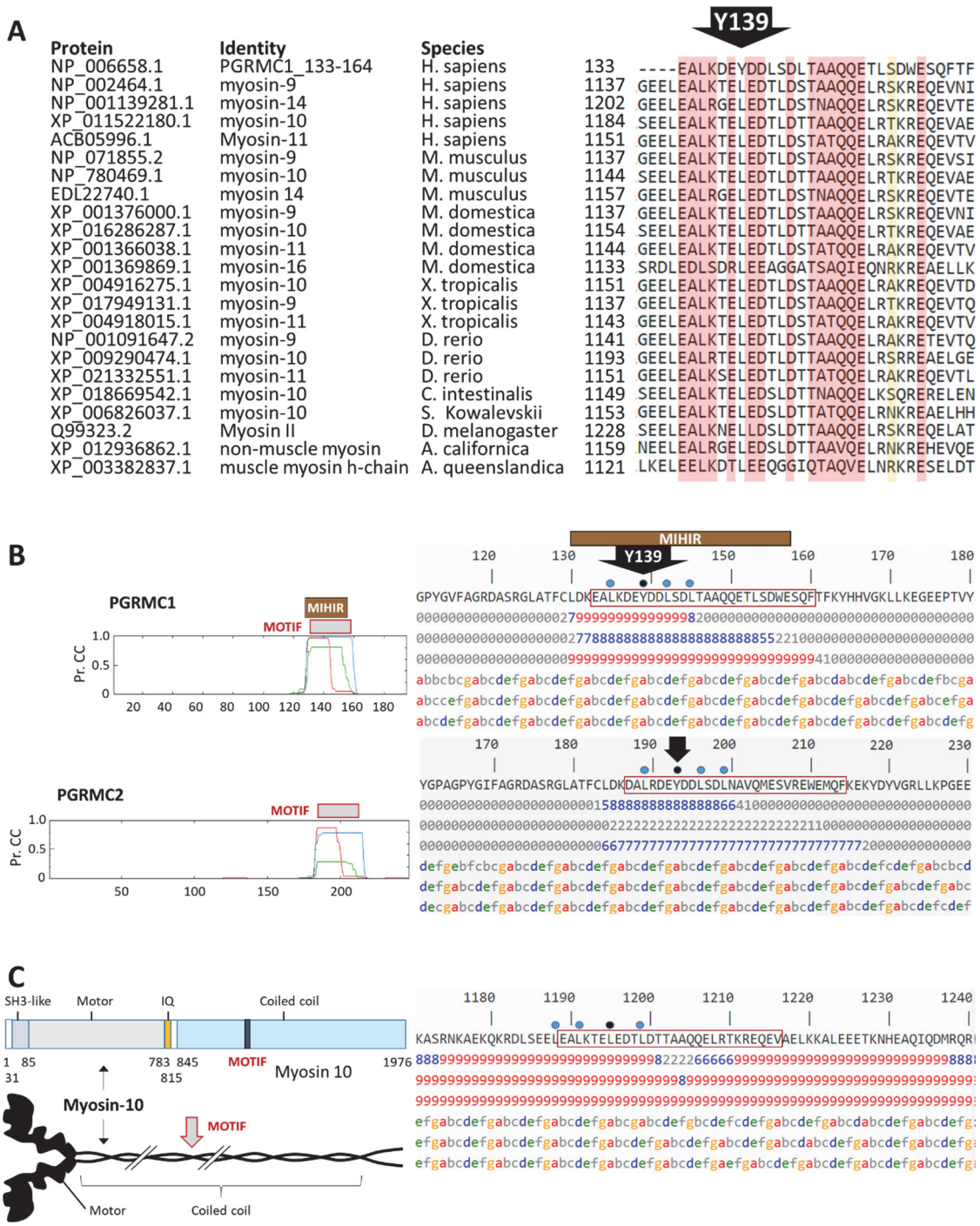


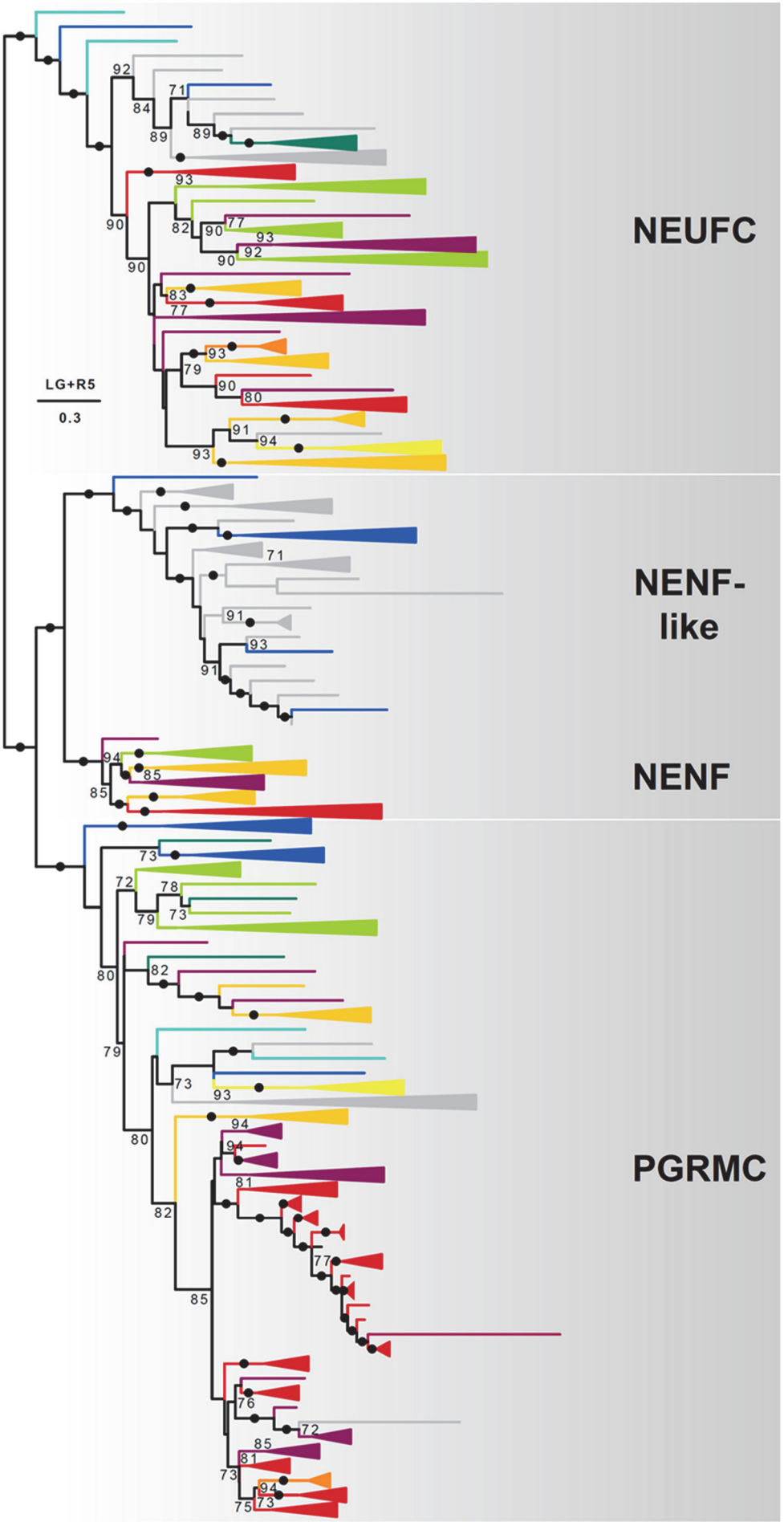
Figure 5. The PGRMC1 MIHIR has predicted coiled-coil character shared with some myosins.

(A) Alignment of PGRMC1 MIHIR residues 133-164 with selected myosin proteins detected by low stringency BLAST.

