Title:

Cell-type-specific expression of a *DCYTB* ortholog enables the choanoflagellate *Salpingoeca rosetta* to utilize ferric colloids

Authors:

Fredrick Leon[1], Jesus M. Espinoza-Esparza[1], Vicki Deng[1,3], Maxwell C. Coyle[2,4], Sarah Espinoza[2], David S. Booth[1*]

Affiliations:

[1]Chan Zuckerberg Biohub &
Department of Biochemistry and Biophysics
University of California, San Francisco School of Medicine
San Francisco, CA 94143

[2]Howard Hughes Medical Institute &
Department of Molecular and Cell Biology
University of California, Berkeley
Berkeley, CA 94720

[3]Current Address: Department of Molecular Biosciences
University of Texas, Austin
Austin, TX 78712

[4]Current Address: Department of Molecular and Cellular Biology
Harvard University
Cambridge, MA 02138

[*]Corresponding author: David.Booth@ucsf.edu

Abstract:

The dynamic life histories of microeukaryotes present a challenge for determining their role in the nutrient cycles that support marine ecosystems. For instance, cellular differentiation is one life history strategy that enables a single celled microeukaryotic species to produce morphologically and physiologically distinct types of cells with different ecological functions. Here we report that cell differentiation in the marine choanoflagellate *Salpingoeca rosetta* endows one of its cell types with the ability to utilize insoluble ferric colloids for improved growth through the expression of a cytochrome b561 iron reductase (*cytb561a*). This gene is an ortholog of the mammalian duodenal cytochrome b561 (*DCYTB*) that reduces ferric cations prior to their uptake in gut epithelia and is part of an iron utilization toolkit that choanoflagellates and their closest living relatives, the animals, inherited from a last common eukaryotic ancestor. In a database of oceanic metagenomes, the abundance of *cytb561a* transcripts from choanoflagellates positively correlates with upwellings, which are a major source of ferric colloids in marine environments. As this predominant form of iron is largely inaccessible to cell-walled microbes, choanoflagellates and other phagotrophic eukaryotes may serve critical ecological roles by first acquiring ferric colloids through phagocytosis and then cycling this essential nutrient through iron utilization pathways. These findings provide insight into the ecological roles choanoflagellates perform and may inform reconstructions of early animal evolution where functionally distinct cell types became an integrated whole at the origin of animal multicellularity.

Text:

Thecates display a different transcriptome profile than other cell types.

In the decade since the first genome and transcriptomes[1] of *S. rosetta* were sequenced, the establishment of a monoxenic strain of *S. rosetta* that feeds on the bacterium *Echinicola pacifica*[2–4] has led to reproducible methods to culture four different cell types from *S. rosetta*[5]: thecates, rosettes, slow swimmers, and fast swimmers (Fig. 1A). Under nutrient replete conditions, sessile thecates differentiate into motile slow swimmers[5,2,3] that detect specific bacterial[6–8] and algal[9] cues to develop into multicellular colonies called rosettes[10]. In response to starvation, rosettes and slow swimmers differentiate into fast swimmers. Stable culturing methods for these cell types motivated us to refine the transcriptomes of each cell type in a common bacterial background. To further improve transcriptomes, we also developed a method to preferentially lyse *S. rosetta* while discarding their feeder bacteria. This preferential lysis allowed us to harvest mRNA from *S. rosetta* with a single round of poly-A selection (Fig. S1) rather than previous methods that relied on multiple rounds of selection to deplete contaminating RNAs from bacteria[1,10]. A principal component analysis of these improved transcriptomes emphasized a stark difference between the thecate cell type compared to slow swimmers, rosettes, and fast swimmers (Fig. 1B) with the first principal component accounting for 86% of the variance and clearly separating thecates from all other cell types. While these results are consistent with previous gene expression profiles that found thecates separated from other types of cells[1,12], the replicate measures in our data do not show a clear difference between slow swimmers, rosettes, and fast swimmers, suggesting that these three cell types are phenotypically plastic states.

The contrast between thecates and other cell types prompted us to investigate what unique functions thecates may perform. By comparing thecates to slow swimmers, we first identified a set of genes that reproducibly ($q < 0.01$, note that q is a P-value adjusted for multiple comparisons) displayed a greater than two-fold change in gene expression in thecates. We analyzed this set of genes for common gene ontologies (GO)[13–16] that indicated functional categories enriched in thecates. Although the incomplete annotation of genomes from non-model organisms can exacerbate caveats from classifying particular genes into discrete functional categories[17], the genes upregulated in thecates were clearly enriched for functional modules (Fig. 1C) implicated in cell signaling (e.g. nucleotide binding, protein kinase, protein phosphatase, calcium binding, metal ion transporter, guanylate cyclase and G-protein coupled receptor), gene regulation (e.g. transcription factor and ribosome component), and metabolism

87 (e.g. nucleotide binding, oxidoreductase, glycosyl hydrolase, metal ion transporter, carbohydrate
88 binding, sulfuric ester hydrolase, and ascorbic acid binding). From GO molecular function
89 categories, we then generated a list of associated genes (GO Slim)[18]. Within this list, we
90 focused on metabolic genes, anticipating that genes from those categories would yield
91 candidate(s) to experimentally test specific functions that thecates perform. This analysis
92 highlighted a gene (PTSG_09715) with 479-fold higher expression in thecates compared to slow
93 swimmers ($q < 10^{-40}$) that was annotated as an oxidoreductase (GO:0016491) comprised of a
94 single cytochrome b561 iron reductase domain (Pfam 03188). Given the role that
95 oxidoreductases perform in diverse eukaryotes to regulate iron acquisition[19–22], the strong
96 induction of PTSG_09715 (henceforth called *cytb561a*) led us to hypothesize that thecates may
97 have a different capacity to acquire iron from the environment compared to other cell types.
98
99 ***cytb561a* expression improves thecate growth with ferric colloids.**
100 Of the 2,820 genes that reproducibly displayed a two-fold or greater change in mRNA
101 expression in thecates compared to slow swimmers, the change in *cytb561a* expression is in
102 the ninety-eighth percentile (Fig. 2A). This large change in *cytb561a* expression may have been
103 coupled to the thecate cell type program or may have been an independent response to
104 external nutrients that differed between the media for culturing slow swimmers and thecates for
105 the transcriptomes (see methods). To distinguish between these possibilities, we measured
106 *cytb561a* mRNA expression in slow swimmers and thecates after standardizing a culturing
107 regime (Fig. 2B and Table S1). In this protocol, slow swimmers and thecates were grown in the
108 same media formulation and then depleted of iron. Afterwards, we reintroduced iron either as
109 soluble ferric EDTA (Fe^{3+}•EDTA) or as insoluble ferric colloids (Fe^{3+}(oxyhydr)oxides)[23] because
110 models based on ecological measurements predict that 40% of the world's surface ocean
111 waters rely on colloidal sources of iron and 18% on soluble iron chelates[24]. Regardless of iron
112 having been provided as ferric EDTA versus ferric colloids (two-way ANOVA, $P > 0.1$), *cytb561a*
113 mRNA expression increased in thecates compared to slow swimmers by 89-fold with ferric
114 EDTA and 66.2-fold with ferric colloids (two-way ANOVA, $P < 0.005$) (Fig. 2C and S2 A-C). This
115 result supports the conclusion that thecates are a stable cell type that expresses *cytb561a* as
116 part of its differentiation regulon.
117 To determine the functional consequence of increased *cytb561a* expression in thecates,
118 we compared the growth of thecates and slow swimmers in the presence of 100 μM ferric EDTA
119 or ferric colloids (Fig. 2D). Thecates grew to a two-fold higher cell density in the presence of
120 ferric colloids (two-way ANOVA, $P < 10^{-4}$), yet slow swimmers showed no change (two-way

121 ANOVA, $P > 0.1$). This effect was observed at lower concentrations of iron (50 μM), but higher
122 concentrations of iron (≥200 μM), favored higher cell proliferation in ferric EDTA (Fig. S3A).
123 Thecates also grew better than slow swimmers with other ferric chelates, such as ferric EGTA
124 and ferric pyoverdine (Fig. S4B). Importantly, the differences in *S. rosetta* growth are very
125 unlikely due to indirect effects from the prey bacteria, for *E. pacifica* grew similarly in all iron
126 conditions, likely due to the small concentration of iron (3.43 ± 1.01 μM) that dissolved from the
127 ferric colloids (Fig. S3 C-E).

128 The increased proliferation of thecates in the presence of ferric colloids did depend on
129 the expression of *cytb561a*. With an improved genome editing pipeline (see methods and Fig.
130 S4 A-B), we altered *cytb561a* expression by introducing a premature termination sequence[25] at
131 position 151 of *cytb561a* to incorporate missense mutations and a polyadenylation signal, which
132 diminished mRNA expression by 10-fold compared to wild-type (Fig. S2D, one-way ANOVA, $P =$
133 0.004). This mutant allele, $cytb561a^{PTS}$, eliminated the difference in cell proliferation between
134 slow swimmers and thecates (Fig. 2D, $P > 0.1$). By comparing the growth dynamics of
135 $cytb561a^{PTS}$ to wild-type (Fig. S5), we also found that the enhancement of thecate growth with
136 ferric colloids was primarily due to a two-fold larger carrying capacity and a shorter doubling
137 time (17.9 h vs 13.2 h, two-way ANOVA, $P = 0.035$). Altogether, these results suggest that the
138 increased expression of *cytb561a* in thecates confers a cell-type-specific function to utilize
139 insoluble ferric colloids for faster and more extensive growth.

140

141 **Choanoflagellates and animals share a core iron acquisition toolkit comprised of genes**
142 **broadly conserved across eukaryotes.**

143 Although *cytb561a* expression was necessary for the increased proliferation of thecates
144 when grown with ferric colloids, we were uncertain how this iron reductase factored into the
145 pathway(s) for iron acquisition in *S. rosetta*, as iron transport pathways have not previously
146 been examined in choanoflagellates. Therefore, we compiled an inventory of iron import and
147 export proteins that have been characterized in model eukaryotes (e.g. humans[22], *S.*
148 *cerevisiae*[26], *P. tricornutum*[19], *C. reinhardtii*[21], Fig. S6) to survey the distribution of those proteins
149 across Holozoa (the taxonomic group that includes Animalia, Choanoflagellata, Filasterea, and
150 Teretosporea) and a curated set of eukaryotes[27]. Overall, we found that nearly all of the proteins
151 for iron import, export, reduction, and oxidation are widely distributed among eukaryotes,
152 suggesting a deep ancestry of iron acquisition pathways from the last common eukaryotic
153 ancestor (Fig. 3A).

