1 Title

2 MAPR origins reveal a new class of prokaryotic cytochrome b₅ proteins and possible
3 role in eukaryogenesis

4

5 Authors

6 Daniel Tamarit^{1,2}, Sarah Teakel³, Michealla Marama⁴, David Aragão^{5,6}, Svetlana Y.
7 Gerdes⁷, Jade K. Forwood³, Thijs J. G. Ettema¹, and Michael A. Cahill^{3,8*}

8

- 9 1. Laboratory of Microbiology, Wageningen University, Wageningen 6708 WE, The
10 Netherlands.
- 11 2. Department of Aquatic Sciences and Assessment, Swedish University of
12 Agricultural Sciences, SE-75007 Uppsala, Sweden
- 13 3. School of Dentistry and Medical Sciences, Charles Sturt University, WaggaWagga,
14 NSW 2678, Australia.
- 15 4. School of Animal and Veterinary Sciences, Charles Sturt University, WaggaWagga,
16 NSW 2678, Australia.
- 17 5. Australian Synchrotron, Australian Nuclear Science and Technology Organisation,
18 800 Blackburn Road, Clayton, VIC, 3168, Australia.
- 19 6. Current Address: Diamond Light Source, Harwell Science and Innovation Campus,
20 Didcot, OX11 0DE, UK
- 21 7. Fellowship for Interpretation of Genomes, Burr Ridge, Illinois, USA.
- 22 8. The John Curtin School of Medical Research, The Australian National University,
23 ACT 2601, Australia.

24

25

26

27 *Address Correspondence to:

28 Dr. Michael Cahill, School of Dentistry and Medical Sciences, Charles Sturt University,
29 Wagga Wagga, NSW 2678, Australia.

30 e-mail: mcahill@csu.edu.au

31

1 **Abstract**

2 The multiple functions of PGRMC1, the archetypal heme-binding eukaryotic MAPR
3 family member, include steroidogenic regulation, membrane trafficking, and steroid
4 responsiveness. The interrelationships between these functions are currently poorly
5 understood. Previous work has shown that different MAPR subclasses were present
6 early in eukaryotic evolution, and that tyrosine phosphorylated residues appeared in the
7 eumetazoan ancestor, coincident with a gastrulation organizer. Here we show that
8 MAPR proteins are related to a newly recognized class of prokaryotic cytochrome-b₅
9 domain proteins. Our first solved structure of this new class exhibits shared MAPR-like
10 folded architecture and heme-binding orientation. We also report that a protein
11 subgroup from Candidate Phyla Radiation bacteria shares MAPR-like heme-interacting
12 tyrosines. Our results support bacterial origins for both PGRMC1 and CYP51A, that
13 catalyze the meiosis-associated 14-demethylation of the first sterol lanosterol from yeast
14 to humans. We propose that eukaryotic acquisition of a membrane-trafficking function
15 related to sterol metabolism was associated with the appearance of MAPR genes early
16 in eukaryotic evolution.

17

18 **Keywords**

19 Membrane-associated progesterone receptor, eukaryotic evolution, cytochrome b₅,
20 steroidogenesis, candidate phyla radiation

21

22 **1. Introduction**

23 Progesterone Receptor Membrane Component 1 (PGRMC1) is the archetypal member of
24 the eukaryotic membrane-associated progesterone receptor (MAPR) family of
25 cytochrome b₅ (cytb₅)-related proteins (Cahill, 2007). MAPR proteins are defined within
26 the cytb₅-superfamily by tyrosinate heme chelation with unique orientation of the heme

1 (Kabe et al., 2016a; Kaluka et al., 2015; Thompson et al., 2007), and by the presence of
2 a small insertion, the MAPR-specific inter-helical insertion region (MIHIR), between
3 helices 3 and 4 of the canonical cytb₅-domain (Cahill, 2007, 2017; Mifsud and Bateman,
4 2002).

5

6 Mammalian MAPR members include Neuferricin (NEUFC), Neudesin (NENF),
7 PGRMC1 and PGRMC2. The latter diverged prior to the evolution of cartilaginous fish,
8 and share many PGRMC properties. It remains unclear how they differ in function. All
9 three MAPR families (PGRMC, NEUFC, NENF) were present in the Opisthikont
10 ancestor of fungi and animals. All three families are also expressed in the nervous system
11 (Hehenberger et al., 2020), and interact with cytochrome P450 (cyP450) enzymes (Ryu
12 et al., 2017). Otherwise, NEUFC and NENF function is poorly understood: for discussion
13 see Hehenberger et al. (2020).

14

15 PGRMC1 is by far the best characterized MAPR protein, and has diverse functions
16 (Cahill et al., 2016). These include some that predate or are potentially ancient in
17 eukaryotes, such as regulation of heme synthesis (Piel et al., 2016), cyP450 interactions
18 (Ryu et al., 2017), and sterol metabolism (Cahill and Medlock, 2017). Some functions
19 arose in eukaryotes, such as membrane trafficking (Cahill et al., 2016; Hampton et al.,
20 2018), cell cycle regulation at the G₀/G₁ checkpoint (Cahill et al., 2016; Griffin et al.,
21 2014; Peluso et al., 2014), and mitotic/meiotic spindle association (Juhlen et al., 2018;
22 Luciano and Peluso, 2016; Terzaghi et al., 2016). The eukaryotic phylogenetic
23 distribution of these properties remains unexplored. Other functions are specialized
24 metazoan developments, including roles in e.g. fertility, embryogenic axon guidance, and
25 membrane trafficking associated with synaptic plasticity (Cahill et al., 2016; Izzo et al.,
26 2014).

1

2 MAPR cyP450-interactions (Ryu et al., 2017) conspicuously feature PGRMC1 regulation

3 of lanosterol-14-demethylase, the most conserved eukaryotic cyP450 (CYP51A), to

4 modify the first sterol (lanosterol) in yeast and mammals in an oxygen requiring reaction

5 (Fig. 1) (Cahill, 2007; Hughes et al., 2007b). McGuire et al. recently reported that

6 PGRMC1 binding to this and multiple cyP450s leads to their stabilization, elevating the

7 protein levels of cyP450s which occurs even for heme-binding deficient PGRMC1

8 mutants. They argue, but do not demonstrate, that the pentacoordinate and tyrosinate

9 chelated PGRMC heme iron ion is unlikely to act as an electron carrier in enzyme

10 catalyzed reactions (McGuire et al., 2021).

11

12 The isoprenoid precursors for lanosterol synthesis are produced by the mevalonate

13 pathway, whose activity in eukaryotes is regulated by the SREBP/Insig/SCAP complex,

14 which interacts with PGRMC1 (Cai et al., 2015; Suchanek et al., 2005). Thus, PGRMC1

15 is involved in the regulation of the mevalonate pathway, the modification of the first

16 sterol, and it elicits responses to steroid/progesterone levels (Cahill, 2007; Cahill and

17 Medlock, 2017; Neubauer et al., 2008).

18

19 PGRMC1 is also involved in endocytosis of low density lipoproteins (LDL) by the LDL

20 receptor (Riad et al., 2020; Riad et al., 2018). Membrane trafficking and sterol biology

21 seem to have been ancient eukaryotic traits, whereas cholesterol transport via LDLs

22 would be an evolutionary adaptation developed by multicellular animals. This both

23 suggested that sterol biology is a central feature of PGRMC1 function, and reveals a gap

24 in knowledge about how membrane trafficking and LDL receptor function arose.

25

1 We have recently shown that the evolutionary acquisition of phosphorylated PGRMC1
2 tyrosines 139 and 180 was coincident with the appearance of the gastrulation organizer
3 and synapsed neurons in the common ancestor of eumetazoans (Hehenberger et al., 2020).
4 Y139 is situated in the MIHIR, and represents one of the heptad repeat residues of a
5 predicted coiled-coil protein interaction domain that shares similarity with coiled-coil
6 motifs from various myosin proteins. Thereby, phosphorylation of Y139 would be
7 predicted to disrupt coiled-coil interactions, and permit new interactions with phospho-
8 tyrosine-binding SH2-domain proteins (Cahill, 2020; Hehenberger et al., 2020).
9
10 PGRMC1 is immunoprecipitated in protein complexes along with with components of
11 the actin cytoskeleton (Salsano et al., 2020; Teakel et al., 2020), which is in turn involved
12 in the membrane trafficking of exo- and endocytosis (Meunier and Gutierrez, 2016). It
13 follows that evolutionarily ancient actin-cytoskeletal interactions of the PGRMC1 MIHIR
14 sequence could be modulated by tyrosine phosphorylation that was acquired by
15 eumetazoans at the same time as the gastrulation organizer developed (Cahill, 2020).
16 Gastrulation involves induction by transcription factor Brachyury of a suite of actin-
17 cytoskeletal genes that have been conserved since our common ancestor with the
18 opisthokont *Capsaspora owczarzaki*, a single celled eukaryote that can switch between
19 motile and sessile life cycle stages. A similar phenotypic switch is observed during early
20 gastrulation when previously sessile epithelial cells undergo epithelial mesenchymal
21 transition, and migrate within the embryo (Sebe-Pedros et al., 2016). Interestingly,
22 PGRMC1 phosphorylation site mutants in cancer cells both altered the abundance of
23 components of the actin cytoskeleton, and affected cell motility (Thejer et al., 2020a) by
24 epigenetically altering gene expression (Thejer et al., 2020b).
25

1 PGRMC1 is induced in hypoxic human breast cancer cells at a time and place where cells
2 switch to anaerobic metabolism (Neubauer et al., 2008), which was suggestive of
3 PGRMC1 modulation of mitochondrial function. PGRMC1 localizes to mitochondria
4 (Xu et al., 2011) where it associates with mitochondrial proteins (Salsano et al., 2020),
5 modulates mitochondrial ferrochelatase (Piel et al., 2016), and is suggested to have co-
6 evolved with a number of genes encoding mitochondrial proteins (Cahill and Medlock,
7 2017).

8

9 Experimental alteration of PGRMC1 phosphorylation status indeed dramatically affects
10 mitochondrial form and function (Thejer et al., 2020a), simultaneously with altered
11 genomic mutation rates and the status of genomic CpG epigenetic methylation (Thejer et
12 al., 2020b). Since a MIHIR-containing MAPR gene appeared early in eukaryotes
13 (Hehenberger et al., 2020), and PGRMC1 affects mitochondria, we set out to investigate
14 a possible role of MAPR proteins in the origins of mitochondria and the eukaryotic cell.

15

16 Eukaryotic sterol biosynthesis arose primarily from a non-mitochondrially-inherited
17 bacterial mevalonate pathway (MVP) (Castelle and Banfield, 2018; Hoshino and
18 Gaucher, 2018). The isoprenoid squalene products of bacterial MVPs are cyclized by a
19 squalene cyclase homolog into steroid-like six ringed structures called hopanoids
20 (Barrantes and Fantini, 2016; Belin et al., 2018). Like mitochondrial cholesterol,
21 hopanoids modulate bacterial membrane order and permeability to protons (Saenz et al.,
22 2015). It has long been recognized that cholesterol decreases the permeability of the inner
23 mitochondrial membrane to protons (Baggetto et al., 1992), thereby increasing the
24 efficiency of electron transport chain activity, which is critical to mitochondrial function.

25

1 Eukaryotic lanosterol production involves the action of the novel enzyme squalene
2 epoxidase, which produces 2,3-oxidosqualene that is then the substrate of squalene
3 cyclase (Fig. 1). This eukaryotic-specific oxygen-consuming epoxidase reaction dictates
4 that steroidogenesis is oxygen dependent (and hence may be regulated by oxygen
5 tension), and is the reason that eukaryotic squalene cyclases produce the characteristic 3-
6 hydroxysterols which bacterial hopanoids do not possess (Belin et al., 2018; Kirschvink
7 and Kopp, 2008). As discussed above, the subsequent CYP51A/PGRMC1 catalyzed 14-
8 demethylation of lanosterol also requires oxygen (Fig. 1). Hence, eukaryotic
9 steroidogenesis can arguably be viewed as an adaptation of an ancestral hopanoid
10 biosynthesis pathway which can modulate membrane lipid composition in the presence
11 of atmospheric oxygen. This could potentially regulate a switch between aerobic and
12 anaerobic metabolism.

13

14 The supply of cytoplasmically-synthesized sterols to the mitochondrion was probably
15 important for eukaryogenesis before presumed proto-mitochondrial hopanoid pathway
16 genes could be lost from the bacterial genome. With its mevalonate pathway and
17 steroidogenic regulatory functions, sensitivity to sterol levels, and membrane trafficking
18 functionality (Cahill and Medlock, 2017), a PGRMC1-like MAPR protein could have
19 been ideally situated to enable this process. Here we investigated the origins of eukaryotic
20 MAPR proteins, showing that the ancestral MAPR gene originated from a newly
21 identified class of prokaryotic (cytb_{5M}) cytb₅-domain proteins. We were unable to
22 conclusively identify the phylogenetic origins of the ancestral eukaryotic proto-MAPR
23 gene, or whether a MAPR gene was present in the LECA.

24

25 **2. Materials and methods**

26 **2.1. Selection of sequences**

1 To search for MAPR-related proteins in bacteria for Fig. A1A, initial NCBI BLASTp
2 analyses (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Johnson et al., 2008) were performed
3 using the following MAPR proteins (Accession and FASTA files obtained from
4 UNIPROT): Q9UMX5 (human NENF), Q8WUJ1 (human NEUF), O00264 (human
5 PGRMC1), O15173 (human PGRMC2), Q9XFM6 (*Arabidopsis thaliana* plant MAPR
6 protein), and Q12091 (*Saccharomyces cerevisiae* yeast MAPR protein), or the single
7 human cytochrome b5 sequence (P00167). The NCBI non-redundant protein sequence
8 data base was searched, restricting phylogenetic taxa using the “Organism” input to the
9 eukaryotes (taxid:2759), Bacteria (taxid:2) or Archaea (taxid:2157). Hits were selected as
10 proteins with schematic graphical alignment across the cyt_{b5} domain as provided by the
11 BLASTp “Color key for alignment scores” output. Typical values were >40% identity,
12 E<10. For Fig. A1A new eukaryotic sets of MAPR and cyt_{b5} sequences were generated
13 with BLASTp query sequences of NP_006658_1 (PGRMC1) and P04166 (rat cyt_{b5}).
14 Three queries were made for each sequence, with “Organism” set to 1) Eukaryotes
15 excluding Opisthokonta (taxid:33154) (the clade containing yeasts and animals), 2)
16 Opisthokonta excluding metazoans (taxid:33208), and 3) Metazoans excluding Chordata
17 (taxid:7711). For each query the top ten BLASTp hits were used to generate sets of 90
18 MAPR and cyt_{b5} proteins. Default BLASTp parameters were used unless otherwise
19 specified: Expect threshold “10”, Word size “6”, Max matches in a query range “0”,
20 Matrix “BLOSUM62”; Gap Costs: “Existence: 11, Extension: 1”, and Conditional
21 adjustments “compositional score matrix adjustment”.

22

23 Another round of sequence similarity searches was performed with the hope to capture a
24 more diverse sequence dataset and ensure that all representatives of cyt_{b5MY} were found.
25 Separate BLASTp searches were performed against NCBI’s NR database using reference
26 sequences for Clade 1 (KUO41884, NP_037481_1 and NP_006658_1), Clade 2 (P04166

1 and pdb_1CXY_A) and cytb_{5MY} (EKD84736, KKQ43725, KKR05718, KKR31495,
2 KKS94036, KKU89055, OGE21801, OHA80798, OHA83705, OIP98122, and
3 PJA14771). Each one of these BLASTp searches was performed using a maximum
4 number of target sequences set to 10,000, and a maximum e-value of 1e-5. These
5 sequences were grouped into a non-redundant fasta file of 2470 sequences. CD-Hit (Fu
6 et al., 2012) was used to reduce redundancy of these sequences using 5 different similarity
7 thresholds, namely 50%, 65%, 70%, 80% and 90%.