(B) The PGRMC1 and PGRMC2 MIHIR regions contain predicted high propensity to form coiled coil. The images to the left depict the probability for a particular residue to form coiled-coil based upon calculation for surrounding windows of 14 (red), 21 (blue) and 28 (green) residues, generated by the PRABI server. Panels to the right present the numerical depiction of the same result. Numbers under the sequence are the probabilities for coiled-coil formation abbreviated to first digit for windows of 14, 21 and 28 residues (i.e. 9 represents $p \geq 0.9$). Letters a-g represent the corresponding coiled coil heptad register. The positons of predicted heptad hydrophobic coiled-coil core residues including PGRMC1 Y139 are indicated.

(C) The motif from A is present in the coiled-coil region of human Myosin 10. The left side shows the position of the motif in the primary and tertiary structure of the protein. The right side format follows the conventions of B.

Supplemental Figures



1

Figure S1. Phylogenetic reconstruction of MAPR proteins in opisthokonts.

Phylogeny of 3 types of MAPR proteins in opisthokonts. Opisthokont lineages are indicated by colored branches/collapsed clades and nomenclature following Figure 1B. The scale bar and the number beneath it indicate the estimated number of substitutions per site, above the scale bar the model for tree reconstruction is indicated. Node support was calculated using 1000 ultrafast bootstrap (UFBoot) replicates. Only support values >70% are shown, black dots indicated support values of ≥ 95%.

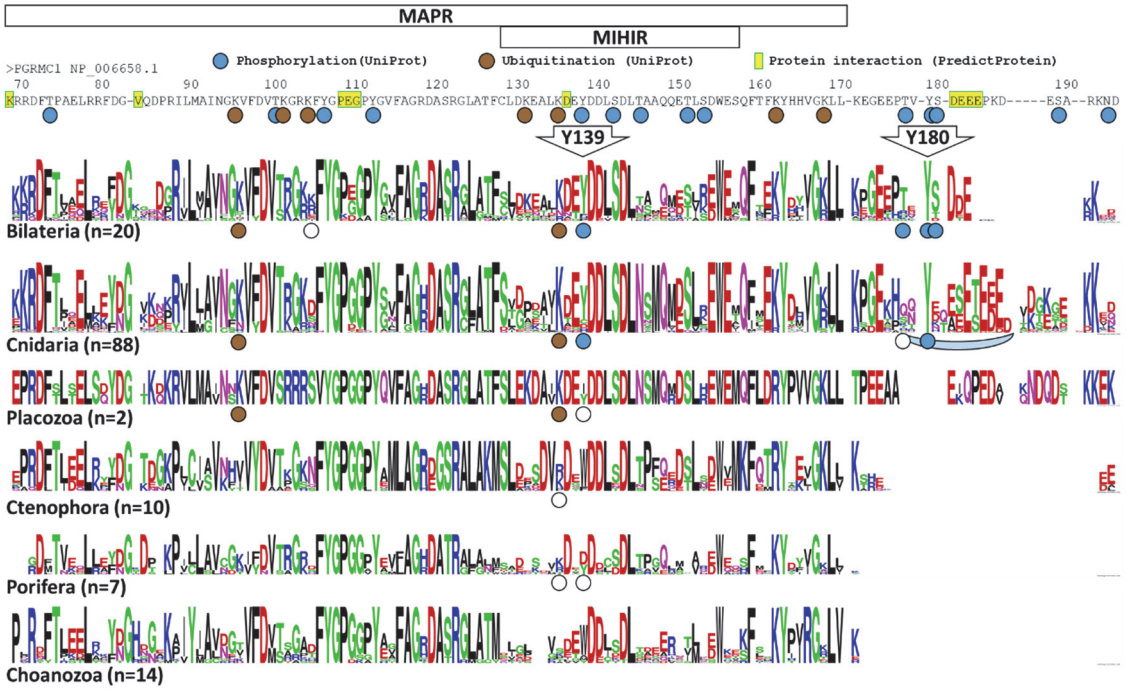


Figure S2. PGRMC MAPR and C-terminus Logo plots.

Conventions follow Figure 2. The human PGRMC1 sequence is provided at the top for reference. Logo plots were constructed from the alignment of Supplemental Information File 1.