154 Our searches also revealed notable evolutionary patterns that correspond to the unique
155 biology of certain eukaryotic lineages. For example, we found significant losses of iron toolkit
156 genes in some lineages: Apicomplexa, a taxon that includes the intracellular parasites
157 *Cryptosporidium parvum*, *Toxoplasma gondii*, and *Plasmodium falciparum* had a severely
158 reduced set of iron acquisition genes, consistent with the origin of *de novo* iron acquisition
159 pathways for siphoning iron from their hosts[28]. Moreover, we observed that some iron utilization
160 genes are more taxonomically restricted than others: Genes first characterized in ascomycete
161 yeasts were either primarily found in Fungi and Archaeplastida (e.g. iron transporter[29]) or only
162 present in a few eukaryotes (e.g. iron permease[30]). Similarly, transferrin and phytotransferrin
163 were respectively restricted to Holozoa and Archaeplastida, respectively, as previously shown[31].
164 Notably, the presence of transferrin in Animalia and Teretosporea indicated that transferrin was
165 present in the last common ancestor of Holozoa, yet the transferrin receptor that facilitates the
166 cellular internalization of iron-bound transferrin appears to have originated in Bilateria,
167 highlighting the evolution of a mechanism in animals to share iron between cells. Finally, we
168 also observed that the presence of genes encoding vacuolar iron transporters and ferritin, which
169 are both involved in iron storage, was slightly anticorrelated (r = -0.380, $P < 10^{-8}$). Overall, this
170 catalog of iron utilization proteins does indicate that eukaryotes have variably retained,
171 amplified, or lost genes from an ancestral toolkit for the acquisition, storage, and export of iron.
172 Choanoflagellates and animals appeared to share a core pathway for acquiring iron from
173 their external environment. Iron acquisition in animals has most extensively been studied in
174 vertebrate gut epithelia[32]. The apical side of those epithelial cells absorbs dietary iron by
175 reducing ferric cations to ferrous cations through the activity of duodenal cytochrome B561
176 (DCYTB) before divalent metal ion transporter (DMT1) brings ferrous cations inside of cells. On
177 the basal end of gut epithelia, ferroportin (FPN) secretes ferrous cations into the bloodstream,
178 where they are oxidized to ferric cations near the membrane by hephaestin (HEPH). In our
179 catalog of iron utilization genes, all choanoflagellates possessed homologs of *DCYTB*, *DMT1*,
180 and *FPN*, but *HEPH* (Fig. 3A, row 8–ferroxidase type1) was only present in *S. rosetta* and 9
181 other species (of 22 surveyed[33])–perhaps due to the lack of complete genomes for many of
182 those species. Differences between choanoflagellates and animals were evident in iron storage
183 mechanisms, as choanoflagellates mostly possessed vacuolar iron transport proteins, which
184 have been studied in *S. cerevisiae*[34,35]; whereas, animals broadly conserved ferritin. As we
185 searched for homologs of iron acquisition genes in *S. rosetta*, we found single copies of *dmt1*,
186 *fpn*, and *heph* and identified two vacuolar iron transporters (*vit1* and *vit2*). Three cytochrome
187 b561 genes were present, yet among all the conserved iron utilization genes, only *cytb561a*

188 exhibited strong differential expression between cell types (Fig. 3B), emphasizing its
189 cell-type-specific role in thecates.
190 Because the apico-basal localization of iron acquisition genes facilitates the
191 unidirectional import of iron in animal epithelia[32], we tested if iron transport in *S. rosetta* shared a
192 similar architecture (Fig. 3C). To do so, we used genome editing at endogenous loci to introduce
193 an epitope tag (ALFA-tag[36]) in all the iron utilization genes. Only Cytb561a, Fpn, and Heph were
194 visible in western blots of thecates (Fig. S7A), so we focused on those proteins to visualize by
195 immunofluorescence. In thecates, Cytb561a localized intracellularly toward the apical half of the
196 cell but not in any pattern resembling organelles[37], but Cytb561a was undetectable in slow
197 swimmers (Fig. 3C, S7B), consistent with its mRNA expression (Fig. 2C). In both slow
198 swimmers and thecates, Fpn and Heph localized near the apical membrane inside of the collar.
199 The similar co-localization of Fpn and Heph would be consistent with their coordinated function
200 as an efflux pump and oxidase as described in animals[32]. Importantly, the basal localization of
201 these proteins appears to be a feature of animal epithelia that is not shared with
202 choanoflagellates and may have evolved alongside the mechanisms to establish and maintain
203 cell polarity[38].
204 Based on the functions of homologous genes and the localization of Cytb561a, Fpn, and
205 Heph, we propose a model for iron acquisition in *S. rosetta* thecates that may extend to other
206 choanoflagellates (Fig. 3D). After internalizing ferric colloids, Cytb561a reduces ferric cations in
207 endosomes where we surmise that Dmt1 transports ferrous cations into cells–although Dmt1
208 was undetectable by immunohistochemistry (Fig. S7). As slow swimmers do not express
209 Cytb561a, *S. rosetta* may acquire iron directly from bacterial prey digested in food vacuoles or
210 rely on Cytb561b or Cytb561c to reduce ferric cations. In both slow swimmers and thecates,
211 Fpn exports ferrous cations at the apical end where Heph can oxidize it.
212
213 **Cytb561 paralogs display distinct biochemical properties and global distributions.**
214 The variation in the number of cytochrome b561 paralogs across eukaryotes led us to
215 question the function and evolution of the three cytochrome b561 paralogs that we identified in
216 *S. rosetta*. While the expression of *cytb561a* in thecates was clearly necessary for improved
217 growth in the presence of colloidal iron, slow swimmers still grew in those conditions as did
218 *cytb561a*^{PTS} thecates (Fig. 2D). These observations and the steady mRNA expression of
219 *cytb561b* and *cytb561c* across cell types would be consistent with these paralogs sustaining a
220 baseline ability to acquire iron, even if we could not detect their protein products by
221 immunohistochemistry (Fig. S7). The potential for functional redundancy among these paralogs

222 also led us to consider if choanoflagellates and other eukaryotes display lineage-specific
223 expansions of this gene family that may have resulted in differing capacities to acquire iron.
224 To better compare the cytochrome b561 paralogs in *S. rosetta*, we first built a
225 phylogenetic tree[39] from an alignment[40] of Cytb561 proteins from Amorphea (Fig. 4A), which is
226 the eukaryotic supergroup that includes Holozoa and Amoebae[27,41]. The unrooted tree of
227 cytochrome b561 proteins robustly supported (ultrafast bootstrap[42] > 88) three distinct groups.
228 Groups B and C contained protein sequences from each major clade in Amorphea and its sister
229 group CrUMS (Collodictyonida, Rigifilida and Mantamonadida). Strikingly, sequences from
230 Holozoa dominated Group A, as a single protein from Amoebozoa was the only sequence from
231 outside of Holozoa that fell into this group. Notably, Group A not only contained mammalian
232 DCYTB proteins but also Cytb561a from *S. rosetta*, supporting the annotation of this gene as an
233 DCYTB ortholog. The Cytb561b and Cytb561c proteins from *S. rosetta* fell into Groups B and C,
234 respectively.
235 Because each of the *S. rosetta* paralogs was part of a different group of cytochrome
236 b561 proteins, we examined[43] their predicted structures[44,45] to evaluate if the deep evolutionary
237 divergence was reflected in their biochemical characteristics (Fig. 4B). Electrostatic maps of
238 their protein surfaces revealed differences in the iron binding site (Fig. 4B, arrows). Cytb561a
239 possessed a mix of positive and negative charges similar to the crystal structure of human
240 DCYTB that shows these charged residues bind iron and coordinated ligands[46]. The anionic
241 surface of the iron binding site in Cytb561c could also plausibly bind iron cations, but the
242 positively-charged pocket of Cytb561b may repel iron cations. With each paralog from *S. rosetta*
243 possessing distinct biochemical properties near the iron binding site and belonging to a different
244 Cytochrome b561 subgroup, we hypothesize that the improvement of thecate growth with ferric
245 colloids may not only be due to the mass action from more Cytb561 proteins by also the distinct
246 biochemical activity of Cytb561a.
247 If the biochemical activity and cell-type-specific expression of Cytb561 paralogs enable
248 choanoflagellates to acquire iron from different environmental sources, then we expect that the
249 expression of each paralog may be enriched in different ecological settings. To test this
250 hypothesis, we searched for each paralog in an open repository (Ocean Gene Atlas[47,48]) of
251 metagenomes and metatranscriptomes from marine environments. When we visualized the
252 distribution of each paralog, we found that they varied in both location and transcript abundance
253 but were similarly present in metagenomic reads (Fig. 4C) The most abundant and
254 cosmopolitan paralog was *cytb561c*. Although present at most of the sites, the transcript
255 abundance of *cytb561b* was ~10-fold lower than *cytb561c*. The abundance of *cytb561a* had a

256 similarly low abundance as *cytb561b,* yet its distinct presence near coastlines and the equator
257 resembled the locations of marine upwellings[49] (Fig. S8A).
258 Importantly, upwellings are a major source of remineralized iron colloids rising from
259 ocean depths[50,51], and upwelling velocities may serve as a better proxy for the presence of
260 colloidal iron than predictions of soluble iron concentrations[52]. Therefore, we correlated the
261 abundance of the genomic and transcript reads for each paralog to average vertical velocities at
262 a depth of 100 meters for each sample location (Fig. 4C, S8B). Of all the comparisons, the only
263 robust correlation was between *cytb561a* transcript abundance and vertical velocity (r = 0.64, P
264 < 10^{-4}), as the transcript abundances of the other paralogs exhibited only weak positive
265 correlations that were probably random (P > 0.3). The positive correlation between *cytb561a*
266 transcripts and upwelling velocities indicate ecological contexts where choanoflagellates may be
267 acquiring insoluble iron colloids for their growth. We cannot, however, determine if the
268 choanoflagellates in those regions are thecates expressing *cytb561a* as in *S. rosetta*, or
269 choanoflagellates that evolved a different regulatory mechanism that uncoupled *cytb561a*
270 expression from thecate development. Nonetheless, these observations connect *cytb561a*
271 expression to an environmental feature and places choanoflagellate biology in a global
272 ecological framework.
273
274 Iron is an essential micronutrient affecting the growth and productivity of
275 phytoplankton[53]. Field measurements and ocean scale models have found that iron particulates
276 are the major source of iron for 40% of the world's surface ocean waters[24]. In addition to
277 photolytic reduction[54,55] and bacterial siderophores[56,57], the ingestion of iron particles by
278 phagotrophic protists is a major mechanism to solubilize iron and can alleviate iron limitation for
279 cell-walled organisms that rely on obtaining iron in soluble forms[58,55,59]. As efficient
280 phagotrophs[60–63], choanoflagellates can ingest iron particulates, yet the need to express
281 *cytb561a* for *S. rosetta* thecates to utilize ferric colloids shows that gene regulation in the
282 context of life history transitions is a critical factor in determining the ability of a microeukaryote
283 to access essential resources from their environment. The complement of iron acquisition genes
284 may further enable or constrain phagotrophs to break down iron, for iron acquisition genes
285 display a variable distribution across eukaryotic lineages (Fig. 3A). Those genes may have also
286 evolved protein biochemical activities for different sources of iron, as shown by the distinctive
287 biochemical properties of cytochrome b561 paralogs (Fig. 4B). Thus, an integrative account of
288 dynamic life histories[64,65] and protein evolution[66] are necessary to determine the ecological roles
289 that microeukaryotes play in acquiring and distributing limiting nutrients in their environment[64,65].

Moreover, the synthesis of cytological, genomic, and ecological data may help to interpret how the rise in oxygen that resulted in the depletion of soluble ferrous cations from the oceans influenced eukaryotic evolution[66–68], including that emergence of animals[69].

The cell-type-specific expression of *cytb561a* not only illustrates how cellular differentiation impacts the environmental function of a choanoflagellate but also informs how cellular differentiation evolved as a defining feature of animal multicellularity[70–75]. Choanoflagellate cell types have been primarily defined by their morphology and motility[5,62]. In addition to those cell biological characteristics, we show here that thecates have a distinct function to utilize ferric colloids. The coupling of this function to cell differentiation stands in contrast to the regulation of iron utilization in other microeukaryotes that transiently regulate iron utilization genes in response to external iron concentrations[35,76,77]. The functional difference between slow swimmers and thecates more closely resembles the stable differentiation of cell types during animal development[78–80]. Taken together, these observations support models for the origin of animal cell differentiation in which the last common holozoan ancestor differentiated into unique cell types at each life history stage[81,70,72,74,82]. Evolution along the animal stem lineage coupled cell differentiation to multicellular development, resulting in multicellular bodies composed of functionally distinct cells[83]. Importantly, mechanisms to distribute resources from cells that acquire nutrients from the external environment (e.g. epithelial cells[84]) to cells that reside internally helped to transform the functions of various cell types into a common organismal physiology[85]. As the evolution of transferrin and its receptor illuminates, the multicellular physiology of animals has likely been cobbled together through the co-option of mechanisms for nutrient acquisition and cellular trafficking.

312 **Materials and Methods:**

313

314 ***Culturing S. rosetta***

315 To grow cell types of *S. rosetta* for the generation of transcriptomes, slow swimmer and rosettes
316 were grown in 5% Sea Water Complete (SWC) Media[2,86] (Table S1), seeded at 10^4 cells/mL and
317 grown for 48 hours at 22°C. Rosettes were induced with outer membrane vesicles (OMVs)
318 harvested from the bacterium *Algoriphagus machipongonensis*[7]. Fast swimmers were prepared
319 the same way as slow swimmers, but were grown for 72 hours at 22°C and transitioned to 30°C
320 for 2 hours and 45 minutes. Thecates were grown from the strain HD1, a strain which is
321 constitutively in the thecate cell type, in 10% Cereal Grass Media 3 (CGM3)[2,86] (Table S1),
322 seeded at 10^4 cells/mL and grown for 48 hours at 22°C.