8

9 Reduced stringency BLASTp was performed with standard NCBI BLASTp settings
10 except: 10,000 max sequences, expect threshold 50, Word size 3 (Organism: bacteria,
11 taxid:2) or Word size 2 (Organism: archaea database, taxid:2157) (Fig. A3). Word size 3
12 and 2 are the available alternative options on the NCBI BLASTp server. Data sets were
13 iteratively refined to yield multiple hundreds of aligned sequences with sufficient gap-
14 free residues for NJ tree building as described case by case in Fig. A3.

15

16 For BLASTp analysis of cytb_{5MY}-resembling proteins in eukaryotic sequences
17 (Supplementary Appendix B) the accession numbers from Fig. A1C were individually
18 entered as successive query sequences to the NCBI BLASTp server, with Organism
19 restricted to Eukaryota (taxid:2759), otherwise employing default BLASTp settings. The
20 top 100 best BLASTp hits per analysis were downloaded as FASTA files and subjected
21 to MAFFT analysis as described below including control reference sequences
22 NP_006658.1 (PGRMC1, clade-1), NP_037481.1 (NENF, clade-1), KUO41884.1
23 (Hadesarchaea archaeon YNP_N21 cytb_{5M}, clade-1), P04166.2 (rat cytochrome b₅, clade-
24 2) and 1CXY_A (*E. Vacuolata* Cytochrome B558, clade-2). The 116 resulting sequences
25 were subjected to MAFFT and refining of the data set by deleting gapped sequences to

1 obtain sufficient gap-free alignments for tree building (between 112 and 116 sequences
2 per analysis: see Supplementary Appendix B).

3

4 BLASTp of non-cytb_{5M} candidate phylum radiation (CPR) bacterial cytb₅-like proteins
5 against eukaryotes were performed using query sequences (IMG identifiers)
6 Ga0075854_11113, Ga0075854_11513, Ga0301032_10388 and Ga0301032_10385,
7 with organism restricted to Eukaryota (taxid:2759) and all other parameter default values.
8 BLASTp searches for cytb_{5M} proteins in CPR were performed separately using the query
9 sequences KUO41884.1 and KKR30394.1 with Organism limited to bacteria candidate
10 phyla (taxid:1783234). Resulting hits with homology over the cytb₅-domain and E value
11 <4 were downloaded. All hits from KKR30394 were also detected by KUO41884.

12

13 Taxon-restricted CYP51A sequences were obtained through BLASTp searches using
14 UniProt Q16850 (human CYP51A) as search string with organism restricted in separate
15 searches to 1) Opisthokonta (taxid:33154) exclude: Bilateria (taxid:33213)
16 (“Eukaryotic”), 2) Archaeobacteria (taxid:2157), 3) Alphaproteobacteria (taxid:28211)
17 (“alphaproteobacteria”), and 4) Bacteria (taxid:2), exclude Alphaproteobacteria
18 (taxid:28211) (“other bacteria”). Additionally, the eukaryotic sequences obtained by this
19 search were aligned using MAFFT L-INS-I and the obtained alignment was used as query
20 for a taxon-unrestricted Psiblast sequence similarity search (Altschul et al., 1997) against
21 NR, with 100,000 maximum target sequences and an e-value of 1e-5. Out of these
22 sequences, 4859 with e-values lower than 1e-50 were selected, and reduced with CD-Hit
23 using a 70% similarity threshold, resulting in 533 sequences.

24

25 2.2. Phylogenetic analyses

26 2.2.1. Phylogenetic reconstruction of cytb₅ sequences

1 All NCBI BLASTp hit CSV files were downloaded and hits combined in Microsoft Excel.
2 Non-redundant FASTA sequences were retrieved using Batch Entrez, and subjected to
3 MAFFT alignment by the L-INS-i method (Kato et al., 2019; Kuraku et al., 2013;
4 Yamada et al., 2016) using the Computational Biology Research Consortium (CBRC)
5 server at <https://mafft.cbrc.jp/alignment/server/index.html>. For reduced stringency
6 BLASTp, sequences which branched outside of MAPR (clade-1) and rat cytb₅ (clade-2)
7 reference sequences on the preliminary guide tree were discarded prior to iterative
8 elimination of aligned sequences and NJ tree-building. Sequences with gaps were deleted
9 manually by inspecting the MAFFT FASTA alignment in AliView (Larsson, 2014) and
10 removing gapped sequences via the “Refine dataset” function of the CBRC server to
11 iteratively generate sequence data sets with sufficient gap free sites for tree-building. The
12 sequence selection was inverted, followed by sequence realignment with MAFFT L-INS-
13 i until sufficient gap-free sites for tree building were iteratively obtained. Inferred
14 phylogenetic trees were constructed using the NJ method (all gap-free sites) and each of
15 JTT, WAG and Poisson substitution models with bootstrap selected and 1,000
16 resamplings, and were processed using Archaeopteryx software through the forester.jar
17 program (Han and Zmasek, 2009), as described on the CBRC server.
18
19 Another round of phylogenetic analyses was performed on the extended set of homologs
20 for cytb₅ sequences, and the 5 datasets generated through sequence redundancy reduction
21 with CD-Hit (see above). The resulting datasets were each aligned using MAFFT L-INS-
22 I and T-Coffee’s regressive algorithm (Garriga et al., 2019), using the heads-and-tails
23 strategy whereby each alignment is performed in the forward and reverse sense of the
24 sequences, and the four resulting alignments were then merged using MergeAlign
25 (Collingridge and Kelly, 2012). Finally, the merged alignments were trimmed using the -
26 gappyout option in trimAl (Capella-Gutierrez et al., 2009). Phylogenetic reconstruction

1 was then performed with IQ-Tree 2.0 (Minh et al., 2020) using ModelFinder
2 (Kalyaanamoorthy et al., 2017) with all variations of the models JTT, WAG, and LG,
3 plus the mixture models generated by LG+C10.C60, with multiple rate heterogeneity
4 (none, +G4, +R2, +R4, +R8) and frequency (none, +F) parameters, and adding the models
5 LG4M and LG4X. Finally, PMSF (Wang et al., 2018) approximations of the chosen
6 models (LG+C60+F+R8 for the 50% and 65% datasets, LG+C50+F+R8 for the 70% and
7 90% datasets, and LG+C40+F+R8 for the 80% dataset) were performed for all alignments
8 to obtain trees with 100 non-parametric bootstrap pseudoreplicates, which were then
9 analysed both using the classical Felsenstein criterion, and the Transfer Bootstrap
10 Expectation (TBE) (Lemoine et al., 2018) variant. One more round was performed on
11 sequence belonging only to Clade 1. The same procedure as above was followed, but
12 more parameter variations were explored. Alignments were obtained with MAFFT L-
13 INS-I. Trimming was performed to remove alignment positions with either over 75% or
14 50% gaps. Phylogenetic reconstructions were performed using LG+C60+R4+F+PMSF,
15 WAG+C60+R4+F+PMSF or the model chosen by ModelFinder among JTT, WAG and
16 LG with C10, C20, C40 and C60 mixtures, plus LG4M, LG4X, UL2 and UL3, in all
17 combinations with extended rate heterogeneity (none, +G4, +G8, +R4, +R8) and
18 frequency (none, +F) parameters), resulting in the WAG+C60+R6 and WAG+C60+R6+F
19 mixture models being chosen for the PMSF approximation of the 50% and 65% reduced
20 datasets, respectively). The 65%, 70% and 90% reduced datasets were also aligned
21 following the heads-and-tails strategy with MAFFT L-INS-I, trimmed at a 75% gap
22 threshold, and used for phylogenetic reconstruction under the LG+C60+R4+F+PMSF
23 model. Critical TBE thresholds were calculated for specific branches following Lemoine
24 et al. (2019), as $(1 - ((p-x)/(p-1)))$, where t is the number of sequences at the lighter side
25 of the bipartition, and x is the number of sequences of the clade of interest (13).

26

2.2.2. Phylogenetic analyses of CYP51A

The 533 CYP51A sequences obtained as above were aligned using MAFFT L-INS-I and trimmed with trimAl removing all columns containing at least 50% gaps. IQ-Tree 2.0 was used for phylogenetic reconstruction, under the model LG+G4+F, with 1000 replicates for SH-like approximate likelihood ratio tests and Ultrafast bootstraps.

2.3. Crystal structure

To achieve protein expression, the codon optimised ORF was subcloned into pGEX-4T-1-H expression vector to create pGEX4T1_KUO41884 (Fig. A2), transformed into *BL21(DE3) pLysS* cells (Novagen) (9F-, *ompT*, *hsdSB*(rB-mB-), *gal*, *dcm* (DE3) pLysS (CamR)), and cultured in an expression base media (1 % Tryptone, 0.5 % yeast extract) containing 30 mL of 2.852 M NaCl and ampicillin (100 µg/ml) until the OD₆₀₀ reached 0.6. Protein expression was induced using 1 mM isopropyl 1-thio-D galactopyranoside (IPTG) and cells were incubated overnight at 37°C at 80 rpm. Cells were harvested by centrifugation at 6,000 rpm at 18°C for 30 minutes. Cell pellets were resuspended in GST (glutathione-S-transferase) cell lysis buffer (50 mM Tris(hydroxymethyl)aminomethane, 125 mM NaCl, pH 7.4). All protein purification steps were performed at room temperature. The soluble whole cell lysate was filtered through a 0.45 µm syringe filter. The soluble cell extracts were injected using a superloop at 2 mL/minute into a GST column equilibrated with GST cell lysis buffer. Proteins were eluted using 10 mM Glutathione. The N-terminal GST tag was cleaved with 100 µL of Tobacco etch virus (TEV) protease. Size exclusion chromatography was performed to further purify proteins using AKTA FPLC using an S200 26/600 filtration column (GE Healthcare), following Khandokar et al. (2017). The protein eluted as a single homogeneous peak, and at a volume consistent with the expected molecular weight of a monomer. The protein was stored at -20°C prior to use for crystallisation.

1

2 Crystals were obtained through sparse matrix screening and the hanging drop vapor
3 diffusion method following Khandokar et al. (2017). Crystals in the space group $P6_3$
4 diffracted to 1.9Å at the Australian Synchrotron microcrystallography beamlines (Aragao
5 et al., 2018; Cowieson et al., 2015). The diffraction data was integrated in Mosflm (Battye
6 et al., 2011), scaled, and reduced in AIMLESS (Evans, 2011; Evans and Murshudov,
7 2013), and the structure determined by molecular replacement using PDB ID 1J03 in
8 Phaser (McCoy et al., 2007), REFMAC (Vagin et al., 2004), PHENIX (Adams et al.,
9 2010), and COOT (Emsley et al., 2010). The final structural model has been refined to an
10 Rwork and Rfree of 0.22 and 0.27 respectively, no Ramachandran outliers, and good
11 stereochemistry (see Table A2). LIGPLOTs (Laskowski and Swindells, 2011), structural
12 superposition (Krissinel and Henrick, 2004), and DALI (heuristic PDB search) protein
13 structure comparison by alignment of distance matrices (Holm and Laakso, 2016) were
14 performed as described.

15

16 2.4. Other protein analysis

17 The degree of conservation of cytb₅-domain residues in the structure of PGRMC1 was
18 visualized using the Consurf server (<http://consurf.tau.ac.il/>) (Ashkenazy et al., 2016).
19 Sequence Logo Plots of amino acids in respective protein clades were generated using
20 WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) (Crooks et al., 2004). Non-PGRMC1
21 residues from the 4X8Y structure (Kabe et al., 2016a) were not included in the PGRMC1
22 structural data for Fig. 3.

23

24 2.5. Gene cluster analyses

25 Conserved gene cluster analysis is based on the observation that functionally related
26 genes are often collocated on the chromosomes in prokaryotes, preserving similar gene

1 context across phylogenetically diverse organisms (Overbeek et al., 1999; Pellegrini et
2 al., 1999). During evolution proteins that function together in a pathway or structural
3 complex evolve in a correlated fashion, and tend to be either preserved collectively, thus
4 ensuring that the pathway or complex remains fully functional, or be eliminated all
5 together. Methods of chromosomal gene context analysis have proved to be valuable for
6 delineation of evolutionary patterns between organisms, as well as for protein function
7 prediction (Mavromatis et al., 2009; Overbeek et al., 1999; Pellegrini et al., 1999).

8

9 To find all genomes containing *cytb_{5M}* genes, we first downloaded all 7645 CPR
10 bacterial genomes from NCBI. Using the alignment employed to reconstruct the
11 phylogeny in Fig A3B, we performed an HMM search against all CPR genomes using
12 HMMER 3.1b2 (hmmer.org). All candidate sequences were joined with those from the
13 original alignment, aligned with MAFFT-auto, trimmed with trimAl to remove all
14 columns with over 90% gaps and used for fast phylogenetic reconstruction using
15 Fasttree2 (Price et al., 2010). Genomes containing genes clustering with the *cytb_{5M}*
16 sequences were selected for further analysis, and annotated using InterProScan v5.48-
17 83.0 (Jones et al., 2014). Taxonomic identification of these genomes was performed
18 against the dereplicated genomes provided by Jaffe et al. (2020) using fastANI (Jain et
19 al., 2018) with default parameters or, if no results were found, by using a fragment length
20 threshold of 1000 bp. Gene synteny analyses were performed using the genoPlotR
21 package (Guy et al., 2010) and the Gene Neighborhood Viewer and Chromosomal
22 Cassette Viewer tools (Mavromatis et al., 2009) of the Integrated Microbial Genomes
23 (IMG) database (Chen et al., 2019).

24

25 **3. Results**

26 **3.1. MAPR related to new prokaryotic *cytb_{5M}***

1 Preliminary BLASTp searches for the presence of MAPR-related proteins in prokaryotes
2 were conducted using eukaryotic MAPR proteins and human cytb₅ as queries. We
3 combined the significant hits from separate BLAST searches into a non-redundant list of
4 176 proteins, and generated an alignment including the reference sequences for human
5 PGRMC1, human NENF, and rat cytb₅. We inferred a preliminary distance-based
6 phylogenetic tree, which produced two distinct eukaryotic clades (Fig. A1A), both of
7 which were subtended by prokaryotic sequences. Notably, clade-1 contained MAPR
8 proteins and the BLASTp hits obtained with those queries, and clade-2 contained
9 eukaryotic cytb₅, and its corresponding BLASTp hits. These results suggest that the
10 eukaryotic cytb₅ protein families are polyphyletic and originated from at least two
11 separate prokaryotic lineages. An alternative explanation could be that we have
12 theoretically sampled MAPR-like and cytb₅-like proteins from a single bacterial
13 homolog, with apparent distinctiveness being an artefact of our BLASTp search strategy
14 using MAPR and eukaryotic cytb₅ sequence search strings.

15

16 To address this possibility we first separately interrogated the archaeal and bacterial
17 sequence data bases by BLASTp with selected prokaryotic sequences from each of clades
18 1 and 2 from the prior analysis to generate independent sets of cytb₅ sequences. MAFFT
19 alignment again produced inferred trees with clade-1 and -2 tree topology, with bootstrap
20 confidence values (BCV) >94% for both archaeal and bacterial trees (see Supplementary
21 Appendix A). BLASTp with string sequences from one clade again did not detect
22 members of the other. Combining all of the sequence from above with additional MAPR
23 sequences yielded similar results when aligning 814 sequences (Fig. A1B), consistent
24 with the cytb₅-domain proteins in clades-1 representing a new type of distinct prokaryotic
25 cytb₅ proteins. MAPR and some bacterial proteins formed a sub-cluster distinct from 348
26 other prokaryotic clade-1 sequences with BCV 87% (Fig. A1B). PDF tree images,

1 FASTA sequence alignments, and xml tree files for preliminary analyses and the panel
2 of Fig. A1A are available in Supplementary Appendix A.