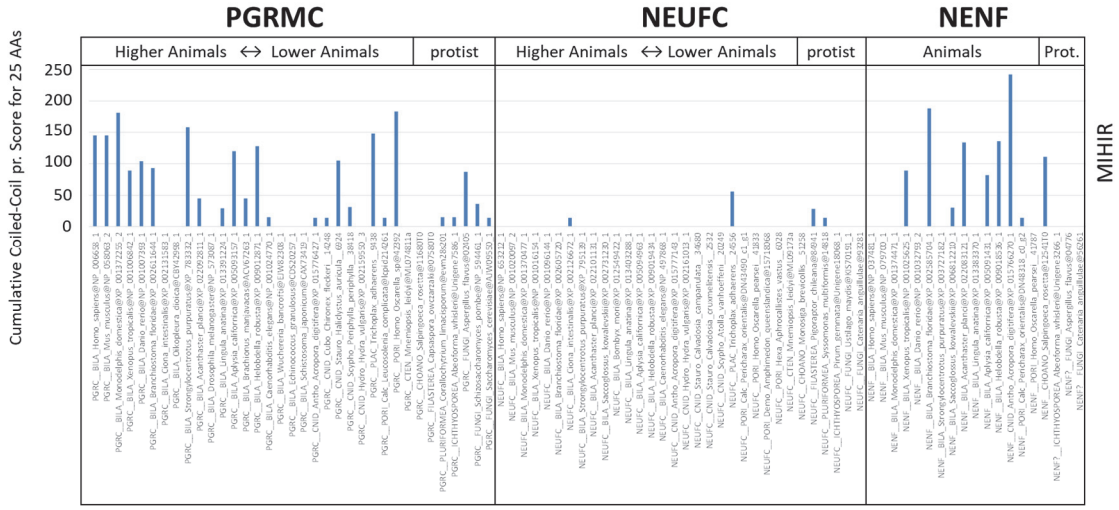


Figure S3. Coiled-coil probability of the MIHIR motif from selected MAPR sequences.

The cumulative predicted probability for coiled-coil (Cumulative coiled-coil Pr.) of 20 residues of each sequence aligned with PGRMC1 MIHIR residues TFCLDKEALKDEYDDLSDLT. By way of example, for human PGRMC1 the score was generated by 0+0+2+7+9+9+9+9+9+9+9+9+9+9+9+9+9+8+2=145 as calculated by the PRABI server.