323

324 For iron acquisition experiments and standard culturing of *S. rosetta*, all cell types were grown in
325 25% Red Algae (RA) Media[87] (Table S1) with the feeder bacteria *Echinicola pacifica*. Cultures
326 were grown at 18°C for maintaining cultures and shifted to 27°C one day prior to conducting iron
327 acquisition experiments.

328

329 ***RNA extraction***

330 In a previous paper[86], we developed a lysis buffer to extract RNA from *S. rosetta* that has
331 subsequently been used in other work for RNA[88–90] and protein extraction[91]. Here we show that
332 this buffer was developed to preferentially lyse *S. rosetta* in cultures that have an abundance of
333 the prey bacterium *E. pacifica*. To preferentially lyse *S. rosetta*, we reasoned that sterol-based
334 detergents, such as digitonin, would more effectively disrupt membranes with sterols, like those
335 of *S. rosetta* and other eukaryotes[92] rather than bacterial membranes that largely lack the
336 sterols[93]. Additionally, the lysis buffer contains RNase inhibitors (RNaseIN and heparin) and
337 translation inhibitors to decrease the degradation of mRNA. We evaluated the efficacy of this
338 buffer by tracking ribosomal RNA from *S. rosetta* and *E. pacifica* in lysed samples that were
339 centrifuged to separate the supernatant and pellet, in which we found that *S. rosetta* ribosomal
340 RNAs were enriched in the supernatant and bacterial ribosomal RNAs were enriched in the
341 pellet (Fig. S1). The optimization of this procedure resulted in the following method to obtain
342 RNA samples from each cell type of *S. rosetta*:

343

344 After counting the cell concentration in a culture of *S. rosetta* feeding on *E. pacifica*, a volume
345 yielding 10^7 cells was centrifuged at 2600xg for 5 min at 4°C. The pellet from this sample was

346 resuspended in 100 µl of preferential lysis buffer (20 mM Tris-HCl, pH 8.0; 150 mM KCl; 5 mM
347 $MgCl_2$; 250 mM Sucrose; 1 mM DTT; 10 mM Digitonin; 1 mg/ml Sodium Heparin; 1 mM Pefabloc
348 SC; 100 µg/ml Cycloheximide; 0.5 U/µl Turbo DNase [Thermo Fisher Scientific, Cat. No.
349 AM2239]; 1 U/µl SUPERaseIN [Thermo Fisher Scientific, Cat. No. AM2696]) and incubated on
350 ice for 10 min. Cells were triturated through a 30G needle ten times. Afterwards, the insoluble
351 debris was pelleted by centrifugation at 6,000 x g for 10 min at 4°C. The supernatant was
352 removed and adjusted to a volume of 100 µl with RNase-free water. Total RNA was purified on a
353 silica membrane with the RNA clean-up protocol from the RNAeasy kit (Qiagen, Cat. No.
354 74104).
355
356 ***RNA library preparation and sequencing***
357 In preparation for sequencing, the quality of total RNA from *S. rosetta* samples was assessed by
358 Agilent Bioanalyzer 2100 (Fig. S1C) with the Agilent RNA 6000 Nano Kit (Agilent, Cat. No.
359 5067-1511). Then total RNA (500 ng per sample) was purified by one round of polyA mRNA
360 selection with oligo-dT magnetic beads (NEB, Cat. No. S1550S), converted to cDNA using
361 KAPA mRNA HyperPrep kit (KAPA biosystems, Cat. No. KK8580), and indexed-adapter ligated
362 with KAPA single-indexed adapter kit (KAPA biosystems, Cat. No. KK8701). Final library quality
363 was determined by Bioanalyzer 2100 using the Agilent High Sensitivity DNA Kit (Agilent, Cat.
364 No. 5067-4626) and pooled together after normalizing samples based on their quantity from
365 qPCR. Sequencing was carried out by the QB3-Berkeley Genomics core (QB3 Genomics, UC
366 Berkeley, Berkeley, CA, RRID:SCR_022170), on the Illumina HiSeq 4000. All samples were
367 pooled and run on a single lane with 12.4 to 61.3 million reads per sample. Reads were
368 demultiplexed, checked for quality by fastqc (Babraham Bioinformatics).
369
370 ***RNA-seq quantification and gene enrichment analysis***
371 RNA-seq reads from each sample were mapped to a reference *S. rosetta* transcriptome
372 (https://ftp.ensemblgenomes.ebi.ac.uk/pub/protists/release-58/fasta/protists_choanoflagellida1_
373 collection/salpingoeca_rosetta_gca_000188695/cdna/) using kallisto[94]. Transcript abundance
374 and differential expression were quantified with sleuth[95]. Gene enrichment for thecates was
375 quantified by selecting genes with $q < 0.01$ and $\log_2$(fold-change + 0.5) > 1. The resulting genes
376 were assigned GO terms using DAVID Bioinformatics[15,16] where *S. rosetta* gene IDs could be
377 identified by “ENSEMBLE_GENE_ID”. For ease of interpretation, assigned GO terms were
378 collapsed down using REVIGO[96] and further assigned to functional categories in *S. rosetta*
379 (TaxID: 946362) by GO Slim[18].

380 ***S. rosetta growth with different sources of iron***

381 **Media Preparation.** To test the ability of *S. rosetta* cell types to differentially utilize iron sources,
382 we developed a method to culture *S. rosetta* under iron limitation. We created an iron depleted
383 media formulation named 4% Peptone Glycerol (PG) Media (Table S1). We added iron to this
384 media from two different sources: Ferric EDTA was prepared by dissolving $FeCl_3$ and EDTA to a
385 final concentration of 1.46 μM in artificial seawater (ASW [Ricca Chemical Company, Cat. No.
386 R8363000-20F]) and sterile filtering through a 0.22 μm polyethersulfone (PES) filter. Ferric
387 colloids were prepared by adding a small volume of 70% (v/v) ethanol to $FeCl_3$(*s*) to sterilize the
388 solid; after evaporating the liquid in a biological safety cabinet, 1.46 μM $FeCl_3$ was prepared in
389 ASW. The solution was heated to 50°C for 10 minutes to precipitate ferric (oxyhydr)oxides, or
390 ferric colloids[23].

391 **Determining iron concentration in media.** To determine the background concentration of
392 soluble iron in our newly formulated media and the amount of labile iron liberated from the ferric
393 colloids, we used a modified protocol[58] to measured soluble iron concentrations in 4% PG and
394 4% PG with 100 μM ferric colloids. To measure these iron concentrations, 50 mL of 4% PG
395 media was incubated at 27°C for 6 hours with and without 100 μM ferric colloids. The media
396 was then centrifuged at 4600xg for 15 minutes at room temperature. The top 45 mL of
397 supernatant was taken for the following steps. To capture and concentrate the soluble iron in the
398 supernatant, 125 μM 8-Hydroxyquinoline (oxine) was added to chelate labile iron, and 5 mL of
399 1M MES, pH 6 was added to aid in oxine's solubility in the media and its binding to iron, which is
400 facilitated by lower pH. These solutions were incubated for 24 hours protected from light, and
401 then filtered through a Sep-Pak C18 column (Waters, Cat. No. WAT036800) to bind oxine-iron
402 complexes. The column was then washed with 5 mL water and eluted with 1.4 mL methanol. To
403 measure iron concentrations in the eluate, 50 μL of eluate was measured with an colorimetric
404 iron detection kit (Sigma-Aldrich, Cat. No. MAK025-1KT) following the provided protocol. Iron
405 concentration was determined by measuring absorbance at 595 nm on a plate reader.

406 **Iron-limited *E. pacifica* pellets** were prepared by growing *E. pacifica* in 4% PG for 72 hours at
407 225 rpm and 30°C, then pelleted, aliquoted, and frozen into 10 mg pellets, and stored at -80°C.
408 *E. pacifica* pellets were resuspended in 1mL of media to achieve a 10mg/mL liquid bacteria
409 stock to add to cultures.

410 **Culturing with different iron sources.** To set up the iron utilization assay, cultures grown in
411 25% RA were passaged at a 1:60 dilution into 25% RA and grown for 72 hours at 18°C. This
412 culture was then passaged at a 1:60 dilution into 25% RA and grown for 24 hours at 27°C.
413 Finally, this culture was passaged at a 1:30 dilution into 4% PG (which had no iron added) and

414 grown at 27°C for 24 hours. Cells from this final passage were twice washed by first centrifuging
415 the culture at 2600xg for 4 minutes and then resuspending the pellet in 30 mL sterile ASW. After
416 the final wash, the pellet was resuspended in 100 µL ASW and counted. These washed cells
417 were used to seed cultures at 10^4 cells/mL in 4% PG with the additions of 100 µM iron (ferric
418 EDTA or ferric colloids) and 50 µg/mL of iron-limited *E. pacifica* pellets (see above). The
419 resulting culture was grown at 27°C for 48 hours and harvested for each experiment.
420
421
422 ***qPCR***
423 **Primer design.** Primers were designed against *cytb561a,* the experimental gene of interest and
424 *cofilin,* a highly expressed eukaryotic specific housekeeping gene with a consistent expression
425 between *S. rosetta* cell types, as shown by the RNAseq analysis. Primers for quantitative
426 polymerase chain reaction (qPCR) were designed by selecting 19-20 bp oligos spanning
427 exon-exon junctions from cDNA sequences, (downloaded from:
428 https://protists.ensembl.org/Salpingoeca_rosetta_gca_000188695/Info/Index). Primer
429 sequences as follows: *cytb561a* forward primer 5'-CATGGAAGAAGCGCATGGTG-3', reverse
430 primer 5'-GGGTCCGCAAGATCATTGAG-3', *cofilin* forward primer
431 5'-CAAGCTCCCCACGGACAAG-3', reverse primer 5'-GGGTCCATGCGAAGAAGAC-3'. In the
432 case of *cytb561a,* primers were selected downstream of the premature termination sequence
433 edited in by CRISPR/Cas9.
434 **Primer validation.** Target amplicons were amplified by polymerase chain reaction (PCR) from
435 the cDNA of slow swimmer and thecate cultures (Fig. S2A) with the Luna Universal qPCR
436 Master Mix (NEB, Cat, No. M3003). The amplicons were run on a 1% (w/v) agarose gel in
437 Tris-Borate-EDTA (TBE) buffer to verify that the primer sets amplified only one amplicon. The
438 amplification efficiency of primer sets was characterized across a serial dilution of ssDNA
439 standards for each target (Fig. S2B-C). ssDNA standards were generated from PCR products
440 amplified with a forward primer that was 5' phosphorylated to promote Lambda exonuclease
441 digestion of that strand (NEB, Cat. No. M0262S) and a reverse primer with phosphorothioate
442 bonds between the first four 5' nucleotides to block digestion. Those PCR products were then
443 digested with Lambda exonuclease following NEB's protocol. ssDNA concentrations were
444 determined by Qubit ssDNA Assay Kit **(**Thermo Fisher, Cat. No. Q10212) by creating a standard
445 curve with the provided standards and then reading fluorescence (Ex = 500 nm / Em=540 nm)
446 on a plate reader (Molecular Devices, SpectraMax iD5). The ssDNA standards were serially