3

4

5 We had still not eliminated the possibility that separate clades were due to BLASTp
6 search string bias. To further investigate this issue, we employed reduced-stringency
7 BLASTp analysis by using ‘word size’ (the number of adjacent residue identity required
8 for alignment) parameters of 3 and 2, so that queries from one clade detected sequences
9 from the other clade (Fig. A2). The results continued to generate two discrete clades
10 connected by a long branch (Fig. A3). We therefore detected no evidence that the two
11 eukaryotic clades detected represented a sampling artefact selected by a biased choice of
12 BLAST query sequences from a broad continuum of prokaryotic cytb₅ proteins. . Instead,
13 these results support the existence of two separate types of prokaryotic cytb₅ domain,
14 which gave rise to eukaryotic MAPR proteins (clade-1) and cytb₅ proteins (clade-2),
15 respectively. We propose the name cytochrome b_{5M} (cytb_{5M}) for the newly recognized
16 “MAPR-like” prokaryotic cytb₅-domain clade-1 proteins.

17

18 3.2. Bacterial cytb_{5M} cluster with MAPR

19 Next, we considered the clade-1 bacterial sequences that clustered closer to MAPR than
20 to other prokaryotic cytb_{5M} proteins (Fig. A1B,C). Three contained MIHIR residues,
21 while a clade of 11 sequences did not. The first three belonged to two 65 and 14 kb-long
22 delta-proteobacterial contigs (KPK16282 and KPK52515) and one acidobacterium
23 chromosome (ANM31058). They were nested within eukaryotic sequences in the tree,
24 thus potentially representing misclassified sequences or horizontally transferred genes.
25 The top eukaryotic BLASTp hits for two MIHIR-containing sequences were plants
26 (KPK16282.1, KPK52515.1; not shown), and for the other (ANM31058.1) were all either

1 choanoflagellates or animals (not shown). We investigated the top BLASTp similarity of
2 the proteins encoded by the two contiguous genes in their respective contigs, and
3 observed that multiple top hits belonged to the same broad taxa, consistent with their
4 classification being correct. Hence, these MIHIR-containing bacterial sequences likely
5 represent HGT originating from eukaryotic plants and holozoans.

6
7 The other 11 MAPR-clustering bacterial sequences lacked MIHIR sequences, but
8 contained the cognate equivalents of PGRMC1 heme-chelating Y113, and heme
9 hydrogen bond donors Y107, K163 and K164 (Kabe et al., 2016a) (Fig. A1C), as opposed
10 to the cytb₅-domain bis-his axial heme ligation of hitherto described bacterial and
11 eukaryotic cytb₅ proteins (Kabe et al., 2016a; Liu et al., 2014). We designate this
12 subgroup of cytb_{5M} proteins with “MAPR-like tyrosines (Y)” as cytb_{5MY}. All cytb_{5MY}
13 sequences were found to originate from CPR bacteria (Table A1). The top 100 BLASTp
14 hits of all cytb_{5MY} sequences from Fig. A1B against eukaryotic sequences returned only
15 proteins with MIHIR sequences (Supplementary Appendix B), indicating that eukaryotic
16 genomes do not encode cytb_{5MY} proteins.

17
18 To further investigate cytb_{5MY} phylogenetic relationships, we performed additional
19 sequence similarity searches using the 11 cytb_{5MY} and additional reference sequences as
20 queries (see Methods). We then reduced the resulting 2486 sequences under 5 different
21 levels of redundancy, and subjected each one of these into more rigorous phylogenetic
22 analyses. The obtained maximum likelihood phylogenies corroborated the previous
23 topology, revealing two major separate eukaryotic clades (containing MAPR and cytb₅
24 sequences, respectively), and supporting the affiliation of CPR-bacterial cytb_{5MY} with the
25 eukaryotic MAPR sequences, albeit under different topologies and without satisfying
26 branch support (Fig. A4). This is important because if cytb_{5MY} formed a sister group to

1 eukaryotic MAPR proteins it could imply that MAPR proteins originated from a cytb_{5MY}
2 protein. However, if cytb_{5MY} proteins are topologically nested within the MAPR family,
3 it would imply that CPR bacteria obtained a MAPR gene from eukaryotes by HGT. We
4 were unable to discriminate between these models.

5

6 In four of the obtained phylogenies, cytb_{5MY} branched within eukaryotes, while the fifth
7 displayed cytb_{5MY} as sister to eukaryotes. For each phylogenetic reconstruction, 100 non-
8 parametric bootstrap pseudoreplicates were obtained, for which both Felsenstein
9 Bootstrap Proportion (FBP) and Transfer Bootstrap Expectation (TBE) (Lemoine et al.,
10 2018) values were calculated to assess the robustness of the obtained trees.

11

12 We performed additional phylogenetic reconstructions using the same 5 reduced datasets,
13 but restricting the analysis to clade-1 sequences only (Fig. 1, Supplementary Appendix
14 D). In these trees, we observed two major topologies with respect to the positioning of
15 the cytb_{5MY} sequences: as sister to MAPR group (Fig. 1A), or nested within it, possibly
16 as sister to the NEUFC clade (Fig. 1B) (see also Supplementary Appendix D). In general,
17 these trees included high TBE support for many relevant branches, and none of the trees
18 can be considered to accurately reflect cytb_{5MY} relationships. TBE values reflect the
19 average number of transfers at either side of the bipartition, thus indicating the
20 consistency of a clade but not necessarily of all its members (Lemoine et al. 2018). For
21 that reason, we calculated for each tree the threshold TBE value that would support a
22 given bipartition without allowing the transfer of the 13 cytb_{5MY} sequences.

23

24 With this consideration, several trees on the sequence datasets reduced at 50% and 90%
25 redundancy levels included a TBE-supported topology with cytb_{5MY} as sister to all
26 eukaryotic MAPR sequences. However, trees reconstructed from the sequence dataset

1 reduced at 65% redundancy level supported the alternative, derived position as sister to
2 the NEUFC eukaryotic sequences. From these results we conclude that cytb_{5MY} may
3 derive from a HGT event from a eukaryote into a CPR bacterium, followed by loss of the
4 MIHIR sequences and further propagation to other CPR species through HGT. However,
5 an alternative scenario that a CPR cytb_{5MY} protein gave rise to eukaryotic MAPR proteins
6 cannot be excluded by our tree-building exercise.

7

8 Furthermore, the relationship between the eukaryotic MAPR protein families indicates
9 that multiple events of duplication occurred early in the evolution of eukaryotes and
10 diversified in various eukaryotic groups (Hehenberger et al., 2020). However, it is not
11 possible to conclude from this analysis whether a MAPR ancestral protein was present in
12 the LECA, or whether it was acquired later during eukaryotic evolution. The analysis is
13 limited by the relatively short homologous region of the cytb₅ domain, the substantial
14 evolutionary distances separating the sequences, and the skewed taxonomic distribution
15 of the domain.

16

17 Both MAPR and cytb_{5MY} proteins are derived cytb_{5M} proteins. However, the primitive
18 MAPR/cytb_{5MY} structural and functional characteristics remain uncertain. We suggest
19 these proteins to be sufficiently distinct to warrant reference to prokaryotic cytb_{5M} and
20 cytb_{5MY} proteins, with the continued use of the MAPR terminology for eukaryotes
21 (although formally, as clade-1 members, MAPR and cytb_{5MY} proteins cladistically belong
22 to the cytb_{5M} clade).

23

24 3.3. Sequence differences between clades

25 Next, we analyzed the nature of conserved sequence differences between cytb₅, cytb_{5M},
26 cytb_{5MY}, and MAPR proteins. Sequence logos (Fig. 3A) revealed altered frequencies of

1 amino acid usage between clades ($\Delta C1:\Delta C2$, Fig. 3A). This included a surface loop
2 between PGRMC1 G83-R88 (Kabe et al., 2016a) (loop-1 in Fig. 3), which is at least two
3 residues larger in all clade-1 than clade-2 proteins. G83, site of a pronounced change of
4 polypeptide backbone direction (Kabe et al., 2016a), is strongly conserved in clade-1 (Fig.
5 3A). A second loop involving rat-cytb₅ heme-binding (Rodriguez-Maranon et al., 1996)
6 was strongly conserved in clade-2 (loop-2 in Fig. 3). Additionally, relative to cytb_{5M},
7 MAPR and cytb_{5MY} proteins exhibit conserved cognates of PGRMC1 Y107, Y113, K163
8 and Y164 (Fig. A1C), all of which are involved in heme interaction (Kabe et al., 2016a).
9 The residues differential between clade-1 and clade-2, or cytb_{5M} and cytb_{5MY}/MAPR
10 proteins are interspersed along the primary PGRMC1 sequence (Fig. 3A, top), yet form a
11 contiguous surface extending from the heme-binding pocket to the surface loop, which
12 rests upon conserved F81 and G83 (Fig. 3B,C). Residues more similar between MAPR
13 and cytb_{5MY} are clustered around the heme-binding pocket (purple in Fig. 3A,B). In
14 contrast, the residues conserved among all proteins in the analysis (conserved between
15 both cytb₅-domain clades) predominantly occupy the protein interior (Fig. 3D).

16

17 3.4. Clade-1 protein structure

18 We next solved the first crystal structure of a prokaryotic cytb_{5M} clade-1 protein from the
19 archaeon Hadesarchaea YNP_N21 (KUO41884.1) (Fig. A1B) (attempts to crystalize
20 multiple cytb_{5MY} proteins were unsuccessful: not shown), to compare this with the
21 structures of representative proteins from each major group of clades-1 and -2, revealing
22 overall shared similarity of the clade-1 proteins (Fig. 4). The HP and HS heme iron
23 chelation sites of rat cytb₅ (Fig. 4E) are strongly conserved in clade-2, with loop-2
24 extending from the conserved HP site (Fig. 3A). Hadesarchaea-cytb_{5M} lacks the MAPR
25 heme-binding tyrosines yet binds heme in a MAPR-like orientation. It shares one heme
26 chelating histidine (H61) with conventional (clade-2) cytb₅ proteins (Fig. 4E, Fig. A5),

1 which is conserved in cytb_{5M} proteins (Fig. 3A). Heme-binding residues differ between
2 cytb_{5M}, cytb_{5MY} and MAPR proteins within clade-1 (Fig. 3A). Excluding known MAPR
3 proteins (Cahill, 2007), the published structures most closely resembling Hadesarchaea-
4 cytb_{5M} belong to clade-2 and exhibit clade-2-like heme-binding (Fig. A6, Table A3),
5 confirming the novelty of our archetypal clade-1 cytb_{5M} structure. Altogether, the
6 requirements for dissimilar heme-binding and folded architecture underlie the presence
7 of two discrete prokaryotic cytb₅ domain-related clades in Fig. 2, one of which gave rise
8 in eukaryotes to classical cytb₅ (clade-2), and the other to MAPR proteins (clade-1).

9

10 3.5. Genomic context of CPR cytb_{5MY} clade-1 proteins

11 To compare the genomic context of CPR cytb_{5M} and cytb_{5MY} genes we investigated the
12 distribution of both protein types in CPR bacterial genomes. To recover all genomes
13 containing cytb_{5MY} sequences regardless of sequence redundancy, we performed another
14 phylogenetic analysis with the sequences shown in Fig. 1AB and all hits with e-value
15 lower than 1e-5 after an HMM search against all 7645 CPR genomes found in NCBI (as
16 of July 5th 2021) (Fig A6). This tree revealed a total of 36 genomes which contained
17 cytb_{5MY} proteins. To visualise the presence of cytb_{5MY} in CPR bacteria, we mapped these
18 36 genomes onto a phylogenomic reconstruction performed by (Jaffe et al., 2020) (Fig.
19 5A). This analysis revealed that these genomes belong to multiple distinct phyla and that
20 the presence of cytb_{5MY} is highly punctuated in CPR bacteria, indicative of horizontal
21 gene transfer as their main mode of evolution.

22

23 We then investigated the gene neighbourhoods of cytb_{5MY} in these 36 genomes and found
24 that in most (29), the cytb_{5MY} genes colocalized with putative ferric reductase (pFre)
25 genes containing a ferric reductase-like transmembrane domain (Fig. 5B, Fig. A7). Other
26 genes found to colocalise with cytb_{5MY} were those containing other cytb₅-domain

1 proteins (28), and two-component inducible signal transduction systems (22) (Fig. 5B,
2 Fig. A7). The predicted phylogenetic tree topology for the *cytb5MY* genes in Fig. A7 is
3 identical to Fig. A8, which expands the *cytb5MY* section of Fig A3B. While the synteny
4 was often lost, a putative operonic structure of these genes following the same disposition
5 was found for multiple distant lineages (e.g., Fig. 5B). Alternative configurations were
6 found in multiple genomes, although these were often highly similar, such as a group of
7 8 Daviesbacterial and 3 Yonathbacterial genomes with over 99% average nucleotide
8 identity (ANI) values, respectively (Fig. 5A).

9
10 Using the Integrated Microbial Genomes and Microbiomes (IMG) platform, we observed
11 that besides *cytb5MY* and pFre, *cytb5MY* -containing operons often involved two novel
12 types of atypical *cytb5*-domain-like proteins which we denote as *cytb5*-like type A
13 (CBLA) and B (CBLB) proteins, additional to the *cytb5MY* protein and a two-component
14 signal transduction element (Fig. 5B). The CBLA and CBLB proteins were distantly
15 related to *cytb5*-domain protein clades-1 and -2, respectively, with barely detectable
16 homology (Fig. A9). Because the deepest branches of the tree of Fig. A7 contain a two
17 component element, a pFre gene, two *cytb5*-related genes as well as a *cytb5MY* gene,
18 this probably represents the ancestral CPR state. All proteins in this putative ancestral
19 operon contain one or more predicted transmembrane helices – hence, we speculate these
20 could potentially form a membrane-associated protein complex.

21
22 Pfamscan (Mistry et al., 2007; Mistry et al., 2021) identified the ferric reductase domain
23 (PF01794) in 1427 of the 7645 CPR genomes, 185 of which contained genes with *cytb5*
24 domain (PF00173) within 10 kb of the pFre genes (164 of which were distinct from the
25 39 genomes described above). Yet colocalization of pFre with *cytb5*-domain genes,
26 including *cytb5MY*, was not observed outside of CPR bacteria in the IMG database,

1 suggesting that a putative transmembrane protein complex containing pFre and cyt_{b5}-like
2 proteins performs an unknown CPR-specific function.

3 4 3.6. CYP51A is of bacterial origin

5 Having considered the prokaryotic origins of MAPR proteins, we next considered the
6 origins of the eukaryotic MAPR-containing enzyme pathway responsible for synthesizing
7 14-demethylated lanosterol. Fig. 1 shows the conversion of squalene, the product of the
8 mevalonate pathway (MVP), to the first sterol lanosterol, and subsequent decarboxylation
9 to 14-dimethyl-14-dehydrolanosterol, also known as follicular fluid meiosis-activating
10 sterol (FF-MAS) (Mitsche et al., 2015). The respective MVP (Castelle and Banfield,
11 2018; Hoshino and Gaucher, 2018) and squalene cyclase enzymes (Barrantes and Fantini,
12 2016; Rajamani and Gao, 2003) are of bacterial origin (Frickey and Kannenberg, 2009).
13 PGRMC1-like proteins bind to and regulate CYP51A to catalyze subsequent lanosterol-
14 14-demethylation in opisthokonts such as yeast and mammals (Hand et al., 2003; Hughes
15 et al., 2007b). The evolutionary origin of CYP51A, however, has thus far remained
16 unclear.