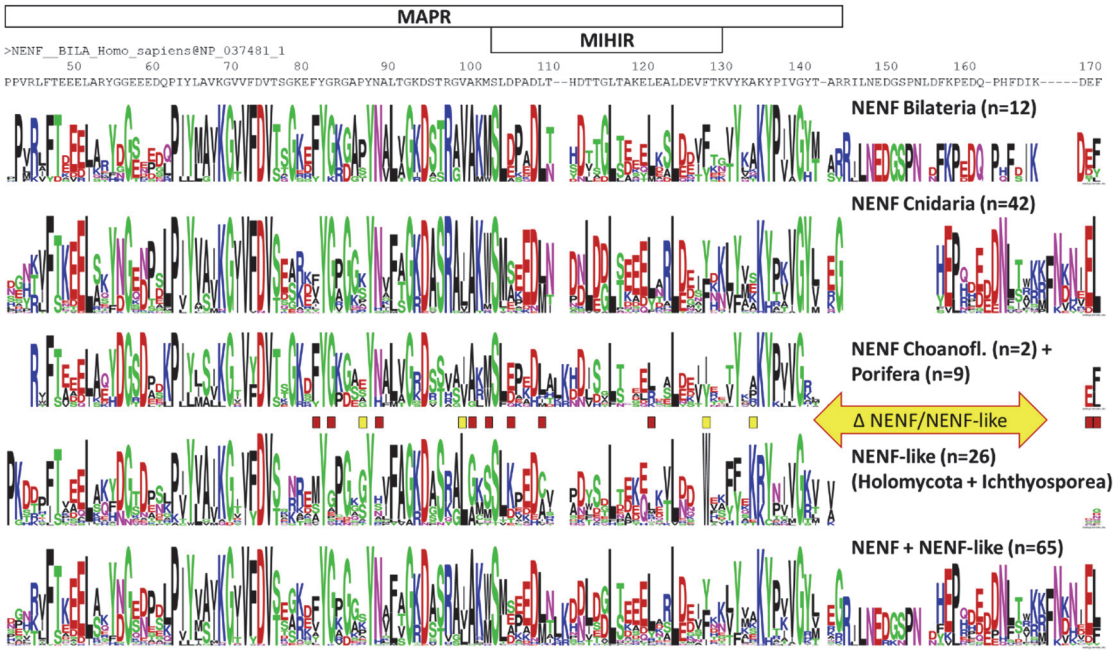


Figure S4. NENF MAPR and C-terminus Logo plot.

Conventions follow previous figures. The human NENF sequence is provided at the top for reference. Logo plots were constructed from the alignment of Supplemental Information File 3. The positions of major differences between NENF and NENF-like consenus sequences are indicated (Δ NENF/NENF-like). Red (dark) boxes represent residues where differences are more conserved in NENF, whereas yellow (light) boxes represent residues where differences are more conserved in NENF-like.

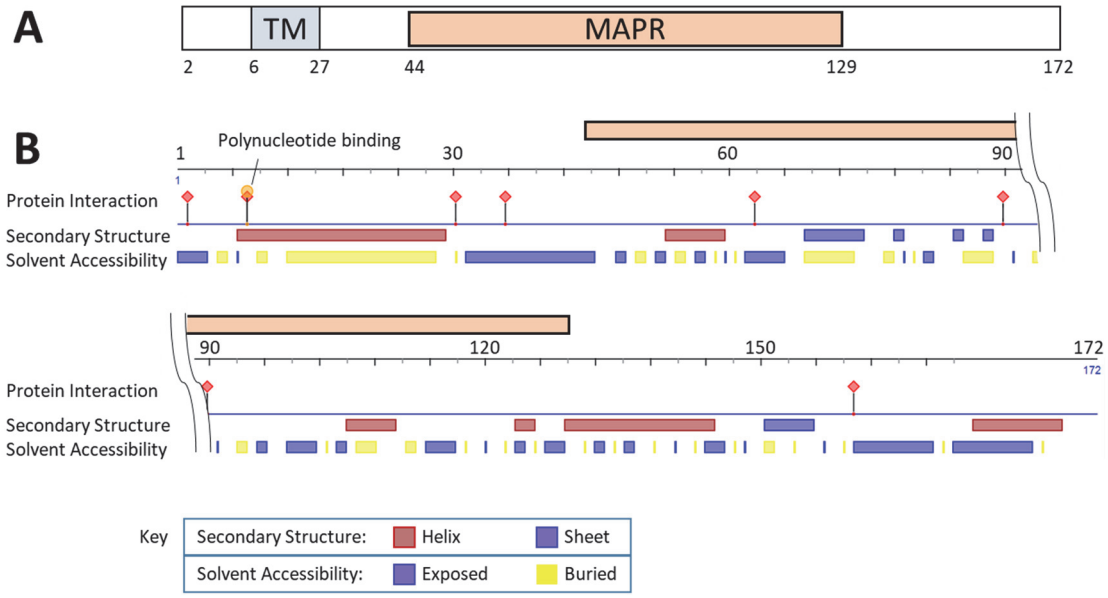


Figure S5. NENF predicted features.

A) Schematic representation of Human NENF, showing the position of the transmembrane helix (TM) and MAPR domain

B) Graphical depiction produced by the ProteinPredict server, showing sites of predicted protein interactions, secondary structure, and solvent accessibility. A single site of predicted polynucleotide binding is indicated at residue 7.