447 diluted from 10^5 to 10^0 copies/µL in a solution with 10 ng/µL of *S. rosetta* RNA (Total RNA after
448 removing mRNAs) to account for any matrix effects.
449 **cDNA preparation.** Samples from slow swimmer or thecate cultures were pelleted and lysed
450 using the preferential lysis protocol, and total RNA was purified by RNeasy MinElute Cleanup
451 Kit (Qiagen, Cat. No. 74204). cDNA was prepared by SuperScript IV Reverse Transcriptase
452 (Thermo FIsher, Cat. No. 18090010) following the provided protocol using oligo $d(T)_{20}$ for
453 synthesis. However, the incubation temperature for cDNA synthesis was increased to 60°C to
454 account for the high GC content of the *S. rosetta* genome.
455 **qPCR.** 3 µl of cDNA samples or ssDNA standards were added to a 20µl reaction with Luna
456 Universal qPCR Master Mix (NEB, Cat, No. M3003) according to the provided protocol. The
457 samples were amplified in a single run on the QuantStudio 3 Real-Time PCR System (Thermo
458 Fisher, Cat. No. A28567).
459
460 ***Genome editing***
461 **Culturing cells for transfection.** Mutant strains were generated with a modified protocol from a
462 previously published method[25]. Cultures of *S. rosetta* slow swimmers were maintained in 15%
463 RA/2% PYG (called 15/2, Table S1) at 22°C, and 48 hours prior to transfection, cultures were
464 seeded at 10^4 cells/mL in 80 mL of 15/2 Media in 300 cm^2 vented tissue culture flasks (VWR,
465 Cat. No. 10062-884) and grown at 22°C.
466 **Cas9 RNP preparation.** On the day of transfection, Cas9 bound to a guide RNA (Cas9 RNP)
467 was prepared by combining 2 µL of 100 µM guide RNA (gRNA) with 2 µL of 20 µM EnGen
468 SpyCas9 NLS (NEB, Cat. No. M0646T), and incubated at room temperature for 2-4 hours.
469 (Note: Our gRNA was ordered as synthetic RNA oligonucleotides from Integrated DNA
470 Technologies that came as an Alt-R CRISPR-Cas9 crRNA and an Alt-R CRISPR-Cas9
471 tracrRNA. These two oligonucleotides were annealed together to form gRNA).
472 **Transfection.** During the Cas9 RNP incubation, cells were harvested for transfection by
473 centrifuging the cultured cells at 2400xg for 3 minutes at 4°C. The supernatant was discarded,
474 and cells were resuspended in 50 mL of Cell Wash Buffer (420 mM NaCl; 50 mM $MgCl_2$; 30 mM
475 Na_2SO_4; 10 mM KCl; titrated to pH 8.0 with ~2.4 mM $NaHCO_3$). The cells were centrifuged again
476 at 2400xg and 4°C for 3 minutes. The supernatant was removed, and the pellet was
477 resuspended in 100 µL of Cell Wash Buffer. Cells were counted and diluted to $5x10^7$ cells/mL,
478 split into 100 µL aliquots, and centrifuged for 2 min at 800xg and room temperature. The
479 supernatant was removed and replaced with 100 µL priming buffer (40 mM HEPES-KOH, pH
480 7.5; 34 mM Lithium Citrate; 15% (w/v) PEG 8000; 50 mM Cysteine; 1.5 µM Papain). Cells were

481 primed for 45 minutes at room temperature. Afterwards, the priming was quenched with 10 µL of
482 50 mg/mL bovine serum albumin. The primed cells were centrifuged at 1200xg for 4 min at
483 room temperature. After removing the supernatant, the cells were resuspended in 25 µL ice cold
484 SF buffer (Lonza Cat. No. V4SC-2096). Nucelofection reactions were prepared by combining 16
485 µL ice cold SF buffer, 4 µL Cas9 RNP, 2 µL of repair oligo, and 2 µL of the resuspended and
486 primed cells. This reaction mix was loaded into a 96-well nucleofection plate (Lonza Cat. No.
487 AAF-1003S, AAF-1003B), and pulsed with CU 154 (Fig. S4). 100 µL of ice cold recovery buffer
488 (10 mM HEPES-KOH, pH 7.5; 900 mM Sorbitol; and 8% (w/v) PEG 8000) was added
489 immediately to each pulsed well and incubated for 5 minutes at room temperature. Afterwards
490 the entire content from each nucleofection well was transferred into 2 mL of 15/2 Media with the
491 addition of 50 µg/mL *E. pacifica* pellet.
492
493 ***Strain isolation and Screening***
494 We adapted the Cas12a DETECTR genotyping assay[97] to screen for cells with desired
495 mutations from Cas9 genome editing. Following overnight recovery post-nucleofection, cells
496 were counted and diluted to 45 cells/mL in 10% RA (Table S1) with 50 µg/mL *E. pacifica* pellet
497 and aliquoted into 96-well plates. Plates were then grown at 27°C for 72 hours to propagate
498 cells. Afterwards 12 µL from each well was added into 36 µL DNAzol Direct (20 mM potassium
499 hydroxide; 60% (w/v) PEG 200, pH 13.3-13.7– Note: It is important to test a range of pH values
500 to establish the optimal pH for your own use)[25] that was pre-aliquoted into 96-well PCR plates,
501 which we called the lysed sample. The lysed samples were then incubated at 80°C for 10
502 minutes. The target DNA was amplified in PCR reactions prepared with GoTaq Clear Master Mix
503 (Promega, Cat. No. M7123) using 2 µL of the lysed samples per 25 µL PCR reaction. Cycling
504 conditions followed Promega guidelines and melt temperatures for oligos were calculated using
505 (https://www.promega.com/resources/tools/biomath/tm-calculator/). During the PCR, a Cas12a
506 mastermix was assembled in two parts: First, Cas12a RNP: (13 µl water; 2 µl r2.1 buffer [NEB,
507 Cat. No. B6002S]; 2.5 µl 100 µM gRNA [ordered as a synthesized oligo]; 2 µl 100 µM
508 LbaCas12a [NEB, Cat. No. M0653T]) was incubated for 5 minutes at room temperature.
509 Second, the Cas12a RNP was combined with the rest of the components (for one 96-well plate:
510 486 µl water; 60 µl r2.1 buffer; 18 µl Cas12a RNP; 36 µl 5 µM ssDNA probe [IDT, Cat. No.
511 11-04-02-03]) and then incubated for 5 minutes at room temperature. 5 µL of the mastermix was
512 added to each 25 µL PCR and incubated at 37°C for 1 hour. The fluorescent signal was then
513 measured in the VIC channel on a QuantStudio 3 Real-Time PCR System. Wells with high
514 signal (≥10-fold above background) were recovered in 25% RA overnight at 27°C, counted,

515 diluted to 3 cells/mL in 10% RA with 50 µg/mL *E. pacifica* pellet, and plated in a 96-well plate
516 with 100 µL/well, which corresponds to 0.3 cells/well. Plates were grown at 27°C for 72-96
517 hours. Finally, these clonal isolates were screened by digesting amplicons with a restriction
518 enzyme (RE). 12 µl per isolate was added to 36 µl DNAzol, incubated at 80°C to make a lysed
519 sample. 2 µl of the lysed samples were added to 25 µl PCR reactions with GoTaq Green
520 mastermix (Promega, Cat. No. M7123). After amplification, 1 µl of RE was added to each 25 µl
521 PCR reaction, and incubated following manufacturers' protocols. RE digestions were run on 1%
522 agarose gels with wild-type or undigested controls. For positive RE digestion hits, PCR products
523 were Sanger sequenced to confirm the sequence of the edited site.
524
525 ***Growth curves***
526 ***S. rosetta*.** We monitored cell density at times points over 48 hours to assess the population
527 dynamics of wild type and *cytb561*^{PTS} thecates in iron utilization assay conditions. Cells were
528 plated in 24-well plates with 100 µM ferric EDTA or ferric colloids sources in 4% PG or in
529 nutrient replete media (25% RA) as a control. At each time point, the entire contents of one well
530 was harvested by scraping the bottom and then transferring the liquid to a microfuge tube. Cells
531 were fixed with 5 µl of 37.5% paraformaldehyde (PFA) and then counted on a Reichert
532 Bright-Line Metallized Hemacytometer (Hausser Scientific, Cat. No. 1483). Growth curves were
533 fit to a logistic growth equation that explicitly models the lag time with a heaviside step function[9].
534 ***E. pacifica*** growth curves were set up in 4% PG media inoculated with 50 µg/mL iron-limited *E.*
535 *pacifica* pellet and a ferric EDTA titration in clear bottom 96-well plates, and grown for 48 hours
536 at 27°C. Growth was assessed by monitoring the optical density at 600 nm on a plate reader
537 with orbital shaking before every reading. Ferric colloids were not included in the growth curves
538 as iron particulates would shift optical density readings. To assess *E. pacifica* growth with ferric
539 colloids, 4% PG media was inoculated with 50 µg/mL iron-limited *E. pacifica* pellet and 100 µM
540 ferric EDTA or ferric colloids in 6-well plates and grown for 48 hours at 27°C. Wells were
541 scraped and 1 mL aliquots were centrifuged at 500xg for 2 minutes to settle iron particulates.
542 The top 500 µl of supernatant was gathered for measuring the optical density at 600 nm.
543
544 ***Immunofluorescence***
545 Cultures of wild-type and ALFA-tagged strains were maintained in 25% RA at 18°C for thecates
546 and 22°C for slow swimmers, as higher temperatures helped to maintain slow swimmers. 24
547 hours prior to imaging, cultures were seeded at 10^4 cells/mL in 6 mL of 25% RA in 25cm^2 vented
548 tissue culture flasks (VWR Cat. No. 10062-874) and grown at 27°C. On the day of imaging, prior

549 to preparing cells, 50 µl of 10 mg/mL Poly D-lysine hydrobromide (MP Biomedicals, Cat. No.
550 102694) was added to each well of a glass bottom 96-well plate (Thermo Scientific, Cat. No.
551 165305) and incubated for ≥15 min at room temperature. To prepare cells, the supernatant of
552 thecate cultures was decanted, and replaced with 6 mL filtered ASW, then scraped with a cell
553 scraper, and centrifuged at 500xg for 5 minutes at room temperature. For swimmer cultures, the
554 cells were centrifuged 500xg for 5 minutes at room temperature. The supernatant removed
555 leaving 0.5 mL remaining, and resuspended in 5 mL filtered ASW and centrifuged 500xg for 5
556 min at room temperature. Both thecate and swimmer cultures follow the same protocol
557 henceforth. The supernatant was removed, leaving 0.5 mL, and the tubes gently swirled to
558 resuspend cells. The poly D-lysine coated wells were washed 3 times with 50 µl ASW, and then
559 50-100 µl of cells were added to each well using wide-bore pipette tips. Cells settled on the well
560 surface for 15 minutes. All subsequent steps use gel loading tips. After settlement, the
561 supernatant was removed leaving 25 µl behind to avoid drying and damaging delicate cell
562 structures. 70 µl of Fix buffer (10 mM MES-KOH, pH 6.1; 138 mM KCl; 3 mM $MgCl_2$; 2 mM
563 EGTA; 3% (v/v) PFA; 15% (w/v) sucrose) was slowly added to the top of each well and
564 incubated for 5 min at room temperature. 70 µl of supernatant was removed from the opposite
565 side from which solution was added. This will be referred to as a “wash”. Next, wells were
566 washed with Fix/tween buffer (Fix buffer with the addition of 0.07% (v/v) Tween-20) and
567 incubated for 5 min at room temperature. Wells were then washed with a quench buffer (10 mM
568 MES-KOH, pH 6.1; 138 mM KCl; 3 mM $MgCl_2$; 2 mM EGTA; 15% (w/v) sucrose; 0.3M glycine)
569 and washed immediately with Permeabilization-MeOH buffer (8% (v/v) MeOH; 1% (v/v)
570 Tween-20; in LICOR Intercept Blocking Buffer [LI-COR Biosciences Cat. No. 927-70001]) and
571 incubated for 15 minutes. (Note: LICOR Intercept Blocking buffer increased detection of AFLA
572 epitopes during testing.) After, wells were washed with Permeabilization buffer (1% (v/v)
573 Tween-20; LICOR Intercept Blocking Buffer) and then washed twice in antibody mix (1:200
574 FluoTag-X2 anti-ALFA Alexa647 [Nanotag, Cat No. N1502-AF647-L]; 1:200 phalloidin Alexa546
575 [Invitrogen, Cat. No. A22283]; 1:500 alpha Tubulin Monoclonal Antibody, DM1A [Invitrogen, Cat.
576 No. 62204]; 1:500 alpaca anti-mouse IgG1, recombinant VHH, Alexa488 [Chromotek, Cat. No.
577 sms1AF488-1]) diluted in the Permeabilization buffer and incubated for 1 hour protected from
578 light. Afterwards, wells were washed twice with PEM (100 mM PIPES-KOH, pH 6.9; 1 mM
579 EGTA; 1 mM MgSO4) and immediately imaged.
580
581
582

583 ***Microscopy***

584 Samples were imaged by widefield microscopy. The microscope was a Nikon Eclipse Ti2-E
585 inverted microscope outfitted with a D-LEDI light source, Chroma 89401 Quad Filter Cube, and
586 60x CFI Plan Apo VC NA1.2 water immersion objective. Z-stacks of samples were imaged with
587 two-fold binning (120 µm/pixel) on a Nikon Digital Sight 50 M Camera with 100-400 msec
588 exposure times for each channel and variable illumination intensities to extend the dynamic
589 range of signal without photobleaching samples. Images were processed in FIJI by cropping
590 individual cells in 150x200 pixel bounds, projecting the maximal intensity, subtracting the
591 background (75 pixel, sliding paraboloid radius), enhancing the contrast (saturation 0.02% of the
592 pixels), and applying a minimum filter of 0.5 pixels. Merged images were combined using
593 maximum intensity composite.