17
18 We performed taxon-restricted maximum-likelihood phylogenies of CYP51A, which
19 revealed a closer relationship between eukaryotic CYP51A and bacterial rather than
20 archaeal proteins with an ultrafast bootstrap support of 91% and an SH-like approximate
21 likelihood ratio test support value of 98.7% (Fig. 6). A larger-scale taxonomically-
22 unrestricted sequence similarity search against the NR database retrieved ca. 3500
23 eukaryotic, ca. 1000 bacterial (of which 30 from Alphaproteobacteria) and only 4 archaeal
24 sequences with an e-value lower than 1e-50. A maximum-likelihood phylogeny of a
25 reduced but phylogenetically diverse dataset obtained from these sequences resulted in a
26 monophyletic eukaryotic clade with no particular affiliation to any specific bacterial

1 group (Fig. 6). Altogether, these results indicate that CYP51A derived from a bacterial
2 gene, indicating that both the MVP (Hoshino and Gaucher, 2018) and genes for the
3 subsequent conversion of squalene to FF-MAS (Fig. 1) may have been inherited via HGT
4 from one or more bacterial donors. This would mean that steroidogenesis was imposed
5 upon the emerging eukaryotic cell by bacteria.

7 **4. Discussion**

8 Here, we provide the first description of distinct clades of MAPR-related cytb_{5M} clade-1
9 and eukaryotic cytb₅-like clade-2 prokaryotic proteins. We show that the distinctive cytb₅-
10 related MAPR proteins and conventional eukaryotic cytb₅ proteins arose from separate
11 ancestral eukaryotic cytb₅-domain genes. MAPR proteins may have arisen from a newly
12 recognized distinct class of CPR cytb_{5MY} proteins, already possessing the residues
13 necessary for the unique tyrosinate-based heme chelation of MAPR proteins. Conversely,
14 cytb_{5MY} protein may alternatively have arisen by HGT of a eukaryotic MAPR protein into
15 CPR bacteria. We cannot confidently discriminate between these opposing models. The
16 cytb₅ domain is rather small, permitting relatively few informative sites for phylogenetic
17 analyses to reconstruct their evolutionary relationships. The relationship between MAPR
18 and cytb_{5MY} proteins remains therefore uncertain, the latter perhaps being either a sister
19 group to all MAPR proteins or alternatively having arisen from the NEUFC clade within
20 MAPR. While genome annotators currently refer to e.g. “Cytochrome b₅-like
21 Heme/Steroid-binding domain-containing protein” to proteins of both clades, we suggest
22 that “Heme/Steroid-binding” should only characterize MAPR-like clade-1.

23

24 The newly identified class of MAPR-like cytb_{5MY} proteins are unique to CPR bacteria,
25 which typically have small genomes characteristic of symbionts (Brown et al., 2015;
26 Castelle and Banfield, 2018; Castelle et al., 2018). Some eukaryotic MVP enzymes

1 resemble those from CPR (Castelle and Banfield, 2018; Hoshino and Gaucher, 2018).
 2 Therefore, a CPR *cytb₅MY* gene may have contributed to a MAPR-influenced eukaryotic
 3 steroidogenic pathway. Genomic organization of CPR *cytb₅MY* genes suggests an
 4 ancestral role related to undefined yet regulated redox reactions, which have no known
 5 eukaryotic counterparts.

6
 7 The presence of a *cytb₅MY*/pFre/*cytb₅*-like operon co-locating with a two-component
 8 inducible system (Fig. 5A) in CPR but not other prokaryotes strongly suggests the
 9 existence of an inducible CPR-specific process which requires *cytb₅MY* proteins.
 10 Regulated bacterial ferric-reductase-like operons are unusual because in prokaryotes
 11 these enzymes are largely constitutively expressed, catalyzing the reduction of free
 12 flavins which in turn could transfer electrons to a variety of substrates (e.g. various ferric
 13 siderophores are known substrates for the ‘classic’ ferric reductases) (Schroder et al.,
 14 2003). While eukaryotic ferric reductases are specific for Fe^{3+} , in prokaryotes they are
 15 merely flavin reductases: reducing diverse flavins, not Fe^{3+} directly. Hence, we can
 16 predict neither flavin-specificity, nor the specificity for the terminal pFre reduction
 17 substrate: which could be Fe^{3+} , copper, other siderophores (Schroder et al., 2003), or
 18 entirely different compounds.

19
 20 The signature tyrosinate heme chelation of MAPR proteins allows them to bind
 21 ferric/ Fe^{3+} heme tightly, but ferrous/ Fe^{2+} heme weakly (Kaluka et al., 2015; Thompson et
 22 al., 2007), potentially discharging their heme upon reduction (“one-shot”, or
 23 stoichiometric reactivity), with heme-chaperone and conditional status-monitoring
 24 implications (Cahill and Medlock, 2017; Kabe et al., 2016b). This specialized
 25 functionality may be involved at CPR pFre/*cytb₅MY* loci, speculatively involving Fe^{3+}
 26 reduction via *cytb₅*-domain proteins, with *cytb₅MY*-released Fe^{2+} -heme perhaps acting as

1 siderophore (Schroder et al., 2003). Involvement of MAPR proteins in similar biology
2 remains unreported, however the heme-transport function of PGRMC2 (Galmozzi et al.,
3 2019) is conceivably related.

4

5 The ancestral MAPR protein had evolved a eukaryotic-specific MIHIR motif, and
6 together with CYP51A was catalysing the oxidative 14-demethylation of lanosterol (Fig.
7 1) early during eukaryotic evolution, at least before the diversification of fungi and
8 metazoans (Hand et al., 2003; Hughes et al., 2007b). Although cytb_{SMY} proteins may have
9 evolved through HGT from a eukaryotic NEUFC gene to CPR bacteria, a bacterially-
10 inherited MAPR-dependent steroidogenic pathway (Fig. 1) could carry profound
11 implications for early eukaryotic evolution.

12

13 Eukaryotic membranes differ from those of archaea, such that a bacterial contribution to
14 their evolution is believed to have been substantial. Cytoplasmically synthesized sterols
15 and membrane lipids required the evolution of new transport pathways to reach the
16 mitochondrion. Through regulation of SREBP1 and SREBP2 activity, PGRMC1
17 regulates the synthesis of fatty acids and sterols respectively (Cai et al., 2015; Lee et al.,
18 2018; Shimano and Sato, 2017; Suchanek et al., 2005), the core constituents of eukaryotic
19 cell membranes. It is possible that MAPR-mediated membrane trafficking (Cahill et al.,
20 2016; Riad et al., 2020) contributed to the properties of eukaryotic cell membrane whose
21 lipids resemble bacterial more than archaeal membranes, or alternatively, to the
22 trafficking of cytoplasmically synthesized novel lipids to modify mitochondrial
23 membrane properties. This line of reasoning could be highly informed if the NEUFC and
24 NENF MAPR proteins were better functionally characterized, yet that is not the case
25 (Hehenberger et al., 2020).

26

1 Four main features hitherto distinguished MAPR from other previously described cytb₅
2 proteins: tyrosinate heme-chelation, heme orientation, a MIHIR, and membrane-
3 trafficking functionality. Our analysis, the resemblance of the MIHIR to an actin-
4 cytoskeleton interaction motif (Hehenberger et al., 2020), and the association of
5 PGRMC1 with actin cytoskeletal components (Salsano et al., 2020; Teakel et al., 2020),
6 together show that heme orientation was inherited by MAPR from a protein within the
7 prokaryotic clade-1, whereas the MIHIR appears to have been a eukaryotic invention
8 which may be associated with membrane trafficking. MAPR and cytb₅MY proteins also
9 share tyrosinate heme chelation, however the primitive state of the two protein classes
10 remains unclear, as discussed above.

11
12 We hypothesise that the original MIHIR-containing MAPR protein may have initially
13 been associated with the heme-related and redox-modulated modification of sterol-
14 containing membranes and their trafficking, perhaps to modulate mitochondrial function.
15 A potential non-vesicular route for mitochondrial sterols is via endoplasmic reticulum
16 (ER)-mitochondrial contacts (EMC): communication mediators between ER and
17 mitochondrial compartments. PGRMC1 is also associated with EMC (Cho et al., 2017).
18 In a preprint, Sabbir and colleagues report that PGRMC1 ablation disrupts the ordered
19 association of mitochondria with the ER at EMC/mitochondrial-associated membranes
20 (MAMs) (Sabbir et al., 2020). Because of the deep evolutionary conservation of the
21 MAPR/CYP51A1 steroidogenic reaction, this MAPR function is possibly related to
22 mitochondrial regulation of an early eukaryote. However, this conjectural hypothesis
23 requires further study.

24
25 Sterols are involved in endocytosis at multiple levels in yeast and animals, both in
26 facilitating membrane properties that enable receptor activation, and by participation in

1 an actin-independent post-internalization process (Heese-Peck et al., 2002). Additionally
2 to catalyzing the CYP51A reaction, PGRMC1 is involved in sensing progesterone and
3 progestogen sterol levels (Cahill et al., 2016; Cahill and Medlock, 2017; Ruan et al.,
4 2012), associates with sigma-2 receptor/TMEM97 in sterol transport via endocytosis
5 (Cahill and Medlock, 2017; Riad et al., 2020; Riad et al., 2018; Xu et al., 2011), and
6 regulates sterol/lipid homeostasis via interaction with the SREBP/Insig/SCAP complex,
7 where it is also involved in transcriptional regulation of SREBP-1 (Cahill and Medlock,
8 2017; Lee et al., 2018; Suchanek et al., 2005).

9

10 An evolutionary ancient role of MAPR-regulated sterol synthesis and mitochondrial
11 oxygen consumption status seems plausible. Whereas bacterial hopanoid synthesis does
12 not require oxygen (Saenz et al., 2015), molecular O₂ is required for steroidogenesis: in
13 particular squalene oxidase epoxidation of squalene (Belin et al., 2018; Kirschvink and
14 Kopp, 2008) and the first lanosterol modification to the 4-methyl sterol FF-MAS by the
15 highly conserved PGRMC1-regulated CYP51A reaction (Stromstedt et al., 1996;
16 Takishita et al., 2017) (Fig. 1). We could therefore model early eukaryotic steroidogenesis
17 as a modulator of oxic/anoxic metabolic responses.

18

19 Accordingly, the synthesis of yeast cholesterol-like ergosterol from lanosterol requires
20 oxygen. The fungal Sre1 (the yeast ortholog of SREBP) is cleaved to activate
21 steroidogenesis in a process regulated by hypoxia and the sensing of 4-methyl sterols by
22 Scp1 (the yeast ortholog of SCAP). Changes in 4-methyl sterol levels can lead to an Scp1-
23 dependent proteolysis of Sre1, in a mechanism that has been proposed (based upon
24 presence in mammals and two fungi) to possibly be conserved among unicellular
25 eukaryotes (Hughes et al., 2007a; Hughes and Espenshade, 2008). The fungal Sre1
26 hypoxic response is accompanied by upregulated increased carbohydrate catabolism and

1 globally reduced transcription and translation in a monumental metabolic switch by the
2 coordination of multiple hypoxic pathways whose interconnectedness remains poorly
3 understood (Burr and Espenshade, 2018).
4
5 Interestingly, these dramatic metabolic changes are reminiscent of PGRMC1-dependent
6 effects on mammalian cell culture cells (Thejer et al., 2020a; Thejer et al., 2020b),
7 however, any mechanistic link remains uninvestigated. Both the *Schizosaccharomyces*
8 *pombe* Dap1 MAPR protein (a PGRMC-like protein (Hehenberger et al., 2020)) and the
9 CYP51A homolog are transcriptionally-induced by hypoxia (Hughes et al., 2007b).
10 PGRMC1 is also induced in the hypoxic zone of human tumors (Neubauer et al., 2008),
11 possibly representing conserved ancestral MAPR hypoxic functionality. That could be
12 related to a proposed role of mitochondria in O₂ detoxification during early eukaryotic
13 evolution (Imachi et al., 2020; Kurland and Andersson, 2000) (either pre- or post-LECA),
14 and to the profound effects of PGRMC1 on mitochondria (Thejer et al., 2020a), and on
15 glycolytic metabolism (Sabbir, 2019; Thejer et al., 2020b).
16
17 MAPR genes are present in at least plants, opisthokonts, *Guillardia*, *Dictyostelium* and
18 Stramenopiles (Hehenberger et al., 2020) (Fig. 2) (there has been no systematic
19 phylogenetic survey), which suggests an early MAPR origin in eukaryotes. Many aspects
20 of modern meiosis reflect functions shared by these groups (Loidl, 2016), which may
21 extend to MAPR proteins and steroids. Yeast meiotic membrane fusion requires sterols
22 (Aguilar et al., 2010). Mammalian PGRMC1 is attached to the kinetochore microtubules
23 of meiotic/mitotic spindles (Juhlen et al., 2018; Luciano and Peluso, 2016), and to
24 metaphase centromeres (Juhlen et al., 2018; Luciano et al., 2010; Terzaghi et al., 2016).
25 The products of the reaction catalyzed by PGRMC1 and CYP51A are either FF-MAS
26 (Fig. 1) or dihydro-FF-MAS, both 4-methylsterol inducers of meiosis (Fakheri and Javitt,

2011; Mitsche et al., 2015). Similar 4-methylsterols activate meiosis in plants as well as animals (Darnet and Schaller, 2019) and are induced by hypoxia in yeast (see above), suggesting the operation of an ancient and conserved mechanism. Mammalian PGRMC1 mediates a progesterone-induced block of meiotic progression (Guo et al., 2016; Luciano and Peluso, 2016). Yeast *Saccharomyces cerevisiae* Dap1 (PGRMC1 ortholog) mutation also leads to cell cycle defects (Hand et al., 2003). As tantalizingly as these PGRMC attributes align with possible ancestral eukaryotic biology, their phylogenetic distribution has not been investigated, and so functional extrapolation to an early eukaryotic ancestral MAPR gene remains speculative.

10

Whereas post-cholesterol steroid hormones evolved in vertebrates, the synthetic pathway from lanosterol to cholesterol involves production of diverse bioactive sterols (Mitsche et al., 2015) which sequentially appeared during evolution, where CYP51A/PGRMC1 catalyze the first step after lanosterol is produced (Fig. 1). The closest related bacterial proteins to eukaryotic CYP51A (Fig. 6) suggest that HGT of both the ancestral MAPR and CYP51A genes to an early eukaryote from genomes other than the proto-mitochondrial alphaproteobacterium to metabolize a lanosterol-like hopanoid and regulate its transport may have occurred early in eukaryotic evolution.

19

The MAPR-dependent first sterol modification (Fig. 1) may have originally been catalyzed by bacterial enzymes to manipulate an associated early eukaryotic cell, and drive a metabolic switch that is still operative in humans. These considerations refocus the context of highly conserved PGRMC1 biology across the spectrum of heme/lipid metabolism and transport, and affected diseases such as metabolism and glucose regulation, cancer, fertility, and neuropathologies such as Alzheimer's disease (Cahill et

1 al., 2016; Cahill and Medlock, 2017; Izzo et al., 2020; Lee et al., 2018; Peluso and Pru,
2 2014; Riad et al., 2020; Riad et al., 2018; Sabbir, 2019; Xu et al., 2011).

3

4 CRediT authorship contribution statement

5 **Daniel Tamarit:** Data curation, Formal analysis, Funding acquisition, Investigation,
6 Methodology, Writing - review & editing.

7 **Sarah Teakel:** Formal analysis, Writing - review & editing.

8 **Michealla Marama:** Formal analysis, Writing - review & editing.

9 **David Aragão:** Methodology, Resources, Writing - review & editing.

10 **Svetlana Y. Gerdes:** Data curation, Formal analysis, Investigation, Methodology,
11 Writing - original draft.

12 **Jade K. Forwood:** Funding acquisition, Methodology, Resources, Supervision, Writing
13 - review & editing.

14 **Thijs J. G. Ettema:** Funding acquisition, Methodology, Resources, Supervision, Writing
15 - review & editing.

16 **Michael A. Cahill:** Conceptualization, Formal analysis, Funding acquisition,
17 Investigation, Methodology, Project administration, Resources, Writing - original draft.