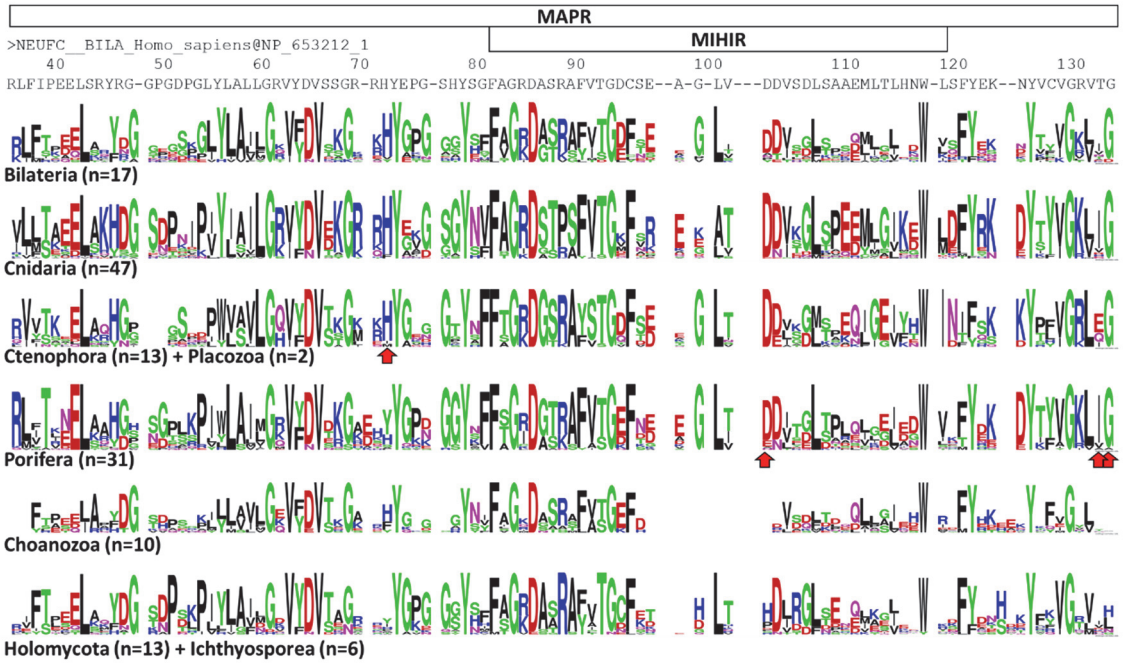
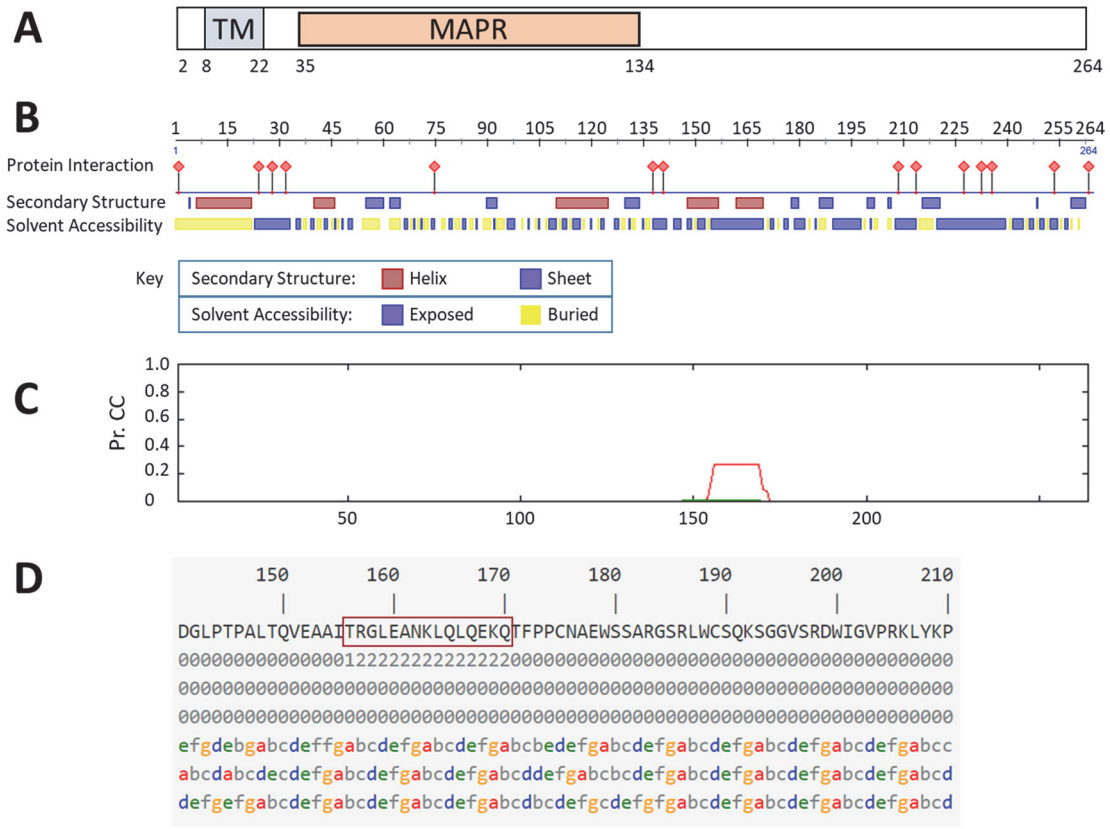