594

595 ***Western blots***

596 Cultures were grown in 25% RA or 4% PG (without supplemented iron) for 24 hours at 27°C to
597 test whether external iron conditions (replete or deplete) resulted in any post-translational
598 regulation of protein expression. Cultures were scraped and then lysed according to the
599 preferential lysis protocol, with the following modifications: TurboDNase and SUPERasein were
600 replaced with 15 µl of Pierce universal nuclease (Thermo Scientific Cat. No. 88702) per 1mL
601 buffer, and heparin and cycloheximide were removed. After clearing the lysate, Tween-20 was
602 added to a final concentration of 1% (v/v), and the sample was incubated for 10 minutes at room
603 temperature to aid in liberating membrane-bound proteins. Protein concentration was
604 determined by Bradford assay (Thermo Fisher Scientific Cat. No. 23236) and samples were
605 normalized to the lowest concentration. To prepare for SDS-PAGE, approximately 20 µg or 40
606 µg of total protein for thecates or slow swimmers, respectively, was denatured with 4x Laemmli
607 SDS loading buffer and incubated for 10 minutes at room temperature (Note: 80°C caused
608 membrane proteins to crash out, so room temperature incubation was essential). Samples were
609 loaded into a Novex wedgewell 10% Tris-Glycine Mini Gel (Thermo Fisher Scientific Cat. No.
610 XP00105BOX) along with a 1:10 diluted PageRuler Plus Prestained Protein Ladder (Thermo
611 Fisher Scientific Cat. No. 26619). Gels were run at 225 V for 30 minutes, and afterwards the gel
612 was transferred to a 0.45 µM nitrocellulose membrane (Bio-rad Cat. No. 1620115) at 50 V for 30
613 minutes, then 100 V for 60 minutes in transfer buffer (25 mM tris base; 192mM glycine; 10%
614 methanol) at 4°C. Afterwards the membrane was stained with total protein stain (LI-COR
615 Biosciences Cat. No. 926-11010) and then imaged in the 700 nm channel with a 30 second
616 exposure on a LI-COR Odysey Fc. Afterwards, the membrane was blocked in LICOR Intercept

617 Blocking Buffer (LI-COR Biosciences Cat. No. 927-70001) for 1 hour at 37°C and then overnight
618 at 4°C. The following day the membrane was immunostained with FluoTag-X2 anti-ALFA
619 LI-COR IRDye 800CW (NanoTag Biotechnologies Cat. No. N1502-Li800-L) diluted 1:10,000 in
620 LI-COR PBS block plus 0.2% Tween-20 for 2 hours at room temperature. (Note: Both the
621 LI-COR PBS block and anti-ALFA IRDye 800CW were essential for visualization, other blocks
622 and fluorophores lead to significantly lower signal.) The membrane was washed 3x in PBST
623 (PBS with 0.2% Tween-20) and then imaged in the 800 nm channel with a 10 minute exposure.
624 This western protocol was adapted from a previous study[91].
625
626 ***Protein database searches and tree construction***
627 We compiled a list of iron acquisition proteins from these model organisms: *Homo sapiens*[22],
628 *Saccharomyces cerevisiae*[26], *Chlamydomonas reinhardtii*[21], *Arabadopsis thaliana*[98], and
629 *Phaeodactylum tricornutum*[19]. For each protein, we retrieved the corresponding Pfam hidden
630 markov model (HMM) from Interpro (https://www.ebi.ac.uk/interpro/)[99]. Using HMMER v3.4
631 (hmmer.org[100]), we searched proteomes from EukProt v3[101], focusing on proteomes from "The
632 Comparative Set" (TCS), which consisted of 196 species of phylogenetic relevance with high
633 BUSCO completion. We supplemented the TCS with the transcriptomes of 18 choanoflagellate
634 species[33] and other Holozoa and Fungi. To count the number of unique proteins that conformed
635 to the HMM profile from 'hmmsearch', we aligned hits with 'hmmalign' and retained only one
636 sequence from hits with >99% identity (Note: this was important for removing multiple isoforms
637 and reference sequences from highly annotated genomes). We additionally filtered hits based
638 on length with 'esl-alimanip' to retain protein sequences that only possessed the HMM profile,
639 which eliminated multidomain proteins that only had the HMM profile as a one of its domains
640 (Table S3 for length filter thresholds). (Note: The number of hits prior to filtering can be found in
641 Table S3) For HMM profiles that yielded a high number of protein hits that were not well-filtered
642 (e.g. Transferrin Receptor), we aligned sequences with MAFFT-DASH[40], removed sites that
643 were uninformative or contained 90% gaps with ClipKIT[102] (kpic-gappy 0.9), and generated
644 phylogenetic trees with IQ-TREE[103] using ModelFinder[104] to choose the appropriate substitution
645 model for each set of protein alignments. This procedure was also used to determine the
646 phylogenetic relationships of Cytb561 paralogs (Fig. 4A), and the chosen substitution model
647 was LG+F+R7. Trees were visualized with iTOL[105].
648
649
650

Gene distribution maps and correlations with upwellings.

***cytb561* paralog search.** The abundance and location of *S. rosetta* cytb561 paralogs was searched for using the Ocean Gene Atlas[47,48]. Protein sequences of each paralog from *S. rosetta* were searched for in the MATOUv1+G (for metagenomes) and MATOUv1+T (for metatranscriptomes) databases, using tblastn with the following search parameters: expected threshold E-values < 1^{-10}, and abundance as percent of total reads. The Ocean Gene Atlas automatically assigns protein sequences to phyla and classifies them on expected organism size, such that the results could be filtered first by size (0.8-5 µm) and then by phyla to identify only choanoflagellate specific hits. Therefore to identify choanoflagellate specific hits and the location they were gathered from, three tables were gathered from each search: the distribution of E-values (E-values per hit and their assigned phyla), abundance matrix for the 0.8-5 µm fraction (the abundance as a fraction of total reads and the sample site ID for each hit), and environmental parameters for the 0.8-5 µm fraction (all measured and inferred environmental measurements for each hits sample location). From these three tables, choanoflagellate assigned hits, abundances, and sample locations were compiled.

Upwelling velocities. To identify average vertical velocities at the sample sites from the Tara Oceans Expedition[106], from which the Ocean Gene Atlas acquires its data, upward seawater velocities were gathered from the Copernicus Marine Services' Global Ocean Physics Analysis and Forecast model's (https://doi.org/10.48670/moi-00016) dataset (cmems_mod_glo_phy-wcur_anfc_0.083deg_P1M-m). The coordinates for each sample site were compiled[106] and vertical velocities were gathered from the dates 10/31/2020 - 5/20/2024. Then the search was done using a modified script offered by the Copernicus Marine Services (https://help.marine.copernicus.eu/en/articles/7970637-how-to-download-data-for-multiple-points-from-a-csv). The vertical velocity at each sample site was then averaged across the 3.5 year range to account for temporal variation (Fig. S8B).

Correlations and maps. Choanoflagellate paralog abundances were tabulated with their sample site ID, sample coordinates, and average upwelling velocity (from above). First, the abundances and average vertical velocities were correlated by the Pearson function in Microsoft Excel. To calculate the associated *P* value for each Pearson correlation, a T statistic was calculated using the formula: $t = (r \cdot [n-2]^{½}) \cdot (1-r^2)^{-½}$, where *r* is the calculated Pearson correlation, *n* is the sample number, *t* is the T statistic. This statistic was then input into the t distribution function (TDIST) in Excel, along with the degrees of freedom (*n*-2), to calculate the *P* values. Next, the abundances and average vertical velocities were plotted in a scatter plot to manually inspect the data for outliers. If a datum was suspected of being an outlier, a Dixon's Q

test was conducted for that datum with the formula: $Q = |gap|/range$, where gap = the difference between the suspected outlier and the next nearest datum, and range = the range of the data set. The value is considered an outlier and removed if $Q > Q_{table}$, where Q_{table} is a reference value determined by the sample size and confidence level. Of all tested outliers, only one datum was removed, belonging to *cytb561a* transcript abundance at TARA site 52 (-16.96°, 53.99°), with a Q = 0.8718 and where Q_{table} = 0.3323 for 99% confidence level for a sample size of 33. Without this outlier removed the statistics for *cytb561a*'s correlation would be r = 0.09 and P = 0.58. Maps were plotted with the compiled table of paralog abundances and coordinates, using matplotlib with modified scripts generated by ChatGPT (https://chat.openai.com).

Data Availability:

RNA sequencing data is available in the public repository NCBI Gene Expression Omnibus (GEO) database under requisition ID: GSE267344

Author Contributions:

FL and DSB conceived of the project, designed experiments, conducted research, performed data analysis, and wrote the manuscript. JMEE and VD generated strains and developed the genotyping pipeline. MCC and SE performed an early analysis of RNA-seq data.

Acknowledgements:

We are enormously thankful to Nicole King for supporting an early phase in this project and providing helpful feedback. We thank Alain Garcia de las Bayonas, Flora Rutaganira, Ben Larson, members of the Booth lab, and the students at the Marine Biological Laboratories (MBL) 2023 Physiology Course for experimental advice and feedback. We also appreciate helpful discussions and feedback from Elizabeth Blackburn, Jeremy Reiter, and Wallace Marshall. FL was supported by a National Science Foundation Graduate Research Fellowships Program, a UCSF Discovery Fellowship, and the UCSF Tetrad Graduate Program Training Grant T32GM139786. This work was supported by awards to DSB from the Chan Zuckerberg Biohub, the UCSF Sandler Program for Breakthrough Biomedical Research, a David and Lucile Packard Foundation Fellowship in Science and Engineering, and a Maximizing Investigators' Research Award for Early Stage Investigators from the National Institute of General Medical Sciences R35GM147404. Sequencing was additionally supported by an instrumentation grant from the National Institutes of Health S10OD018174 to the QB3 Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley.

718 **Figures**:

719

720 **Figure 1:** Thecates from the choanoflagellate *S. rosetta* are a distinct cell type.
721 **(A)** *S. rosetta* differentiates into morphologically distinct cell types. This schematic shows four of
722 the cell types from *S. rosetta* that were stably cultured to produce transcriptome profiles. All of
723 these cell types display the common choanoflagellate cell architecture in which an apical
724 flagellum is encircled by a collar of actin-filled microvilli (indicated on the slow swimmer) that
725 enables choanoflagellates to phagocytose bacteria and particulate matter. In the study, the only
726 provided prey bacterium was *E. pacifica.* When nutrients are abundant, slow swimmers (green)
727 respond to bacterial and algal cues to develop into multicellular rosettes (blue) that form by
728 serial cell division. Cultures of slow swimmers also form chains of cells through serial cell
729 divisions, but those chains are easily disrupted by mechanical force. Under starvation, slow
730 swimmers and rosettes become fast swimmers (grey), which have a reduced cell body and
731 collar with a longer flagellum. Under sustained nutrient deprivation, fast swimmers differentiate
732 into thecates (red), a type of cell that adheres to substrates by constructing an extracellular
733 apparatus called a theca. Thecates still proliferate in low nutrient conditions by dividing into
734 swimmers that build their own theca. By taking the supernatant of thecate cultures in high
735 nutrient conditions, thecates can then differentiate into slow swimmers.
736 **(B)** The transcriptome profile of thecates stands apart from all other cell types. A principal
737 component analysis of triplicate RNA-seq profiles from each cell type (panel A) shows that the
738 86% of the variance between samples is attributed to the thecate transcriptome profile. All other
739 cell types cluster closely together.
740 **(C)** Genes with increased expression in thecates encode proteins that mediate signal
741 transduction, nutrient acquisition, and gene regulation. Genes with transcript abundances that
742 reliably ($q < 0.01$) changed more than two-fold in thecates were analyzed for the enrichment of
743 Gene Ontology (GO) categories, focusing on GO molecular function categories (y-axis). Genes
744 that were upregulated in thecates and associated with an enriched GO molecular function
745 category were counted (x-axis). Each bar on the graph is colored by the *P*-value, which reflects
746 the probability that a given category was represented by a random selection of genes from the
747 *S. rosetta* genome.