18

19 Declaration of Competing Interest

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

24

25 Acknowledgements

1 Acknowledgements. This research was supported by Charles Sturt University (CSU)
2 School of Biomedical Sciences (SBMS) internal awards to M.A.C. and PhD scholarship
3 support to S.T; by the Swedish Research Council (VR grants 2015-04959 to T.J.G.E., and
4 2018-06609 to D.T.), by the European Research Council (ERC consolidator grant 817834
5 to T.J.G.E.). Computation was performed on resources provided by the Swedish National
6 Infrastructure for Computing through the Uppsala Multidisciplinary Center for Advanced
7 Computational Science (UPPMAX) under projects SNIC-2020-5-37 (to T.J.G.E. and
8 Felix Homa) and SNIC-2020-15-113 (to D.T.). This research was partly undertaken on
9 the macromolecular crystallography beamlines at the Australian Synchrotron, part of
10 ANSTO. Elisabeth Hehenberger (Department of Botany, University of British Columbia,
11 Vancouver, Canada) provided protist MAPR sequences. Aya Narunsky (The Nir Ben-Tal
12 Group at the Department of Biochemistry and Molecular Biology of Tel-Aviv University)
13 assisted with Consurf results. Amy Medlock (University of Georgia), George John
14 (CSU), and Mike Dyll-Smith (University of Melbourne) commented upon draft
15 manuscripts.

16

17 Supplementary Information is linked to the online version of the paper.

18

19 Research data for this article

20 The structure of Hadesarchaea YNP_N21 cytb_{5M} (KUO41884.1) has been deposited in
21 the Protein Data Bank with accession number 6NZX. The PDB validation report is
22 included as Supplementary Appendix G. Other data and materials are supplied as
23 supplemental data or are available upon request.

24

25

1 References

- 2 Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd,
3 J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W.,
4 Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart,
5 P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular
6 structure solution. *Acta crystallographica. Section D, Biological crystallography* 66, 213-
7 221.
- 8 Aguilar, P.S., Heiman, M.G., Walther, T.C., Engel, A., Schwudke, D., Gushwa, N.,
9 Kurzchalia, T., Walter, P., 2010. Structure of sterol aliphatic chains affects yeast cell
10 shape and cell fusion during mating. *Proceedings of the National Academy of Sciences*
11 *of the United States of America* 107, 4170-4175.
- 12 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman,
13 D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search
14 programs. *Nucleic Acids Res* 25, 3389-3402.
- 15 Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J.,
16 Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H.,
17 Hubbard, S.S., Banfield, J.F., 2016. Thousands of microbial genomes shed light on
18 interconnected biogeochemical processes in an aquifer system. *Nature communications*
19 7, 13219.
- 20 Aragao, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N.P.,
21 Ericsson, D.J., Gee, C.L., Macedo, S., Mudie, N., Panjekar, S., Price, J.R., Riboldi-
22 Tunnicliffe, A., Rostan, R., Williamson, R., Caradoc-Davies, T.T., 2018. MX2: a high-
23 flux undulator microfocus beamline serving both the chemical and macromolecular
24 crystallography communities at the Australian Synchrotron. *Journal of synchrotron*
25 *radiation* 25, 885-891.
- 26 Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016.
27 ConSurf 2016: an improved methodology to estimate and visualize evolutionary
28 conservation in macromolecules. *Nucleic Acids Res* 44, W344-350.
- 29 Baggetto, L.G., Clottes, E., Vial, C., 1992. Low mitochondrial proton leak due to high
30 membrane cholesterol content and cytosolic creatine kinase as two features of the deviant
31 bioenergetics of Ehrlich and AS30-D tumor cells. *Cancer research* 52, 4935-4941.
- 32 Barrantes, F.J., Fantini, J., 2016. From hopanoids to cholesterol: Molecular clocks of
33 pentameric ligand-gated ion channels. *Progress in lipid research* 63, 1-13.
- 34 Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G., 2011.
35 iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.
36 *Acta crystallographica. Section D, Biological crystallography* 67, 271-281.
- 37 Belin, B.J., Busset, N., Giraud, E., Molinaro, A., Silipo, A., Newman, D.K., 2018.
38 Hopanoid lipids: from membranes to plant-bacteria interactions. *Nature reviews.*
39 *Microbiology* 16, 304-315.
- 40 Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins,
41 M.J., Wrighton, K.C., Williams, K.H., Banfield, J.F., 2015. Unusual biology across a
42 group comprising more than 15% of domain Bacteria. *Nature* 523, 208-211.
- 43 Burr, R., Espenshade, P.J., 2018. Oxygen-responsive transcriptional regulation of lipid
44 homeostasis in fungi: Implications for anti-fungal drug development. *Semin Cell Dev*
45 *Biol* 81, 110-120.
- 46 Cahill, M.A., 2007. Progesterone receptor membrane component 1: an integrative review.
47 *J Steroid Biochem Mol Biol* 105, 16-36.
- 48 Cahill, M.A., 2017. The evolutionary appearance of signaling motifs in PGRMC1.
49 *Bioscience trends* 11, 179-192.

- 1 Cahill, M.A., 2020. Choose your partners for the next dance: implied PGRMC1 roles in
2 membrane trafficking and mitochondrial modulation. *Fertility and sterility* 113, 938-941.
- 3 Cahill, M.A., Jazayeri, J.A., Catalano, S.M., Toyokuni, S., Kovacevic, Z., Richardson,
4 D.R., 2016. The emerging role of progesterone receptor membrane component 1
5 (PGRMC1) in cancer biology. *Biochimica et biophysica acta* 1866, 339-349.
- 6 Cahill, M.A., Medlock, A.E., 2017. Thoughts on interactions between PGRMC1 and
7 diverse attested and potential hydrophobic ligands. *J Steroid Biochem Mol Biol* 171, 11-
8 33.
- 9 Cai, H.L., Tan, Q.Y., Jiang, P., Dang, R.L., Xue, Y., Tang, M.M., Xu, P., Deng, Y., Li,
10 H.D., Yao, J.K., 2015. A potential mechanism underlying atypical antipsychotics-induced
11 lipid disturbances. *Translational psychiatry* 5, e661.
- 12 Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for
13 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*
14 (Oxford, England) 25, 1972-1973.
- 15 Castelle, C.J., Banfield, J.F., 2018. Major New Microbial Groups Expand Diversity and
16 Alter our Understanding of the Tree of Life. *Cell* 172, 1181-1197.
- 17 Castelle, C.J., Brown, C.T., Anantharaman, K., Probst, A.J., Huang, R.H., Banfield, J.F.,
18 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and
19 DPANN radiations. *Nature reviews. Microbiology* 16, 629-645.
- 20 Chen, I.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M.,
21 Varghese, N., White, J.R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S.P., Woyke,
22 T., Elie-Fadrosh, E.A., Ivanova, N.N., Kyrpides, N.C., 2019. IMG/M v.5.0: an integrated
23 data management and comparative analysis system for microbial genomes and
24 microbiomes. *Nucleic Acids Res* 47, D666-D677.
- 25 Cho, I.T., Adelmant, G., Lim, Y., Marto, J.A., Cho, G., Golden, J.A., 2017. Ascorbate
26 peroxidase proximity labeling coupled with biochemical fractionation identifies
27 promoters of endoplasmic reticulum-mitochondrial contacts. *The Journal of biological*
28 *chemistry* 292, 16382-16392.
- 29 Collingridge, P.W., Kelly, S., 2012. MergeAlign: improving multiple sequence alignment
30 performance by dynamic reconstruction of consensus multiple sequence alignments.
31 *BMC bioinformatics* 13, 117.
- 32 Cowieson, N.P., Aragao, D., Clift, M., Ericsson, D.J., Gee, C., Harrop, S.J., Mudie, N.,
33 Panjikar, S., Price, J.R., Riboldi-Tunnicliffe, A., Williamson, R., Caradoc-Davies, T.,
34 2015. MX1: a bending-magnet crystallography beamline serving both chemical and
35 macromolecular crystallography communities at the Australian Synchrotron. *Journal of*
36 *synchrotron radiation* 22, 187-190.
- 37 Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo
38 generator. *Genome research* 14, 1188-1190.
- 39 Darnet, S., Schaller, H., 2019. Metabolism and Biological Activities of 4-Methyl-Sterols.
40 *Molecules* (Basel, Switzerland) 24, 451.
- 41 Deng, B., Parthasarathy, S., Wang, W., Gibney, B.R., Battaile, K.P., Lovell, S., Benson,
42 D.R., Zhu, H., 2010. Study of the individual cytochrome b5 and cytochrome b5 reductase
43 domains of Ncb5or reveals a unique heme pocket and a possible role of the CS domain.
44 *The Journal of biological chemistry* 285, 30181-30191.
- 45 Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of
46 Coot. *Acta crystallographica. Section D, Biological crystallography* 66, 486-501.
- 47 Evans, P.R., 2011. An introduction to data reduction: space-group determination, scaling
48 and intensity statistics. *Acta crystallographica. Section D, Biological crystallography* 67,
49 282-292.
- 50 Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution?
51 *Acta crystallographica. Section D, Biological crystallography* 69, 1204-1214.

- 1 Fakheri, R.J., Javitt, N.B., 2011. Autoregulation of cholesterol synthesis: physiologic and
- 2 pathophysiologic consequences. *Steroids* 76, 211-215.
- 3 Frickey, T., Kannenberg, E., 2009. Phylogenetic analysis of the triterpene cyclase protein
- 4 family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer.
- 5 *Environmental microbiology* 11, 1224-1241.
- 6 Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the
- 7 next-generation sequencing data. *Bioinformatics (Oxford, England)* 28, 3150-3152.
- 8 Galmozzi, A., Kok, B.P., Kim, A.S., Montenegro-Burke, J.R., Lee, J.Y., Spreafico, R.,
- 9 Mosure, S., Albert, V., Cintron-Colon, R., Godio, C., Webb, W.R., Conti, B., Solt, L.A.,
- 10 Kojetin, D., Parker, C.G., Peluso, J.J., Pru, J.K., Siuzdak, G., Cravatt, B.F., Saez, E., 2019.
- 11 PGRMC2 is an intracellular haem chaperone critical for adipocyte function. *Nature* 576,
- 12 138-142.
- 13 Garriga, E., Di Tommaso, P., Magis, C., Erb, I., Mansouri, L., Baltzis, A., Laayouni, H.,
- 14 Kondrashov, F., Floden, E., Notredame, C., 2019. Large multiple sequence alignments
- 15 with a root-to-leaf regressive method. *Nat Biotechnol* 37, 1466-1470.
- 16 Griffin, D., Liu, X., Pru, C., Pru, J.K., Peluso, J.J., 2014. Expression of progesterone
- 17 receptor membrane component-2 within the immature rat ovary and its role in regulating
- 18 mitosis and apoptosis of spontaneously immortalized granulosa cells. *Biology of*
- 19 *reproduction* 91, 36.
- 20 Guo, M., Zhang, C., Wang, Y., Feng, L., Wang, Z., Niu, W., Du, X., Tang, W., Li, Y.,
- 21 Wang, C., Chen, Z., 2016. Progesterone Receptor Membrane Component 1 Mediates
- 22 Progesterone-Induced Suppression of Oocyte Meiotic Prophase I and Primordial
- 23 Folliculogenesis. *Scientific reports* 6, 36869.
- 24 Guy, L., Kultima, J.R., Andersson, S.G., 2010. genoPlotR: comparative gene and genome
- 25 visualization in R. *Bioinformatics (Oxford, England)* 26, 2334-2335.
- 26 Hampton, K.K., Anderson, K., Frazier, H., Thibault, O., Craven, R.J., 2018. Insulin
- 27 Receptor Plasma Membrane Levels Increased by the Progesterone Receptor Membrane
- 28 Component 1. *Molecular pharmacology* 94, 665-673.
- 29 Han, M.V., Zmasek, C.M., 2009. phyloXML: XML for evolutionary biology and
- 30 comparative genomics. *BMC bioinformatics* 10, 356.
- 31 Hand, R.A., Jia, N., Bard, M., Craven, R.J., 2003. *Saccharomyces cerevisiae* Dap1p, a
- 32 novel DNA damage response protein related to the mammalian membrane-associated
- 33 progesterone receptor. *Eukaryotic cell* 2, 306-317.
- 34 Heese-Peck, A., Pichler, H., Zanolari, B., Watanabe, R., Daum, G., Riezman, H., 2002.
- 35 Multiple functions of sterols in yeast endocytosis. *Molecular biology of the cell* 13, 2664-
- 36 2680.
- 37 Hehenberger, E., Eitel, M., Fortunato, S.A.V., Miller, D.J., Keeling, P.J., Cahill, M.A.,
- 38 2020. Early eukaryotic origins and metazoan elaboration of MAPR family proteins.
- 39 *Molecular phylogenetics and evolution* 148, 106814.
- 40 Holm, L., Laakso, L.M., 2016. Dali server update. *Nucleic Acids Res* 44, W351-355.
- 41 Hoshino, Y., Gaucher, E.A., 2018. On the Origin of Isoprenoid Biosynthesis. *Molecular*
- 42 *biology and evolution* 35, 2185-2197.
- 43 Hughes, A.L., Lee, C.Y., Bien, C.M., Espenshade, P.J., 2007a. 4-Methyl sterols regulate
- 44 fission yeast SREBP-Scap under low oxygen and cell stress. *The Journal of biological*
- 45 *chemistry* 282, 24388-24396.
- 46 Hughes, A.L., Powell, D.W., Bard, M., Eckstein, J., Barbuch, R., Link, A.J., Espenshade,
- 47 P.J., 2007b. Dap1/PGRMC1 binds and regulates cytochrome P450 enzymes. *Cell*
- 48 *metabolism* 5, 143-149.
- 49 Hughes, B.T., Espenshade, P.J., 2008. Oxygen-regulated degradation of fission yeast
- 50 SREBP by Ofd1, a prolyl hydroxylase family member. *The EMBO journal* 27, 1491-
- 51 1501.

- 1 Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano,
2 Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai,
3 S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., Kamagata, Y., Tamaki, H.,
4 Takai, K., 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature*
5 577, 519-525.
- 6 Izzo, N.J., Colom-Cadena, M., Riad, A.A., Xu, J., Singh, M., Abate, C., Cahill, M.A.,
7 Spires-Jones, T.L., Bowen, W.D., Mach, R.H., Catalano, S.M., 2020. Proceedings from
8 the Fourth International Symposium on sigma-2 Receptors: Role in Health and Disease.
9 *eNeuro* 7.
- 10 Izzo, N.J., Xu, J., Zeng, C., Kirk, M.J., Mozzoni, K., Silky, C., Rehak, C., Yurko, R.,
11 Look, G., Rishton, G., Safferstein, H., Cruchaga, C., Goate, A., Cahill, M.A., Arancio,
12 O., Mach, R.H., Craven, R., Head, E., LeVine, H., 3rd, Spires-Jones, T.L., Catalano, S.M.,
13 2014. Alzheimer's therapeutics targeting amyloid beta 1-42 oligomers II: Sigma-
14 2/PGRMC1 receptors mediate Abeta 42 oligomer binding and synaptotoxicity. *PloS one*
15 9, e111899.
- 16 Jaffe, A.L., Castelle, C.J., Matheus Carnevali, P.B., Gribaldo, S., Banfield, J.F., 2020.
17 The rise of diversity in metabolic platforms across the Candidate Phyla Radiation. *BMC*
18 *biology* 18, 69.
- 19 Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High
20 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
21 *Nature communications* 9, 5114.
- 22 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., Madden, T.L.,
23 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res* 36, W5-9.
- 24 Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H.,
25 Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A.,
26 Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-
27 scale protein function classification. *Bioinformatics (Oxford, England)* 30, 1236-1240.
- 28 Juhlen, R., Landgraf, D., Huebner, A., Koehler, K., 2018. Triple A patient cells suffering
29 from mitotic defects fail to localize PGRMC1 to mitotic kinetochore fibers. *Cell division*
30 13, 8.
- 31 Kabe, Y., Nakane, T., Koike, I., Yamamoto, T., Sugiura, Y., Harada, E., Sugase, K.,
32 Shimamura, T., Ohmura, M., Muraoka, K., Yamamoto, A., Uchida, T., Iwata, S.,
33 Yamaguchi, Y., Krayukhina, E., Noda, M., Handa, H., Ishimori, K., Uchiyama, S.,
34 Kobayashi, T., Suematsu, M., 2016a. Haem-dependent dimerization of PGRMC1/Sigma-
35 2 receptor facilitates cancer proliferation and chemoresistance. *Nature communications*
36 7, 11030.
- 37 Kabe, Y., Yamamoto, T., Kajimura, M., Sugiura, Y., Koike, I., Ohmura, M., Nakamura,
38 T., Tokumoto, Y., Tsugawa, H., Handa, H., Kobayashi, T., Suematsu, M., 2016b.
39 Cystathionine beta-synthase and PGRMC1 as CO sensors. *Free radical biology &*
40 *medicine* 99, 333-344.
- 41 Kaluka, D., Batabyal, D., Chiang, B.Y., Poulos, T.L., Yeh, S.R., 2015. Spectroscopic and
42 mutagenesis studies of human PGRMC1. *Biochemistry* 54, 1638-1647.
- 43 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermin, L.S., 2017.
44 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*
45 14, 587-589.
- 46 Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence
47 alignment, interactive sequence choice and visualization. *Briefings in bioinformatics* 20,
48 1160-1166.
- 49 Khandokar, Y.B., Srivastava, P., Cowieson, N., Sarker, S., Aragao, D., Das, S., Smith,
50 K.M., Raidal, S.R., Forwood, J.K., 2017. Structural insights into GDP-mediated
51 regulation of a bacterial acyl-CoA thioesterase. *The Journal of biological chemistry* 292,
52 20461-20471.