Figure S6. NEUFC MAPR domain Logo plot.

Conventions follow previous figures. The human NEUFC sequence is provided at the top for reference. Logo plots were constructed from the alignment of Supplemental Information File 4.



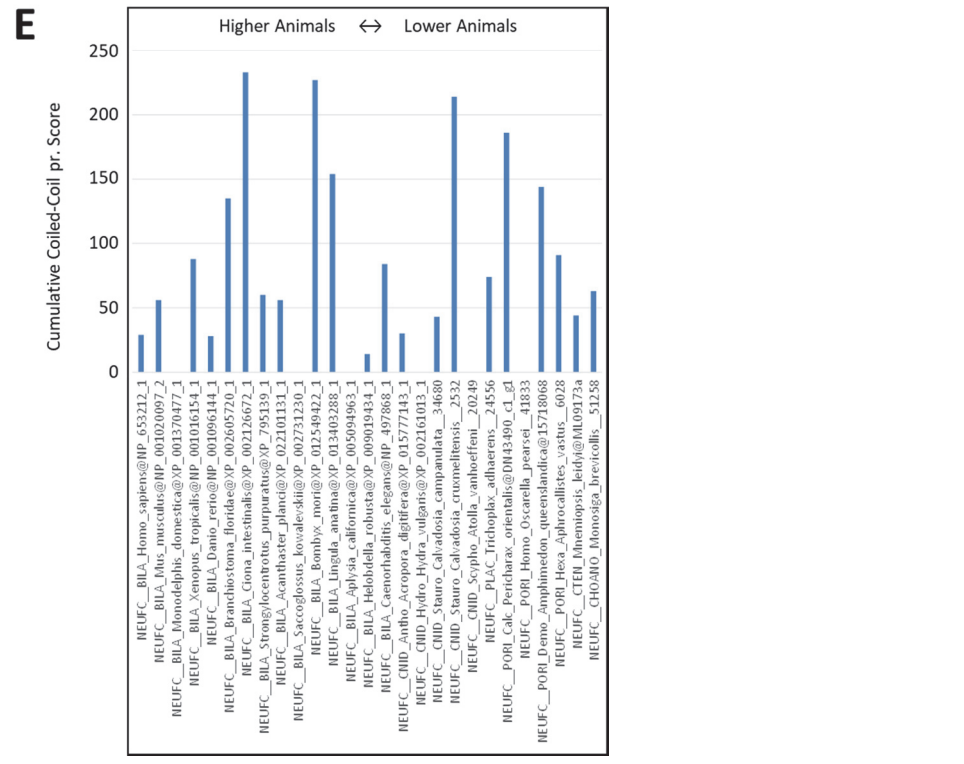


Figure S7. NEUFC predicted features.

A) Schematic representation of human NEUFC. Conventions follow Figure S5A.

B) Graphical depiction produced by the ProteinPredict server, showing sites of predicted protein interactions, secondary structure, and solvent accessibility.