748

749 **Figure 2:** Thecates utilize ferric colloids through the expression of an iron reductase, *cytb561a.*
750 **(A)** Thecates highly express an iron reductase. The transcript abundances (counts) from
751 thecates were divided by the abundances from slow swimmers samples and $\log_2$ transformed

752 (x-axis) and the *P*-values for gene expression comparisons were corrected for multiple
753 comparisons (*q*, y-axis). Among the genes with transcript abundances that reliably ($q < 0.01$)
754 changed more than two-fold in thecates (red) compared to swimmers (green), PTSG_09715 is
755 in the 98th percentile. This gene, which is annotated as an iron reductase, encodes a single
756 domain: cytochrome b561 (Pfam 03188).
757 **(B)** A method to swap iron sources that support choanoflagellate growth. In preparation for iron
758 limitation assays, *S. rosetta* was passaged in low-nutrient media and depleted of iron. Cells
759 were then washed multiple times and inoculated with iron-limited feeder bacteria. At the same
760 time, ferric EDTA (Fe^{3+}•EDTA) or ferric colloids (Fe^{3+}) were provided. Cultures grew for 48 hours
761 at 27°C before assaying *cytb561a* expression (Fig. 2C and S2) and growth (Figs. 2D and S4)
762 **(C)** *Cytb561a* expression is part of the thecate regulon, and not determined by external iron
763 conditions. The expression of *cytb561a* was monitored by RT-qPCR in cultures of slow
764 swimmers (green) or thecates (red) that were grown with either ferric EDTA or ferric colloids.
765 The expression of *cytb561* was normalized to *cofilin* (PTSG_01554), a eukaryotic gene that
766 displays high, consistent expression across all *S. rosetta* cell types. Independent triplicates were
767 performed and *P*-values were calculated from a two-way ANOVA.
768 **(D)** Thecates require *cytb561* for increased proliferation with ferric colloids. To account for
769 different growth parameters between slow swimmers (green) and thecates (red), the cell density
770 of cultures grown with ferric colloids was normalized to those grown with ferric EDTA. With this
771 metric, a ratio greater than one indicates that the cell type displays increased growth with ferric
772 colloids; whereas, a ratio less than one indicates the converse. A premature termination
773 sequence introduced at position 151 in *cytb561* with CRISPR/Cas9 genome editing produced a
774 mutant allele (*cytb561*^{PTS}) with a ten-fold reduction in *cytb561a* mRNA levels and stop codons
775 that would truncate the protein translated from this transcript (Fig. S2 and Table 2). Unlike the
776 wild-type strain (circles), thecate cells with *cytb561a*^{PTS} displayed no improved growth with ferric
777 colloids (triangles).
778
779 **Figure 3:** Choanoflagellates and animals share core iron acquisition genes.
780 **(A)** Distribution of iron acquisition proteins across diverse eukaryotes. Iron acquisition proteins
781 characterized in model eukaryotes were compiled into a list to search a curated database of
782 proteomes predicted from the genomes and/or transcriptomes of diverse eukaryotes. The iron
783 acquisition proteins are sorted into these categories from the top to bottom rows: transporters,
784 reductases, oxidases, iron storage, and sharing/binding. Each column indicates a single

785 species, and major eukaryotic lineages are denoted by color. The number of unique protein hits
786 identified in each species is shown in greyscale.
787 **(B)** Cell-type expression of iron acquisition genes identified *S. rosetta* cell types. Gene names
788 are presented as accession numbers and newly denoted gene names, based on their homology
789 with animal iron acquisition genes. Expression values are averaged triplicate values of
790 transcripts per million (TPM) from cell type transcriptomes (Fig. 1). Of all genes and paralogs,
791 *cytb561a* exhibits the most striking differential regulation.
792 **(C)** The localization of iron acquisition proteins in thecates and slow swimmers. An ALFA
793 epitope was engineered into endogenous genetic loci at the carboxy terminus of the protein
794 coding sequences. Although this tag was engineered into all iron acquisition genes (except the
795 vacuolar transporters), only the proteins displayed in this panel were visible in western blots
796 (Fig. S7). Immunofluorescent staining of each ALFA-tagged strain was performed with tubulin
797 and phalloidin to serve as markers of the cellular architecture. Note that in slow swimmers,
798 Cytb561a has no detectable expression, consistent with transcript abundance. Images of
799 individual cells are representative of more than three independent staining experiments in which
800 six or more images were captured of fields containing more than ten cells. Scale bar is 5 μm.
801 **(D)** Diagram of the proposed iron acquisition pathway in *S. rosetta*. Iron is internalized by
802 phagocytic vesicles derived at the base of the collar. Inside, Cytb561a reduces iron for transport
803 into the cytosol by Dmt1. Iron is then metabolized, incorporated, and utilized as needed by the
804 cell. To maintain iron homeostasis, Fpn exports iron out of the cell, and Heph aids in maintaining
805 a favorable concentration gradient for export.
806
807 **Figure 4:** Cytb561 paralogs possess distinct biochemical properties and appear at different
808 oceanic locations.
809 **(A)** Phylogeny of iron reductases reveals Cytb561a is an ortholog of animal DCYTB. Each
810 paralog of *S. rosetta* falls into distinct clades of cytochrome b561. Circles on each branch are
811 proportional to bootstrap values greater than sixty. Scale bar indicates the average number of
812 substitutions per site. Branch colors correspond to clades as shown in the legend.
813 **(B)** Predicted structures of paralogs show differences in the substrate-binding interface.
814 Predicted electrostatic surfaces from AlphaFold predictions of Cytb561 paralogs show different
815 charge distributions around the substrate binding pocket. Structure models angled 30° for
816 viewing
817 **(C)** *cytb561a* transcript abundance correlates with marine upwellings. Maps (left panels) show
818 the genomic (grey) and transcript (blue) abundance of each paralog. Plots show genomic

819 (middle panels) and transcript (right panels) abundance on the y-axis and average upwelling
820 velocities on the x-axis. Note, the x-axis for *cytb561c* transcript abundance is scaled differently
821 than for *cytb561a* and *cytb561b*. *r* indicates Pearson correlation and *P*-values were calculated
822 from a t-test.

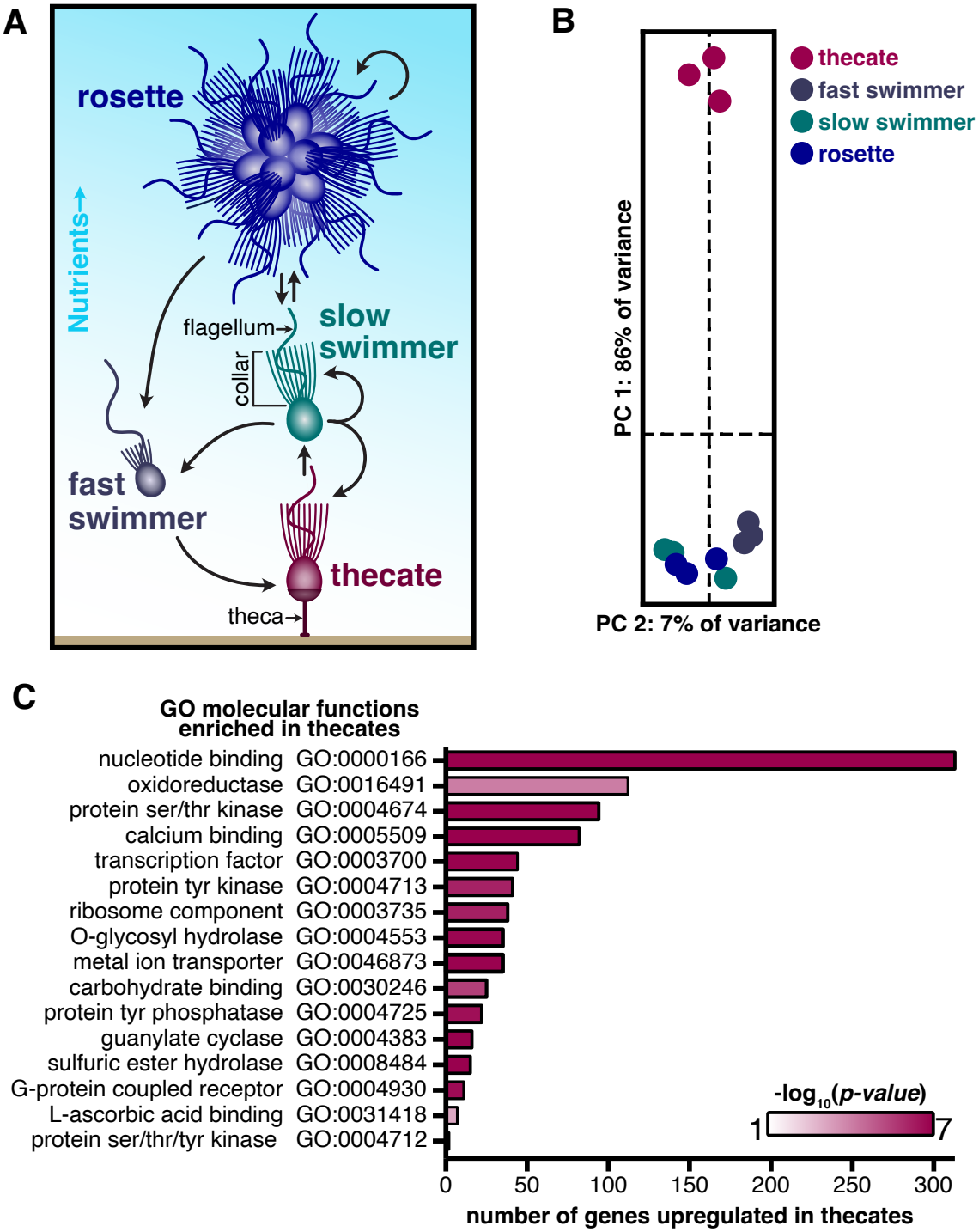


Figure 1: Thecates from the choanoflagellate *S. rosetta* are a distinct cell type.