- 1 Kirschvink, J.L., Kopp, R.E., 2008. Palaeoproterozoic ice houses and the evolution of
2 oxygen-mediating enzymes: the case for a late origin of photosystem II. Philosophical
3 transactions of the Royal Society of London. Series B, Biological sciences 363, 2755-
4 2765.
- 5 Kisker, C., Schindelin, H., Pacheco, A., Wehbi, W.A., Garrett, R.M., Rajagopalan, K.V.,
6 Enemark, J.H., Rees, D.C., 1997. Molecular basis of sulfite oxidase deficiency from the
7 structure of sulfite oxidase. Cell 91, 973-983.
- 8 Kostanjevecki, V., Leys, D., Van Driessche, G., Meyer, T.E., Cusanovich, M.A., Fischer,
9 U., Guisez, Y., Van Beeumen, J., 1999. Structure and characterization of
10 Ectothiorhodospira vacuolata cytochrome b(558), a prokaryotic homologue of
11 cytochrome b(5). The Journal of biological chemistry 274, 35614-35620.
- 12 Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast
13 protein structure alignment in three dimensions. Acta crystallographica. Section D,
14 Biological crystallography 60, 2256-2268.
- 15 Kuraku, S., Zmasek, C.M., Nishimura, O., Katoh, K., 2013. aLeaves facilitates on-
16 demand exploration of metazoan gene family trees on MAFFT sequence alignment server
17 with enhanced interactivity. Nucleic Acids Res 41, W22-28.
- 18 Kurland, C.G., Andersson, S.G., 2000. Origin and evolution of the mitochondrial
19 proteome. Microbiology and molecular biology reviews : MMBR 64, 786-820.
- 20 Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large
21 datasets. Bioinformatics (Oxford, England) 30, 3276-3278.
- 22 Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: multiple ligand-protein interaction
23 diagrams for drug discovery. Journal of chemical information and modeling 51, 2778-
24 2786.
- 25 Lee, S.R., Kwon, S.W., Kaya, P., Lee, Y.H., Lee, J.G., Kim, G., Lee, G.S., Baek, I.J.,
26 Hong, E.J., 2018. Loss of progesterone receptor membrane component 1 promotes
27 hepatic steatosis via the induced de novo lipogenesis. Scientific reports 8, 15711.
- 28 Lemoine, F., Domelevo Entfellner, J.B., Wilkinson, E., Correia, D., Davila Felipe, M.,
29 De Oliveira, T., Gascuel, O., 2018. Renewing Felsenstein's phylogenetic bootstrap in the
30 era of big data. Nature 556, 452-456.
- 31 Liu, J., Chakraborty, S., Hosseinzadeh, P., Yu, Y., Tian, S., Petrik, I., Bhagi, A., Lu, Y.,
32 2014. Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers.
33 Chemical reviews 114, 4366-4469.
- 34 Loidl, J., 2016. Conservation and Variability of Meiosis Across the Eukaryotes. Annual
35 review of genetics 50, 293-316.
- 36 Luciano, A.M., Lodde, V., Franciosi, F., Ceciliani, F., Peluso, J.J., 2010. Progesterone
37 receptor membrane component 1 expression and putative function in bovine oocyte
38 maturation, fertilization, and early embryonic development. Reproduction (Cambridge,
39 England) 140, 663-672.
- 40 Luciano, A.M., Peluso, J.J., 2016. PGRMC1 and the faithful progression through mitosis
41 and meiosis. Cell cycle (Georgetown, Tex.) 15, 2239-2240.
- 42 Mavromatis, K., Chu, K., Ivanova, N., Hooper, S.D., Markowitz, V.M., Kyrpides, N.C.,
43 2009. Gene context analysis in the Integrated Microbial Genomes (IMG) data
44 management system. PloS one 4, e7979.
- 45 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read,
46 R.J., 2007. Phaser crystallographic software. Journal of applied crystallography 40, 658-
47 674.
- 48 McGuire, M.R., Mukhopadhyay, D., Myers, S.L., Mosher, E.P., Brookheart, R.T.,
49 Kammers, K., Sehgal, A., Selen, E.S., Wolfgang, M.J., Bumpus, N.N., Espenshade, P.J.,
50 2021. Progesterone receptor membrane component 1 (PGRMC1) binds and stabilizes
51 cytochromes P450 through a heme-independent mechanism. The Journal of biological
52 chemistry, 101316.

- 1 Meunier, F.A., Gutierrez, L.M., 2016. Captivating New Roles of F-Actin Cortex in
- 2 Exocytosis and Bulk Endocytosis in Neurosecretory Cells. *Trends Neurosci* 39, 605-613.
- 3 Mifsud, W., Bateman, A., 2002. Membrane-bound progesterone receptors contain a
- 4 cytochrome b5-like ligand-binding domain. *Genome Biology* 3, RESEARCH0068.
- 5 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von
- 6 Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for
- 7 Phylogenetic Inference in the Genomic Era. *Molecular biology and evolution* 37, 1530-
- 8 1534.
- 9 Mistry, J., Bateman, A., Finn, R.D., 2007. Predicting active site residue annotations in the
- 10 Pfam database. *BMC bioinformatics* 8, 298.
- 11 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer,
- 12 E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A.,
- 13 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* 49, D412-D419.
- 14 Mitsche, M.A., McDonald, J.G., Hobbs, H.H., Cohen, J.C., 2015. Flux analysis of
- 15 cholesterol biosynthesis in vivo reveals multiple tissue and cell-type specific pathways.
- 16 *eLife* 4, e07999.
- 17 Neubauer, H., Clare, S.E., Wozny, W., Schwall, G.P., Poznanovic, S., Stegmann, W.,
- 18 Vogel, U., Sotlar, K., Wallwiener, D., Kurek, R., Fehm, T., Cahill, M.A., 2008. Breast
- 19 cancer proteomics reveals correlation between estrogen receptor status and differential
- 20 phosphorylation of PGRMC1. *Breast cancer research : BCR* 10, R85.
- 21 Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., Maltsev, N., 1999. The use of
- 22 gene clusters to infer functional coupling. *Proceedings of the National Academy of*
- 23 *Sciences of the United States of America* 96, 2896-2901.
- 24 Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., Yeates, T.O., 1999.
- 25 Assigning protein functions by comparative genome analysis: protein phylogenetic
- 26 profiles. *Proceedings of the National Academy of Sciences of the United States of*
- 27 *America* 96, 4285-4288.
- 28 Peluso, J.J., Griffin, D., Liu, X., Horne, M., 2014. Progesterone receptor membrane
- 29 component-1 (PGRMC1) and PGRMC-2 interact to suppress entry into the cell cycle in
- 30 spontaneously immortalized rat granulosa cells. *Biology of reproduction* 91, 104.
- 31 Peluso, J.J., Pru, J.K., 2014. Non-canonical progesterone signaling in granulosa cell
- 32 function. *Reproduction (Cambridge, England)* 147, R169-178.
- 33 Piel, R.B., 3rd, Shiferaw, M.T., Vashisht, A.A., Marcero, J.R., Praissman, J.L., Phillips,
- 34 J.D., Wohlschlegel, J.A., Medlock, A.E., 2016. A Novel Role for Progesterone Receptor
- 35 Membrane Component 1 (PGRMC1): A Partner and Regulator of Ferrochelatase.
- 36 *Biochemistry* 55, 5204-5217.
- 37 Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2--approximately maximum-
- 38 likelihood trees for large alignments. *PloS one* 5, e9490.
- 39 Probst, A.J., Castelle, C.J., Singh, A., Brown, C.T., Anantharaman, K., Sharon, I., Hug,
- 40 L.A., Burstein, D., Emerson, J.B., Thomas, B.C., Banfield, J.F., 2017. Genomic
- 41 resolution of a cold subsurface aquifer community provides metabolic insights for novel
- 42 microbes adapted to high CO2 concentrations. *Environmental microbiology* 19, 459-474.
- 43 Probst, A.J., Ladd, B., Jarett, J.K., Geller-McGrath, D.E., Sieber, C.M.K., Emerson, J.B.,
- 44 Anantharaman, K., Thomas, B.C., Malmstrom, R.R., Stieglmeier, M., Klingl, A., Woyke,
- 45 T., Ryan, M.C., Banfield, J.F., 2018. Differential depth distribution of microbial function
- 46 and putative symbionts through sediment-hosted aquifers in the deep terrestrial
- 47 subsurface. *Nature microbiology* 3, 328-336.
- 48 Rajamani, R., Gao, J., 2003. Balancing kinetic and thermodynamic control: the
- 49 mechanism of carbocation cyclization by squalene cyclase. *Journal of the American*
- 50 *Chemical Society* 125, 12768-12781.
- 51 Riad, A., Lengyel-Zhand, Z., Zeng, C., Weng, C.C., Lee, V.M., Trojanowski, J.Q., Mach,
- 52 R.H., 2020. The Sigma-2 Receptor/TMEM97, PGRMC1, and LDL Receptor Complex

- 1 Are Responsible for the Cellular Uptake of Abeta42 and Its Protein Aggregates.
- 2 *Molecular neurobiology* 57, 3803-3813.
- 3 Riad, A., Zeng, C., Weng, C.C., Winters, H., Xu, K., Makvandi, M., Metz, T., Carlin, S.,
- 4 Mach, R.H., 2018. Sigma-2 Receptor/TMEM97 and PGRMC-1 Increase the Rate of
- 5 Internalization of LDL by LDL Receptor through the Formation of a Ternary Complex.
- 6 *Scientific reports* 8, 16845.
- 7 Rodriguez-Maranon, M.J., Qiu, F., Stark, R.E., White, S.P., Zhang, X., Foundling, S.I.,
- 8 Rodriguez, V., Schilling, C.L., 3rd, Bunce, R.A., Rivera, M., 1996. ¹³C NMR
- 9 spectroscopic and X-ray crystallographic study of the role played by mitochondrial
- 10 cytochrome b5 heme propionates in the electrostatic binding to cytochrome c.
- 11 *Biochemistry* 35, 16378-16390.
- 12 Ruan, X., Schneck, H., Schultz, S., Fehm, T., Cahill, M.A., Seeger, H., Chen, R., Yu, Q.,
- 13 Mueck, A.O., Neubauer, H., 2012. Nomegestrol acetate sequentially or continuously
- 14 combined to estradiol did not negatively affect membrane-receptor associated
- 15 progestogenic effects in human breast cancer cells. *Gynecological endocrinology : the*
- 16 *official journal of the International Society of Gynecological Endocrinology* 28, 863-866.
- 17 Ryu, C.S., Klein, K., Zanger, U.M., 2017. Membrane Associated Progesterone Receptors:
- 18 Promiscuous Proteins with Pleiotropic Functions - Focus on Interactions with
- 19 Cytochromes P450. *Frontiers in pharmacology* 8, 159.
- 20 Sabbir, M.G., 2019. Progesterone induced Warburg effect in HEK293 cells is associated
- 21 with post-translational modifications and proteasomal degradation of progesterone
- 22 receptor membrane component 1. *J Steroid Biochem Mol Biol* 191, 105376.
- 23 Sabbir, M.G., Taylor, C.G., Zahradka, P., 2020. Progesterone-induced Warburg Effect is
- 24 Regulated by Cell-type-specific Interaction of Progesterone Receptor Membrane
- 25 Component 1 and Hexokinases. *Research Square* (non-peer reviewed preprint)
- 26 <https://doi.org/10.21203/rs.3.rs-55518/v1>.
- 27 Saenz, J.P., Grosser, D., Bradley, A.S., Lagny, T.J., Lavrynenko, O., Broda, M., Simons,
- 28 K., 2015. Hopanoids as functional analogues of cholesterol in bacterial membranes.
- 29 *Proceedings of the National Academy of Sciences of the United States of America* 112,
- 30 11971-11976.
- 31 Salsano, S., Gonzalez-Martin, R., Quinonero, A., Lopez-Martin, S., Gomez-Escribano,
- 32 A.P., Perez-Deben, S., Yanez-Mo, M., Dominguez, F., 2020. Novel nonclassic
- 33 progesterone receptor PGRMC1 pulldown-precipitated proteins reveal a key role during
- 34 human decidualization. *Fertility and sterility* 113, 1050-1066 e1057.
- 35 Schroder, I., Johnson, E., de Vries, S., 2003. Microbial ferric iron reductases. *FEMS*
- 36 *microbiology reviews* 27, 427-447.
- 37 Sebe-Pedros, A., Ballare, C., Parra-Acero, H., Chiva, C., Tena, J.J., Sabido, E., Gomez-
- 38 Skarmeta, J.L., Di Croce, L., Ruiz-Trillo, I., 2016. The Dynamic Regulatory Genome of
- 39 *Capsaspora* and the Origin of Animal Multicellularity. *Cell* 165, 1224-1237.
- 40 Shimano, H., Sato, R., 2017. SREBP-regulated lipid metabolism: convergent physiology
- 41 - divergent pathophysiology. *Nat Rev Endocrinol* 13, 710-730.
- 42 Song, J., Vinarov, D., Tyler, E.M., Shahan, M.N., Tyler, R.C., Markley, J.L., 2004.
- 43 Hypothetical protein At2g24940.1 from *Arabidopsis thaliana* has a cytochrome b5 like
- 44 fold. *Journal of biomolecular NMR* 30, 215-218.
- 45 Stromstedt, M., Rozman, D., Waterman, M.R., 1996. The ubiquitously expressed human
- 46 CYP51 encodes lanosterol 14 alpha-demethylase, a cytochrome P450 whose expression
- 47 is regulated by oxysterols. *Archives of biochemistry and biophysics* 329, 73-81.
- 48 Suchanek, M., Radzikowska, A., Thiele, C., 2005. Photo-leucine and photo-methionine
- 49 allow identification of protein-protein interactions in living cells. *Nature methods* 2, 261-
- 50 267.