C) Graphical depiction of predicted probability of forming coiled-coil generated by the PRABI server, following Figure 5B.

D) Numerical output of the results from C), following Figure 5B, with amino acid sequence boxed.

E) Cumulative predicted coiled-coil scores for the residues aligned with human NEUFC TRGLEANKLQLQEKQ from D), following the convention of Figure S3.

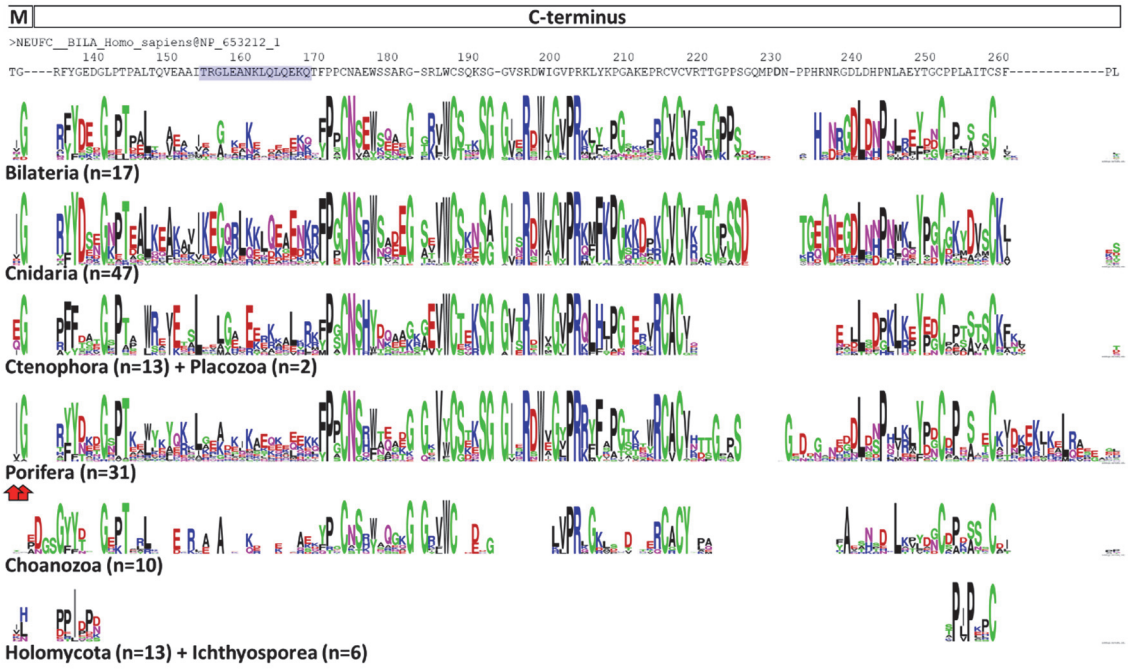


Figure S8. NEUFC C-terminus Logo plots.

Conventions follow previous figures. The human NEUFC sequence is provided at the top for reference. Shaded residues (purple) are those referred to in Figure S7C-E. Logo plots were constructed from the alignment of Supplemental Information File 5. Boxes at the top denote two residues fdorm the MAPR domain (M) and the C-terminus of the NEUFC consensus Logo plot (C-terminus).

Supporting Information files

Supplemental Information File 1. PGRMC MAPR domain and C-terminus alignment used for Logo plots of Figure S2. (PGRMC-LOGO_20181026.fasta)

Supplemental Information File 2. NENF N-terminal alignment. No Logo plot produced. (NENF-BIG_20181026.fasta)

Supplemental Information File 3. NENF MAPR domain and C-terminal alignment used for for Logo plots of Figure S4. (NENF-small_20181026.fasta)

Supplemental Information File 4. NEUFC MAPR domain alignment used for for Logo plots of Figure S6. (NEUFC-MAPR_20181026.fasta)

Supplemental Information File 5. NEUFC C-terminal alignment used for for Logo plots of Figure S8. (NEUFC-C-term_20181026.fasta)