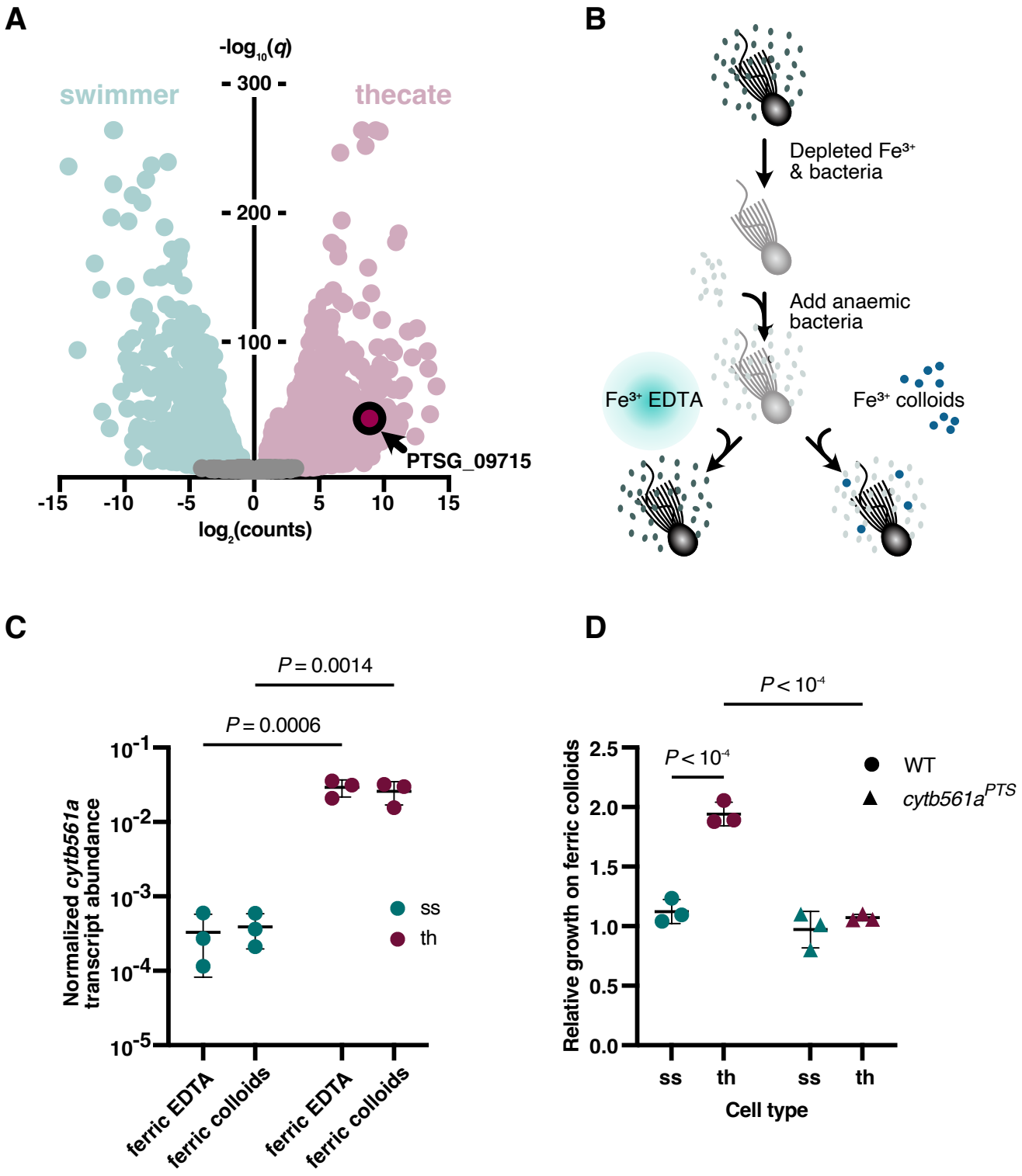


Figure 2: Thecates utilize ferric colloids through the expression of an iron reductase, *cytb561a*.

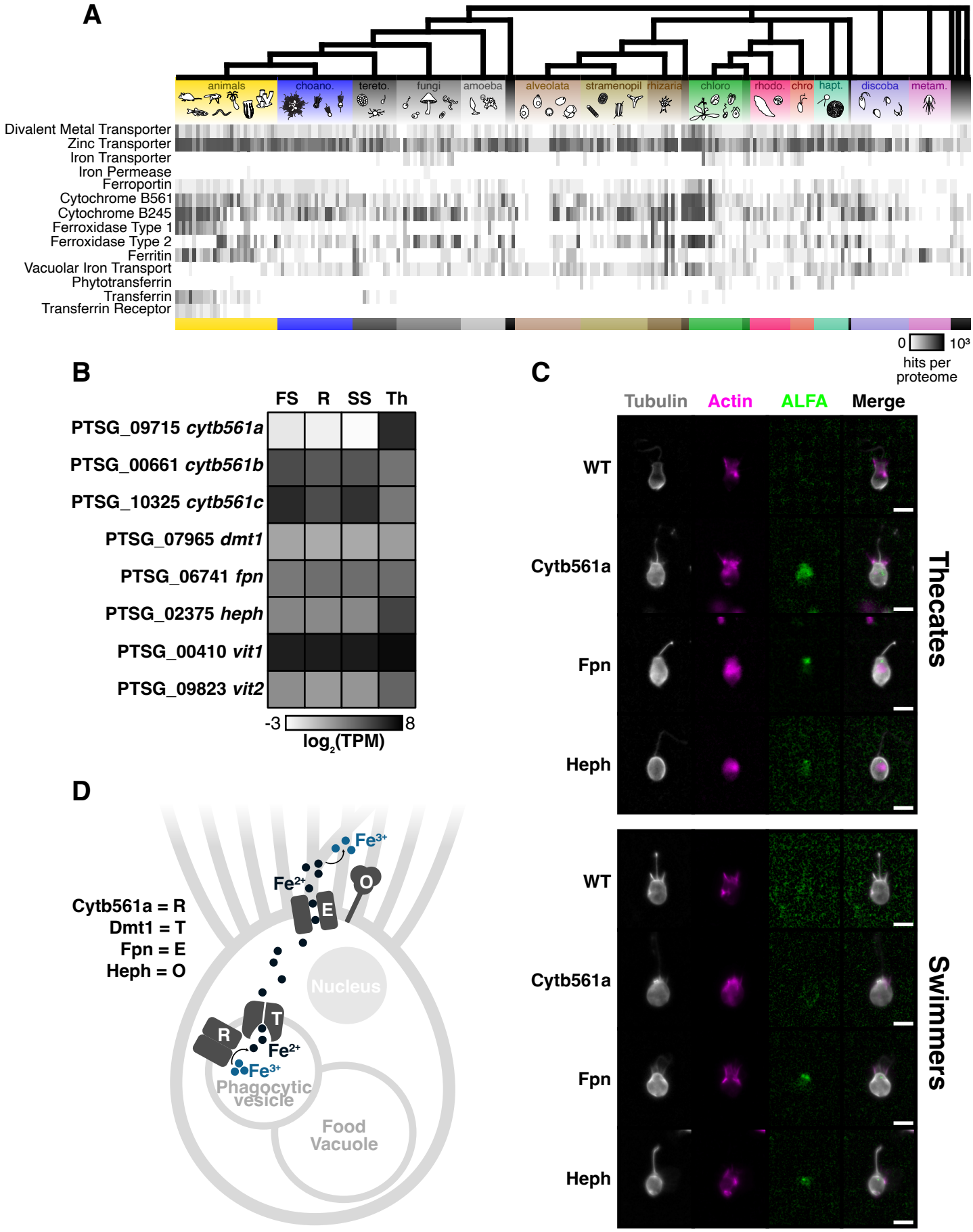


Figure 3: Choanoflagellates and animals share core iron acquisition genes.

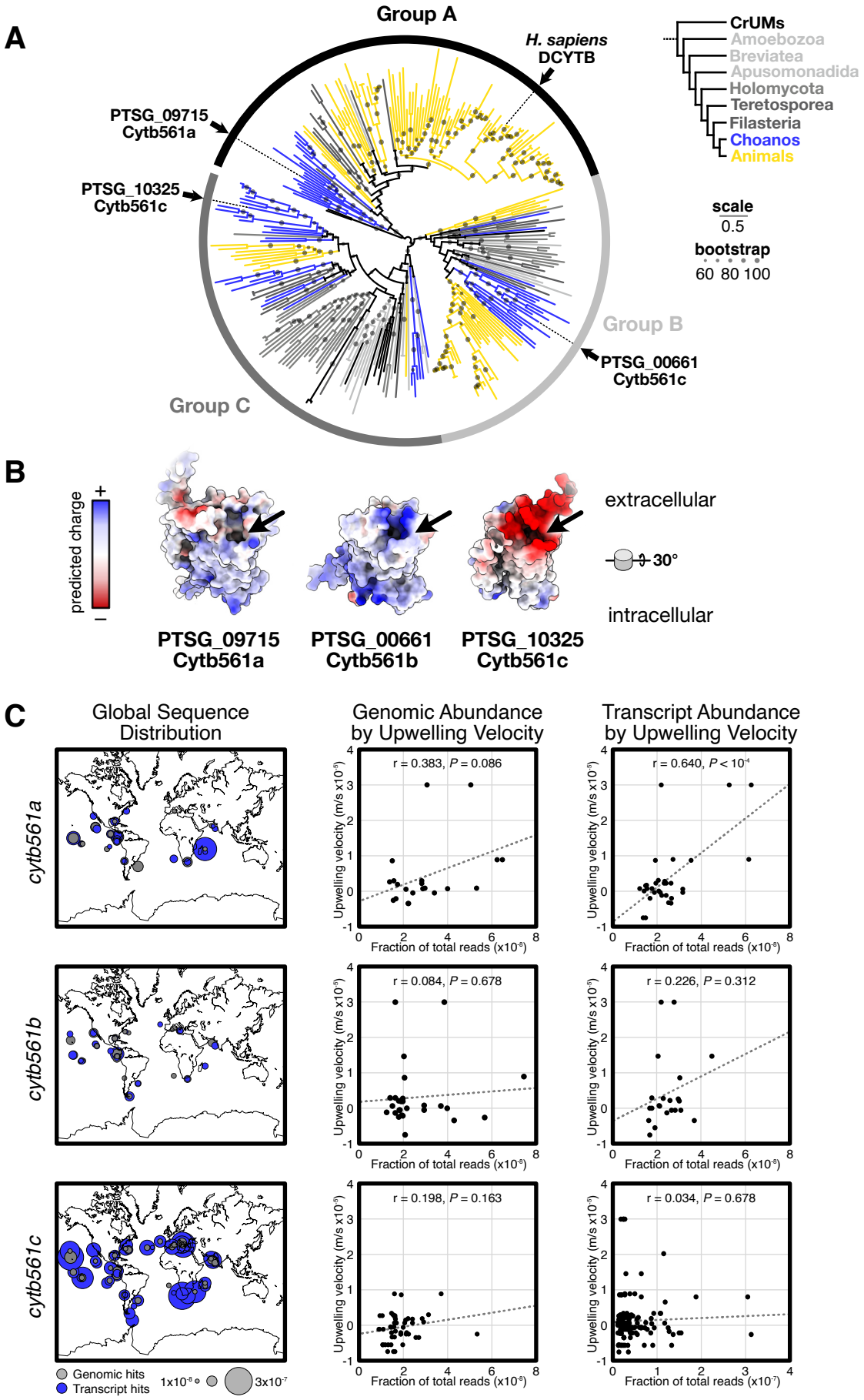


Figure 4: Cytb561 paralogs possess distinct biochemical properties and appear at different oceanic locations.

823 **Supplemental Figures:**

824

825 **Figure S1.** A preferential lysis buffer enriches RNA from *S. rosetta* to improve transcriptomes.
826 **(A)** A schematic showing the separation of *S. rosetta* RNA from *E. pacifica* RNA.
827 **(B)** The preferential lysis procedure enriches RNA from *S. rosetta*. After performing preferential
828 lysis, RNA was purified from the supernatant using Trizol LS reagent. The RNA from the pellet
829 of preferential lysis, an unlysed sample of *S. rosetta* feeding on *E. pacifica*, or a sample of *E.*
830 *pacifica* was purified with Trizol LS. The purified RNA samples were combined with an equal
831 volume of formamide and incubated at 95°C for 5 minutes. After cooling to room temperature,
832 the samples were run on a 2% (w/v) agarose gel in TBE buffer at 100 V. Afterwards, the gel was
833 stained with SYBR Gold to detect RNA. In the gel, the supernatant (S) from whole lysates (W)
834 mostly contains *S. rosetta* rRNA (28S and 18S); the pellet (P) contains E. pacifica rRNA (23S
835 and 16S).
836 **(C)** Total RNA prepared for RNA-seq primarily contains rRNA from *S. rosetta.* Bioanalyzer
837 traces with predicted prokaryotic and eukaryotic rRNA subunit sizes indicated. Replicate RNA
838 samples used for sequencing demonstrate little to no carryover from bacterial contamination, as
839 shown by the low amount of 16S and 23S rRNA.