- 1 Takishita, K., Chikaraishi, Y., Tanifuji, G., Ohkouchi, N., Hashimoto, T., Fujikura, K.,
- 2 Roger, A.J., 2017. Microbial Eukaryotes that Lack Sterols. *The Journal of eukaryotic*
- 3 *microbiology* 64, 897-900.
- 4 Teakel, S.L., Ludescher, M., Thejer, B.M., Poschmann, G., Forwood, J.K., Neubauer, H.,
- 5 Cahill, M.A., 2020. Protein complexes including PGRMC1 and actin-associated proteins
- 6 are disrupted by AG-205. *Biochemical and biophysical research communications* 524,
- 7 64-69.
- 8 Terzaghi, L., Tessaro, I., Raucci, F., Merico, V., Mazzini, G., Garagna, S., Zuccotti, M.,
- 9 Franciosi, F., Lodde, V., 2016. PGRMC1 participates in late events of bovine granulosa
- 10 cells mitosis and oocyte meiosis. *Cell cycle (Georgetown, Tex.)* 15, 2019-2032.
- 11 Thejer, B.M., Adhikary, P.P., Kaur, A., Teakel, S.L., Van Oosterum, A., Seth, I., Pajic,
- 12 M., Hannan, K.M., Pavy, M., Poh, P., Jazayeri, J.A., Zaw, T., Pascovici, D., Ludescher,
- 13 M., Pawlak, M., Cassano, J.C., Turnbull, L., Jazayeri, M., James, A.C., Coorey, C.P.,
- 14 Roberts, T.L., Kinder, S.J., Hannan, R.D., Patrick, E., Molloy, M.P., New, E.J., Fehm,
- 15 T.N., Neubauer, H., Goldys, E.M., Weston, L.A., Cahill, M.A., 2020a. PGRMC1
- 16 phosphorylation affects cell shape, motility, glycolysis, mitochondrial form and function,
- 17 and tumor growth. *BMC Mol Cell Biol* 21, 24.
- 18 Thejer, B.M., Adhikary, P.P., Teakel, S.L., Fang, J., Weston, P.A., Gurusinghe, S.,
- 19 Anwer, A.G., Gosnell, M., Jazayeri, J.A., Ludescher, M., Gray, L.A., Pawlak, M.,
- 20 Wallace, R.H., Pant, S.D., Wong, M., Fischer, T., New, E.J., Fehm, T.N., Neubauer, H.,
- 21 Goldys, E.M., Quinn, J.C., Weston, L.A., Cahill, M.A., 2020b. PGRMC1 effects on
- 22 metabolism, genomic mutation and CpG methylation imply crucial roles in animal
- 23 biology and disease. *BMC Mol Cell Biol* 21, 26.
- 24 Thompson, A.M., Reddi, A.R., Shi, X., Goldbeck, R.A., Moenne-Loccoz, P., Gibney,
- 25 B.R., Holman, T.R., 2007. Measurement of the heme affinity for yeast dap1p, and its
- 26 importance in cellular function. *Biochemistry* 46, 14629-14637.
- 27 Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F.,
- 28 Murshudov, G.N., 2004. REFMAC5 dictionary: organization of prior chemical
- 29 knowledge and guidelines for its use. *Acta crystallographica. Section D, Biological*
- 30 *crystallography* 60, 2184-2195.
- 31 Wang, H.C., Minh, B.Q., Susko, E., Roger, A.J., 2018. Modeling Site Heterogeneity with
- 32 Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation.
- 33 *Syst Biol* 67, 216-235.
- 34 Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes,
- 35 N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., Long, P.E., Banfield,
- 36 J.F., 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated
- 37 bacterial phyla. *Science (New York, N.Y.)* 337, 1661-1665.
- 38 Xu, J., Zeng, C., Chu, W., Pan, F., Rothfuss, J.M., Zhang, F., Tu, Z., Zhou, D., Zeng, D.,
- 39 Vangveravong, S., Johnston, F., Spitzer, D., Chang, K.C., Hotchkiss, R.S., Hawkins,
- 40 W.G., Wheeler, K.T., Mach, R.H., 2011. Identification of the PGRMC1 protein complex
- 41 as the putative sigma-2 receptor binding site. *Nature communications* 2, 380.
- 42 Yamada, K.D., Tomii, K., Katoh, K., 2016. Application of the MAFFT sequence
- 43 alignment program to large data-reexamination of the usefulness of chained guide trees.
- 44 *Bioinformatics (Oxford, England)* 32, 3246-3251.
- 45 Yokota, T., Nakajima, Y., Yamakura, F., Sugio, S., Hashimoto, M., Takamiya, S., 2006.
- 46 Unique structure of *Ascaris suum* b5-type cytochrome: an additional alpha-helix and
- 47 positively charged residues on the surface domain interact with redox partners. *The*
- 48 *Biochemical journal* 394, 437-447.
- 49 Yoshitani, N., Satou, K., Saito, K., Suzuki, S., Hatanaka, H., Seki, M., Shinozaki, K.,
- 50 Hirota, H., Yokoyama, S., 2005. A structure-based strategy for discovery of small ligands
- 51 binding to functionally unknown proteins: combination of in silico screening and surface
- 52 plasmon resonance measurements. *Proteomics* 5, 1472-1480.

1 Figures

2

3

4 **Fig. 1. The production of FF-MAS from the mevalonate pathway. Enzymes**

5 involved in mammalian production and 14-demethylation of of the 4 methyl-sterol

6 lanosterol. MVP: mevalonate pathway. The CYP51A/PGRMC1 reaction produces 14-

7 dimethyl-14-dehydrolanosterol, also known as follicular fluid meiosis-activating sterol

8 (FF-MAS). For CYP51A reaction stoichiometry see UniProt Q16850.

9

1

2

3 **Fig. 2. Eukaryotic cytochrome b5 domain proteins are polyphyletic.** Maximum-
4 likelihood phylogenetic reconstruction using IQ-Tree2 under LG+C60+R4+F+PMSF
5 model of cytb5 sequences reduced at 90% (A) and 65% (B) redundancy. Both trees
6 were rooted on the branch leading to Clade 1 bacterial sequences. Branch symbols
7 indicate Felsenstein Bootstrap Proportions (FBP; upper left half) and Transfer Bootstrap
8 Expectation (TBE; lower right half) interpretations of 100 non-parametric bootstrap
9 pseudoreplicates. Critical thresholds for TBE were calculated as explained in the
10 methods. Additional phylogenetic reconstructions under these and other levels of
11 redundancy, alternative alignment and trimming strategies, and alternative evolutionary
12 models, are all summarized in Supplementary Appendix D. Sequences, alignments and
13 trees are included in Supplementary Appendix C.

14

Fig. 3. A conserved clade-1 surface. (A) Sequence logos of the alignment of Fig. A1A. The box and “?” in clade-2 show a potentially incorrect inter-clade alignment, in which case the adjacent position (H165) is absent from clade-2. “Loop-1” and “loop-2” are shown. (B) A conserved clade-1 surface. Loop 1 is shown as open mesh to display conserved underlying F81 and G83. (C) Rotation through 90° relative to “B”, showing conserved exposed surface residues as filled spheres. (D) Residues conserved between all cytb₅-domain sequences in Fig. A1A (both clades) constitute the protein interior, as generated by Consurf (Ashkenazy et al., 2016), based upon 4X8Y.

Fig. 4. The crystal structure of Hadesarchaea cytb_{5M}. (A) Hadesarchaea YNP_N21 (NCBI KUO41884.1) cytb_{5M} structure. (B) Topological organization of the structure from A). Numbering follows KUO41884.1. (C) Alignment of loop-1 regions of bacterial cytb₅ (1CXY, *Ectothiorhodospira vacuolata*) (Kostanjevecki et al., 1999), rat mitochondrial cytb₅ (1B5M_A) (Rodriguez-Maranon et al., 1996), human PGRMC1 (4X8Y) (Kabe et al., 2016a), and KUO41884.1. The position of KUO41884.1 G14 and PGRMC1 G83 is arrowed at the base of the clade-1-specific loop (PGRMC1 84-88). (D) Heme orientation differs by typically 90 degrees between clades-1 and -2. (E)

- 1 Contacts with heme. The four proteins are shown in the alignment from Fig. A1A,
- 2 showing heme-interacting residues. See Fig. A5 for interaction details.

3

Fig. 5. Phylogenetic distribution and CPR genomic context of $cytb_{5MY}$ genes. (A)

Phyletic distribution of $cytb_{5MY}$, mapped onto a maximum likelihood tree obtained from a concatenated alignment of 16 ribosomal proteins by Jaffe et al. (2020). Genome assignment was performed through ANI value calculation: circles indicate assignment with an ANI value larger than 70%; circles are outlined if they indicate an assignment larger than 99%), and larger if they indicate assignment of more than one genome. Only the taxonomy of groups with $cytb_{5MY}$ -containing genomes is indicated. (B) 5-kb gene neighbourhoods of $cytb_{5MY}$ and pFre genes of selected genomes. Each arrow represents

- 1 a gene and they are colored based on assigned function according to the legend.
- 2 Connecting lines are drawn between homologous genes found as best reciprocal Blastp
- 3 hits with e-values under $1e-5$.
- 4
- 5

1

2

3

4

Fig. 6. CYP51A is of bacterial origin. Maximum-likelihood phylogenetic

5

reconstruction using IQ-Tree2 under the LG+G4+F model. Branch symbols represent

6

SH-like Approximate Ratio Test (upper left) and Ultrafast bootstrap (lower right), both

7

using 1000 pseudoreplicates. Input alignment contained 533 sequences and 488 sites

8

(Supplementary Appendix F).

9

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

Fig. A1. BLASTp hits from MAPR or cytochrome b5 protein search strings form two discrete clades and MAPR proteins are related to prokaryotic cytb_{5MY}. (A) Neighbour-Joining 1,000 bootstrap maximum likelihood phylogenetic tree under the Poisson substitution model, constructed from the alignment of 176 protein sequences by MAFFT. (B) Neighbour-Joining phylogenetic tree under the Poisson substitution model, constructed as per (A) from the alignment of 814 protein sequences. Sequences from preliminary analyses are combined with independent eukaryotic MAPR and cytb₅ sequences (see methods) in a 65 gap-free site alignment. 11 proteins clustering closer to MAPR than other bacterial proteins are indicated (cytb_{5MY}). See Supplementary Appendix A for FASTA and tree files of all preliminary investigations, as well as the combined tree depicted in this panel. (C) MAFFT alignment of MAPR with bacterial HGT and cytb_{5MY} proteins. Heme-interacting residues and the MIHIR in PGRMC1 are indicated. The heme interaction key follows the PGRMC1 4X8Y structure (Kabe et al., 2016a). Alignment with representative clade-2 cytb_{5M} sequence (KUO41884.1, Hadesarchaea cytb_{5M}), and eukaryotic (NP_085075.1, rat cytb_{5B}) or bacterial (1CXY_A *Ectothiorhodospira vacuolata*) clade-1 proteins is provided below.

1 **A**
2 **GGATCCA**ATATT**GAGAACCTGTACTTCCAG\AGC**ATGCGTGTGTTTACCAAAGAGGAGCTGAGC
3 CGTTATAACGGTAAAGAAGGCGCGCCGGCGTACGTTGCGTATAACGGCAAGGTGTACGACGTTA
4 CCGGCAGCTTCCACTGGAAGGGTGGCAAACACCACGTGCTGCACGACGCGGGTCAGGACCTGAC
5 CGAGAGCATTGGTCGTGCGCCGCACACCGCGGAGCTGCTGGAAAAATTTCCGGTGGTTGGTGT
6 CTGCGTGGCTAAAAATATT**GAATTC**
7 **B**
8 **\S**MRVFTKEELSRNGKEGAPAYVAYNGKVYDVTGSFHWKGGKHHVLHDAGQDLTESIGRAPHT
9 AELLEKFPVVGVLRG
10

11 **Fig. A2. Sequences of pGEX4T1_KUO41884 and the crystallised protein. (A)**
12 KUO41884.1 sequence was cloned into the BamH1 and EcoRI sites (bold) of pGEX4T1
13 with added Tobacco Etch Virus (TEV) cleavage site (boxed, cleavage site in translated
14 protein marked by \). The stop codon is underlined. (B) Sequence of the crystallized
15 protein, corresponding to KUO41884.1 with additional N-terminal S residue after TEV
16 cleavage.

17

1

2

3

4

Fig. A3. Reduced stringency BLAST does not detect a sequence continuum

(A) Schematic of reduced stringency BLASTp analysis using Word Size 3 (see Methods) against the bacterial data base. The indicated sequences from clade-1 and clade-2 were used as BLASTp search strings, and combined with three clade-1 (PGRC1 NP_006658.1, NENF NP_037481.1, KXH77621.1) and two clade-2 (CYB5B_RAT P04166, and WP_103012903.1) reference sequences, tagged “>RC2_” and “>RC2_” respectively. Iterative processing of the data as shown led to 818 gap-free sequences which were used for NJ treebuilding. Bootstrap support for separate clades 1 and 2 was 94% for JTT (depicted), 89% (WAG), and 93% (Poisson). Of the 159 sequences in clade-2 of this tree (excluding two clade-2 reference sequences), only 4 (WP_093087318.1, KKQ27723.1, KXK26085.1, GBD33942.1) were detected by both the clade-1 (KUO41884.1) and clade-2 (P04166) BLAST query sequences. (B) Using the same BLASTp search strings as panel a, the smaller archaeal data base was searched with Word Size 2 (WS2), followed by iterative MAFFT data refinement as schematically depicted. 103 WS2 hits were combined with the sequences from Fig. A1A and iteratively refined to provide a robust tree architecture of 932 sequences used for NJ tree building. Bootstrap support for the existence of separate clades 1 and 2 was 85% (JTT), 76% (WAG), and 86% (Poisson, depicted). Of the 34 WS2 sequences present in clade-2, 10 (AQS28389.1, RLI96648.1, RME54213.1, PIN98782.1, RME31891.1, RME52388.1, PIN95630.1, PIZ83262.1, OIO41921.1, RME53069.1) were also BLASTp hits for clade-1 search string KUO41884.1. (C) Word size 2 BLASTp with independent query sequences to panel b. Because of the relatively low hit number of panel b the analysis was repeated using the indicated two outlying clade-1 and clade-2 sequences from Fig. A1A. Iterative data refinement 103 WS2 hits were combined with sequences from Fig. A1A to yield ultimately 912 sequences with 57 gap-free sites for tree building. (D) NJ/Poisson tree (1,000 bootstrap) of the process from panel c, including nine WS2 sequences with 100% bootstrap support as outliers to clades 1 and 2 (green). Bootstrap support for clades 1 and 2 is below 70% for JTT, WAG and Poisson NJ trees (depicted), however a group of cytb5-related proteins form an outlying distinct clade with 100% bootstrap support. Of the 11 WS2 sequences

1 present in clade-2, none were also BLASTp hits for clade-1 search string OHE40815.1.
2 Of the 9 outlying WS2 sequences, one (KQM10489) was also a BLASTp hit for clade-1
3 search string OHE40815.1. (E) Identical tree to panel d, with nine WS2 outlying
4 sequences removed prior to MAFFT alignment. This tree contains 78 WS2 sequences
5 that fall within clades 1 or two, but none that cluster between those clades. Bootstrap
6 support for clades 1 and two is 93% (JTT), 89% (WAG) and 97% (Poisson, depicted).
7 11 proteins were detected by both clade-1 and clade-2 search query sequences. No
8 proteins intermediate in sequence between clades 1 and 2 were found. For identities of
9 all cytb-L proteins detected in CPR bacteria see Supplementary Appendix C.

10

1

2

Fig. A4. Phylogenetic reconstructions of *cytb₅MY* tree position with alternative sequence datasets. Maximum-likelihood phylogenetic reconstructions using IQ-Tree2 under the PMSF approximation of the LG+C60+R8+F model (A,B), LG+C50+R8+F (C), LG+C40+R8+F (D) and LG+C50+R8+F (E). Input alignments were based on the reduction of the 2470 gathered sequences at different levels of similarity: 50% (A), 65% (B), 70% (C), 80% (D) and 90% (E), resulting in 274, 570, 702, 968 and 1384 sequences, respectively. Branch symbols indicate Felsenstein Bootstrap Proportions (FBP; upper left half) and Transfer Bootstrap Expectation (TBE; lower right half) interpretations of 100 non-parametric bootstrap pseudoreplicates. Sequence details are available as Supplementary Appendix D.

1

A

Hadesarchaea YNP_N21, KUO41884 (clade-1)

Hydrogen bonds

<----- A T O M 1 ----->						<----- A T O M 2 ----->						
	Atom	Atom	Res	Res	Chain		Atom	Atom	Res	Res	Chain	Distance
	no.	name	name	no.			no.	name	name	no.		
1.	265	OG	SER	34	A	-->	603	O1D	HEM	101	A	2.88
2.	535	NZ	LYS	68	A	-->	601	O2D	HEM	101	A	2.67

2

B

4X8Y PGRMC1 (clade-1)

Hydrogen bonds

<----- A T O M 1 ----->						<----- A T O M 2 ----->					
Atom	Atom	Res	Res			Atom	Atom	Res	Res	Chain	Distance
no.	name	name	no.	Chain		no.	name	name	no.		
1.	337	OH	TYR	107	A -->	1815	O1D	HEM	500	A	2.69
2.	773	NZ	LYS	163	A -->	1816	O2D	HEM	500	A	2.53
3.	785	OH	TYR	164	A -->	1815	O1D	HEM	500	A	2.54

1

C

1B5M_A rat cytb₅ (clade-2)

1

D

1CXY *E. vacuolata* (clade-2)

Hydrogen bonds

<----- A T O M 1 ----->						<----- A T O M 2 ----->					
	Atom	Atom	Res	Res			Atom	Atom	Res	Res	
	no.	name	name	no.	Chain		no.	name	name	no.	Chain
1.	476	NZ	LYS	64	A	-->	655	O1A	HEM	204	A
2.	476	NZ	LYS	64	A	-->	682	O1D	HEM	204	A
3.	494	OH	TYR	66	A	-->	656	O2A	HEM	204	A
4.	521	N	SER	71	A	-->	682	O1D	HEM	204	A
5.	526	OG	SER	71	A	-->	683	O2D	HEM	204	A
											Distance
											2.96
											2.83
											2.51
											2.86
											2.62

1
2 **Fig. A5. Heme-interacting residues from Fig. 4E.** H-Bonds with heme propionate
3 groups are given at the top of respective panels, with LIGPLOTS below. (A)
4 Hadesarchaea YNP_N21, cytb_{5M} KUO41884.1 (clade-1). (B) Human PGRMC1 4X8Y
5 (Clade-1). Non-PGRMC1 N-terminal residues in the 4X8Y structure from the bacterial
6 expression vector that make contact with heme are indicated by a circled X. The N-
7 terminus is not proximal to heme in the dimeric structure (Fig. 3C). These residues
8 make contact from adjacent symmetry units. The PGRMC1 MAPR domain of 4X8Y

1 was also N-terminally truncated relative to Fig. A1B by 2 residues. N-terminal non-
2 PGRMC1 residues (from the expression vector fusion protein) of adjacent crystal
3 symmetry units contacting heme in the PGRMC1 4X8Y structure could artefactually
4 stabilize the heme-dependent dimer. A native PGRMC1 MAPR-domain structure would
5 be desirable to confirm biological heme-dependent PGRMC1 dimerization.
6 Hydrophobic interactions between heme and S68, P69 and E70 from 4X8Y PGRMC1
7 structure are not shown for Fig. 4E because these represent non-PGRMC1 residues from
8 the bacterial expression vector. (C) Rat mitochondrial cytb_{5B} 1B5M_A (Clade-2).
9 Numbering is according to the corresponding rat cytb₅ sequence NP_085075.1. (D) *E.*
10 *vacuolata* 1CXY (clade-2).

11

Fig. A6. The most similar non-MAPR cytb₅-domain published structures to Hadesarchaea cytb_{5M} are from clade-2. (A) The thirty PDB structures most similar by DALI to KUO41884.1 (Hadesarchaea cytb_{5M}, this study) were from seventeen proteins

1 (colored boxes). MAFFT alignment followed by NJ Poisson putative tree construction
2 (1,000 bootstrap) showed that all except the recognised MAPR proteins (4X8Y, 1J03,
3 1T0G) belonged to clade-2. Reference sequences include clade-1 proteins NP_006658.1
4 PGRMC1, NP_037481.1 NENF, and archaean cytb_{5M} KXH77621.1 (no structure
5 available), and clade-2 NP_085075.1 rat cytb₅. Note the KRR N-terminal PGRMC1
6 MAPR residues absent from 4X8Y. Other conventions follow Fig. A1. (B) Clade-2
7 proteins with solved structures most similar to KUO41884.1 all shared the HPG heme
8 iron chelation motif involving loop 2 (Fig. 4E). The HS motif was also present in all but
9 chicken sulfite oxidase (1SOX) (Kisker et al., 1997), the cytb₅-reductase domain of
10 NADH cytochrome b₅ oxidoreductase (3LF5) (Deng et al., 2010), and *Ascaris suum*
11 cytb₅ (1X3X). The latter is known to be an atypical cytb₅ protein (Yokota et al., 2006),
12 and also exhibited divergent tree clustering in “A”. Red: iron atom chelation. Grey:
13 adjacent conserved heme hydrophobic interaction residue (following Fig. 4E). The color
14 key corresponds to proteins shown in a. (C) All clade-2 proteins except 1X3X exhibit
15 similar heme orientation to reference 1CXY (left panel). In 1X3X the propionates are
16 rotated approximately 90° to the left (centre panel, arrows), whereas in clade-1 proteins
17 the propionates are rotated to the right relative to clade-2 (arrows, right panel), and bind
18 at a tilted plane relatively to clade-2, prominently permitted by the absence of loop 2.
19 Superimposable Arabidopsis MAPR structures 1J03 (Yoshitani et al., 2005) and 1T0G
20 (Song et al., 2004) do not include heme. Only 1J03 is included.

21

1 **Fig. A7.** Synteny of *cytb*_{5MY} and pFre gene clusters. Gene neighbourhoods are shown at
2 a 10-kb distance from *cytb*_{5MY} and pFre genes. Functional annotation of selected genes
3 is shown according to the legend colors. Arrows represent genes and thick vertical lines,
4 contig boundaries. Connecting lines indicate similarity found as best reciprocal blastp
5 hits with e-values lower than 1e-5. The tree shown at the left side is the phylogeny
6 shown in Fig. A8.

7
8
9

1

2

3 **Fig. A8. CPR cytb₅MY gene similarities.** Phylogenetic tree obtained by Fasttree2
4 including all sequences used for the phylogenetic tree in Fig A3B. Sequences included
5 in the previous analysis are marked in red. Only the cytb₅MY clade is shown expanded.

6

1

2

3

4 **Fig. A9. Affiliation of pFre-associated Cytb₅-L proteins.** (A) Neighbour-joining tree
5 under the Poisson model and showing 1,000 bootstrap pseudoreplicates, inferred from
6 916 sequences, including cytb₅-like type A (CPR_BLA) and or cytb₅-like type B1
7 (CPR1/2_BLB1) proteins (Fig. 5A) as well as the proteins from Fig. A1A. It was
8 impossible to align clades-1 and 2 together with both cytb₅-L type B1 and B2 proteins
9 and obtain sufficient gap-free aligned sites for tree building. This panel shows the tree

1 with CPR1/2_BLB2 sequences omitted. These sequences include those from Fig. A1A,
2 supplemented with CPR cytb₅-domain proteins detected in IMG analyses (labelled:
3 “CPRx”, and the results of BLASTp of a cytb₅MY search string (KUO41884.1) against
4 the NCBI CPR database (taxid:1783234) (labelled “tax_”). Identical sequences from the
5 three input sources were not removed. (B) All details are identical to “A”, except that
6 CPR1/2_BLB1 sequences were omitted and CPR2_BLB2 sequences included in the
7 alignment to give 912 MAFFT-aligned sequences. The position of the CPR_BLA clade
8 relative to cytb₅M is uncertain with only 32% and 34% bootstrap support in both
9 analyses. The corresponding WAG tree for panel “A” positioned the CPR1_BLA clade
10 deep within the cytb₅M branch (not shown). Complete tree files and PDF printouts of the
11 expanded trees for this figure are available as Supplementary Appendix E.
12

1

2 **Table A1. Details of cytb₅MY proteins from Fig. A1B.** All species are CPR bacteria
3 (Castelle and Banfield, 2018). The details of all cytb₅MY and cytb₅M proteins detected in
4 CPR bacteria are provided in

5 Supplementary Appendix C.

Accession	Protein name [species]	Reference
KKR05718.1	hypothetical protein UT34_C0002G0225 [candidate division WS6 bacterium GW2011_GWF2_39_15]	(Brown et al., 2015)
OHA80798.1	hypothetical protein A2675_01620 [Candidatus Yonathbacteria bacterium RIFCSPHIGHO2_01_FULL_51_10]	(Ananthar aman et al., 2016)
OHA83705.1	hypothetical protein A2408_00355 [Candidatus Yonathbacteria bacterium RIFOXYC1_FULL_52_10]	(Ananthar aman et al., 2016)
KKS94036.1	putative steroid-binding protein 3 [Candidatus Collierbacteria bacterium GW2011_GWC2_43_12]	(Brown et al., 2015)
KKU89055.1	hypothetical protein UY18_C0004G0020 [Microgenomates group bacterium GW2011_GWF2_47_9]	(Brown et al., 2015)
OIP98122.1	hypothetical protein AUK40_01915 [Candidatus Wirthbacteria bacterium CG2_30_54_11]	(Probst et al., 2017)
PJA14771.1	hypothetical protein COX64_01640 [Candidatus Dojkabacteria bacterium CG_4_10_14_0_2_um_filter_Dojkabacteria_WS6_41_15]	(Probst et al., 2018)
EKD84736.1* HCE30597.1*	Predicted heme/steroid binding protein [Candidatus Daviesbacteria bacterium]*	(Wrighton et al., 2012)
OGE21801.1	hypothetical protein A2778_04905 [Candidatus Daviesbacteria bacterium RIFCSPHIGHO2_01_FULL_40_24]	(Ananthar aman et al., 2016)
KKQ43725.1	hypothetical protein US62_C0043G0008 [Candidatus Woesebacteria bacterium GW2011_GWA1_37_8]	(Brown et al., 2015)
KKR31495.1	hypothetical protein UT63_C0068G0005 [Candidatus Gottesmanbacteria bacterium GW2011_GWC2_39_8]	(Brown et al., 2015)

6 * EKD84736.1 and HCE30597.1 are identical. Protein name and species are provided
7 for HCE30597.1.

8

9

1 **Table A2.** Crystallography statistics.

Resolution range	19.28-1.9 (1.968-1.9)
Space group	P 63
Unit cell	72.17 72.17 22.81 90 90 120
Total reflections	445959 (41994)
Unique reflections	5544 (519)
Multiplicity	80.4 (80.9)
Completeness (%)	99.87 (100.00)
Mean I/sigma(I)	49.00 (4.27)
Wilson B-factor	25.28
R-pim	0.09944 (0.3724)
CC1/2	0.968 (0.643)
CC*	0.992 (0.885)
Reflections used in refinement	5544 (519)
Reflections used for R-free	280 (24)
R-work	0.2228
R-free	0.2702
Number of non-hydrogen atoms	663
Macromolecules	599
Ligands	43
Solvent	21
Protein residues	76
RMS(bonds)	0.021
RMS(angles)	1.89
Ramachandran favored (%)	97.30
Ramachandran allowed (%)	2.70
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	4.77
Average B-factor	29.60
Macromolecules	29.26
Ligands	31.56
Solvent	35.27

2

3

Table A3. Proteins with the highest structural homology to archaeal Hadesarchaea YNP_N21 cytb_{5M} (KUO41884.1, PDB accession number 6NZX). The protein structures are those from Fig. A6. 3MUS is the Ref_RatCyb5 sequence of Fig. A5. z: z-score confidence in similarity significance, RMSD: root-mean-square deviation of atomic positions across the aligned sequences (Å), LALI: Total number of aligned residues, Nres: Total number of residues, %ID: percentage sequence identity.

PDB ID	z	RMSD	LALI	Nres	%ID	Protein name	Gene	Organism
MAPR: Clade-1								
4X8Y	8.5	1.9	69	112	30	PGRMC1	PGRMC1	<i>Homo sapiens</i>
1J03	8.5	2	72	102	38	Putative steroid binding protein	At2g24940	<i>Arabidopsis thaliana</i>
1T0G	8.4	2.1	73	109	38	Cytochrome b5 domain-containing protein	At2g24940	<i>Arabidopsis thaliana</i>
Clade-2								
4HIL	8	2.2	67	85	27	Cytochrome b5 type B	Cyb5b	<i>Rattus norvegicus</i>
1EUE	8	2.2	67	86	27	Cytochrome b5	Cyb5b	<i>Rattus norvegicus</i>
3MUS	7.9	2.2	67	85	27	Cytochrome b5 type B	Cyb5b	<i>Rattus norvegicus</i>
1ICC	8.1	2.1	67	87	27	Cytochrome b5 mitochondrial	Cyb5b	<i>Rattus norvegicus</i>
2I89	8	2.2	67	85	25	Cytochrome b5 type B	Cyb5b	<i>Rattus norvegicus</i>
1U9M	8	2.2	67	82	30	Cytochrome b5	CYB5A	<i>Bos taurus</i>
1M2M	8.1	2.3	68	82	28	Cytochrome b5	CYB5A	<i>Bos taurus</i>
1LR6	8	2.2	67	82	30	Cytochrome b5	CYB5A	<i>Bos taurus</i>
3X32	8	2.4	68	87	29	Cytochrome b5	CYB5A	<i>Sus scrofa</i>
4HIN	8.3	2.2	67	81	31	Cytochrome b5	CYB5A	<i>Bos taurus</i>
3OZZ	8.3	2.2	67	82	31	Cytochrome b5	CYB5	<i>Bos taurus</i>
1X3X	8	2.2	66	82	30	Cytochrome b5	N/A	<i>Ascaris suum</i>
3LF5	8.1	2.3	69	87	25	Cytochrome b5 reductase 4	CYB5R4	<i>Homo sapiens</i>
1CXY	7.4	2.4	66	81	30	Cytochrome b558	CYB558	<i>Ectothiorhodospira shaposhnikovii</i>
1SOX	8.6	2.1	68	463	29	Sulfite oxidase	SUOX	<i>Gallus gallus</i>

1

2 Supplementary Appendix A. Zip archive containing full PDF images, FASTA sequence
3 alignments, and xml tree files for preliminary studies leading to and including Fig. A1B.

4

5 Supplementary Appendix B. Zip archive containing top BLASTp hits from eukaryotes,
6 using each respective cytb_{5M} protein from Fig. A1B as query sequence in BLASTp
7 against the eukaryotic data base. See the methods section and the README file in the
8 zip archive for details.

9

10 Supplementary Appendix C. Sequences, alignments and phylogenetic trees for the
11 results summarised in Fig. A4, Fig. 2 and Supplementary Appendix D.

12

13 Supplementary Appendix D. Summary table of the phylogenetic analysis of Clade 1
14 topologies for Fig. A4 using TBE branch support values analysis.

15

16 Supplementary Appendix E. Zip archive containing full PDF images, FASTA sequence
17 alignments, and xml tree files for Fig. A9.

18

19 Supplementary Appendix F. 533 sequences of Cyp51A from Fig. 6.

20

21 Supplementary Appendix G. Protein Data Bank validation reports for the structure of
22 Hadesarchaea YNP_N21 cytb_{5M} (KUO41884.1), with PDB accession number 6NZX.