840

841 **Figure S2:** Validated qPCR primers confirm the disruption of *cytb561a* expression with the
842 *cytb561a*^{PTS} allele.
843 **(A)** qPCR primers for *cytb561a* and *cofilin* produce a single amplicon. Gel image of PCR
844 products using qPCR primers designed for *cytb561a* and *cofilin*. Note the diminished band from
845 swimmer cDNA likely resulted from lower mRNA copies from slow swimmer cultures, as shown
846 in Fig. 2C. Shadow at 75 bp is bleed-over from loading dye.
847 **(B)** Standard curve for qPCR primers for *cytb561a*. To generate a standard curve, a serial
848 dilution of ssDNA standards was performed in triplicate. The concentration of the ssDNA
849 standard was determined by Qubit. Cycle thresholds (Ct) were converted to transcript copy
850 number by fitting a linear equation to the standard curve. This curve was used to determine the
851 abundance of transcripts from three independent experiments.
852 **(C)** Standard curve for qPCR primers for *cofilin*. The graph is laid out as for panel B.
853 **(D)** *Cytb561a*^{PTS} exhibits lowered *cytb561a* expression compared to wild-type thecates. We
854 compared the normalized expression of *cytb561a* from the strain bearing the *cytb561a*^{PTS} allele
855 to wild-type strains. The wild-type strains were slow swimmers cultured with ferric EDTA or

856 thecates cultured with ferric colloids, for we anticipated that these two conditions would result in
857 the greatest difference in *cytb561a* expression.
858
859 **Figure S3:** *S. rosetta* cell types, but not their feeder bacteria, grow differently with variable iron
860 sources.
861 **(A)** Thecate cell types exhibit differential growth with low concentrations of ferric colloids
862 compared to slow swimming cell types. *S. rosetta* cell density of slow swimmer and thecate
863 cultures when grown with titrations of ferric EDTA or ferric colloids.
864 **(B)** A variety of ferric chelates differentially impact *S. rosetta* growth. Slow swimmers and
865 thecates exhibit different growth characteristics, so to better compare their growth with iron
866 bound to different chelators, we used a ratio of the final cell density of cultures grown with
867 iron-chelator complexes to the final cell density of cultures grown without any supplemental iron.
868 With this metric, a ratio greater than one indicates that the cell type displays increased growth
869 with the iron-chelator complex; whereas, a ratio less than one indicates the converse. All
870 conditions were tested at 100 μM Fe^{3+}, and pyoverdines and deferroxamine are bacterial
871 siderophores.
872 **(C)** *E. pacifica* feeder bacteria growth is easily rescued by low concentrations of iron. Optical
873 density at 600 nm (OD600) of *E. pacifica* measured approximately every 100 minutes for 48
874 hours. Note standard deviations overlap for all conditions except for a no iron control.
875 **(D)** *E. pacifica* grows similarly with ferric EDTA or ferric colloids. Because ferric colloids would
876 influence OD600 measurements in a standard growth curve experiment, *E. pacifica* cultures
877 were grown for 48 hours with 100 μM ferric EDTA or ferric colloids and then lightly centrifuged at
878 500xg for 2 minutes at room temperature to settle iron particulates. Afterwards, the supernatant
879 OD600 was measured.
880 **(E)** Ferric colloids release a small amount of labile iron. When media is prepared with 100 μM
881 ferric colloids, only 3.43 ± 1.01 μM iron is liberated without biological activity. The amount of iron
882 in 4% PG media alone was 0.27 ± 0.036 μM iron.
883
884 **Figure S4:** A screen for optimal nucleofection pulses to improve cargo delivery into *S. rosetta*.
885 **(A)** Nanoluc luminescence values from *S. rosetta* cultures transfected with different pulse codes.
886 Because of changes to the growth media for *S. rosetta*, we conducted a new screen for an
887 optimal nucleofection pulse. Cells were transfected with a plasmid that drives nanoluc
888 expression with a promoter cloned from elongation factor L. Pulse codes correspond to Lonza

4D-Nucleofector codes. The screen was performed in four independent trials. Dot indicates prior pulse code used, asterisk denotes optimized pulse code.

(B) Re-ordered chart of luciferase activity from the pulse screen reveals the landscape to optimize nucleofection pulses. After the first two trials for screening pulses, we hypothesized that the letter and number codes for the nucleofection system are orthogonal parameters, so we selected pulses for the third and fourth trials to more broadly sample the optimization landscape. The maximum luminescence value for each pulse from one of the four trials was used to visualize transfection efficiency as a function of the letter codes (x-axis) or pulse numbers (y-axis). Dot indicates prior pulse code used, asterisk denotes optimized pulse code.

Figure S5: Thecates grow slight faster and to a larger population size with ferric colloids

(A-C) Growth curves of wild type and *cytb561a*^{PTS} thecates show wild type thecates grow to a larger population size when grown with ferric colloids. Growth curves of cultures grown in (A) nutrient replete conditions (25% RA), (B) 4% PG with ferric EDTA, and (C) 4% PG with ferric colloids. Cell densities were counted every 6 hours for 48 hours.

(D-F) Comparisons of growth curve parameters highlight the increase in carrying capacity as the main factor for higher wild type thecate cell type growth with ferric colloids. (D) Carrying capacity, (E) doubling time, and (F) lag time of the growth curves shown in figures (S5B) and (S5C).

Figure S6: Simplified iron acquisition pathways in model eukaryotes.

(A) Mammals. Dietary iron is absorbed by the epithelium of the gut. Ferric cations are reduced and imported at the apical end of epithelial cells. Ferrous cations are exported and oxidized at the basal end into the bloodstream. Ferric cations are bound to transferrin in the bloodstream and distributed to other cells in the body by receptor mediated endocytosis when transferrin binds to the transferrin receptor.

(B) *Saccharomyces cerevisiae.* In the main iron import pathway, ferric cations are reduced, then oxidized, and finally imported in its ferric form for increased substrate specificity. In a minor secondary pathway ferrous cations can be imported in its reduced form by a low affinity importer. Iron is stored intracellularly in vacuoles, via another iron specific transporter.

(C) *Phaeodactylum tricornutum.* Ferric cations can be reduced directly and imported, but they can also be directly bound by phytotransferrin and endocytosed.

921 **(D)** *Chlamydomonas reinhardtii.* Ferric cations are first reduced, then oxidized, and finally
922 imported in its ferric form for increased substrate specificity. Iron can also be directly bound by
923 phytotransferrin, which is thought to be released into the periplasm.
924
925 **Figure S7:** Thecates detectably express Cytb561a, Fpn, and Heph in western blots.
926 **(A)** Western blot of thecates cultures of ALFA-tagged strains, grown in nutrient rich media (25%
927 RA) and iron depleted media (4% PG). Strains are named after the ALFA-tagged gene.
928 **(B)** Western blot of swimmer cultures of alfa-tagged strains, grown in nutrient rich media (25%
929 RA) and iron depleted media (4% PG). Arrow indicates faint Fpn band visible only in the 25%
930 RA condition. The expression of Fpn in media with high iron concentrations and the lack of
931 expression in low iron media is consistent with Fpn regulating homeostatic iron levels through
932 post-transcriptional regulation.
933
934 **Figure S8:** Maps of marine upwelling velocities show increased upwelling near coastlines and
935 the equator
936 **(A)** Global upwelling velocities at 100m depth averaged over a three year span, measured every
937 10° latitude and longitude. Missing values are due to shallow locations or landmasses.
938 **(B)** Upwelling velocities at 100m depth averaged over a three year span at every location
939 sampled during the Tara Oceans Expeditions.
940
941 **Table S1:** Media table
942 Components and recipes for all media used: 5% seawater complete (SWC0, 10% cereal grass
943 media (CGM3), 10% red algae (RA), 25% red algae (RA), 15% red algae, 2% peptone, yeast
944 extract, glycerol (15/2), and 4% peptone, glycerol (PG).
945
946 **Table S2:** Strains, crRNAs, repair templates, and primers used in this study
947 The strains of *S. rosetta* used and created during this study, along with the crRNAs, repair
948 templates, Cas12a screening gRNAs, and screening primers for generating each strain.
949
950 **Table S3:** List of paralog hits for each gene family and tabulation of number of hits per
951 organism.
952 The Pfam ID of each HMM used to search for paralogs, along with the associated protein
953 names used in Fig. 3A, and the length filtering criteria used, when applicable. The next sheet
954 contains paralog hits filtered for uniqueness and length, which is the raw data used in Fig. 3A.

The last sheet is the same paralog heatmap, but without the additional length filtering, resulting in more hits per proteome. Colors of species match Fig. 3A, as does the protein order.

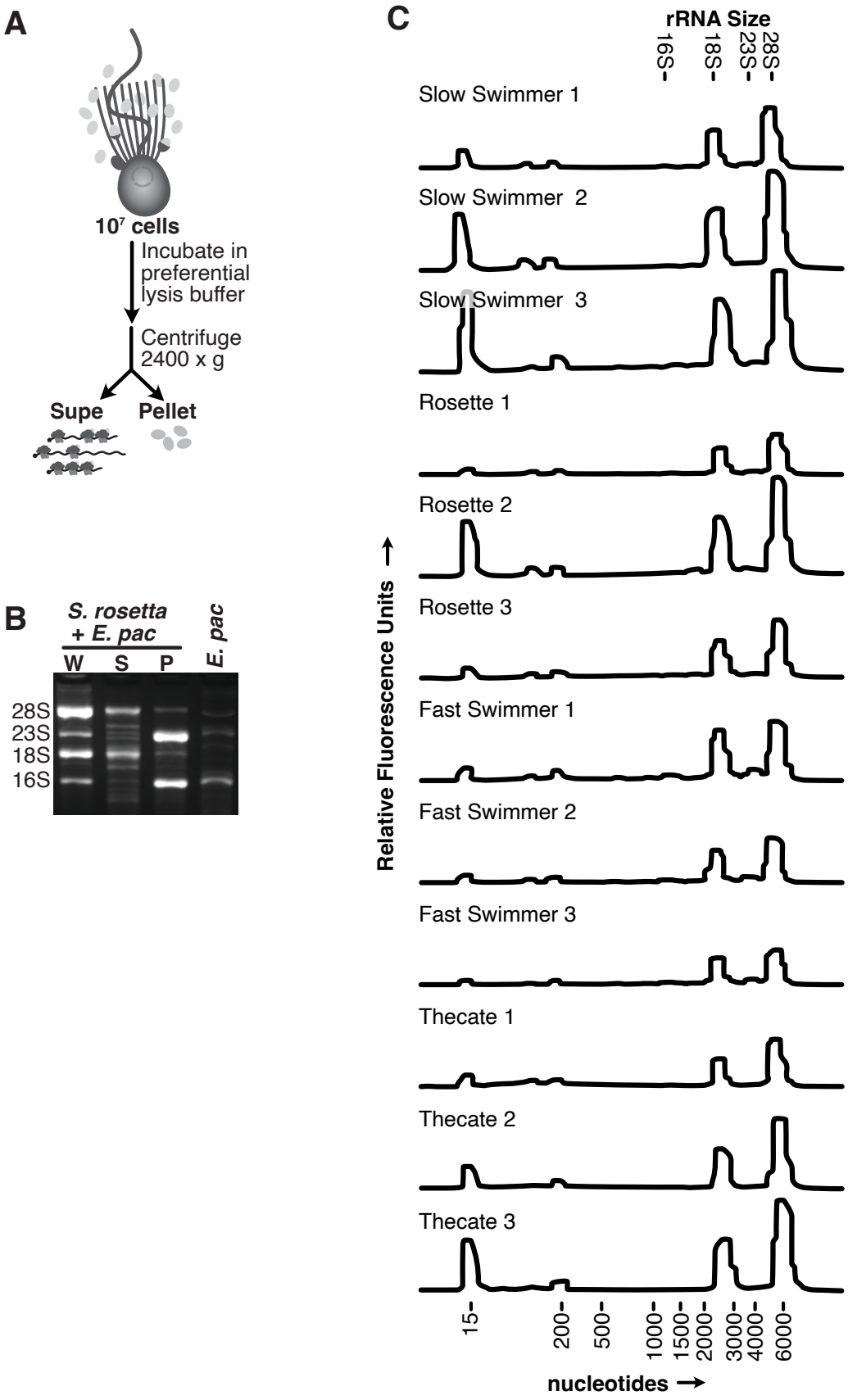


Figure S1. A preferential lysis buffer enriches RNA from *S. rosetta* to improve transcriptomes.

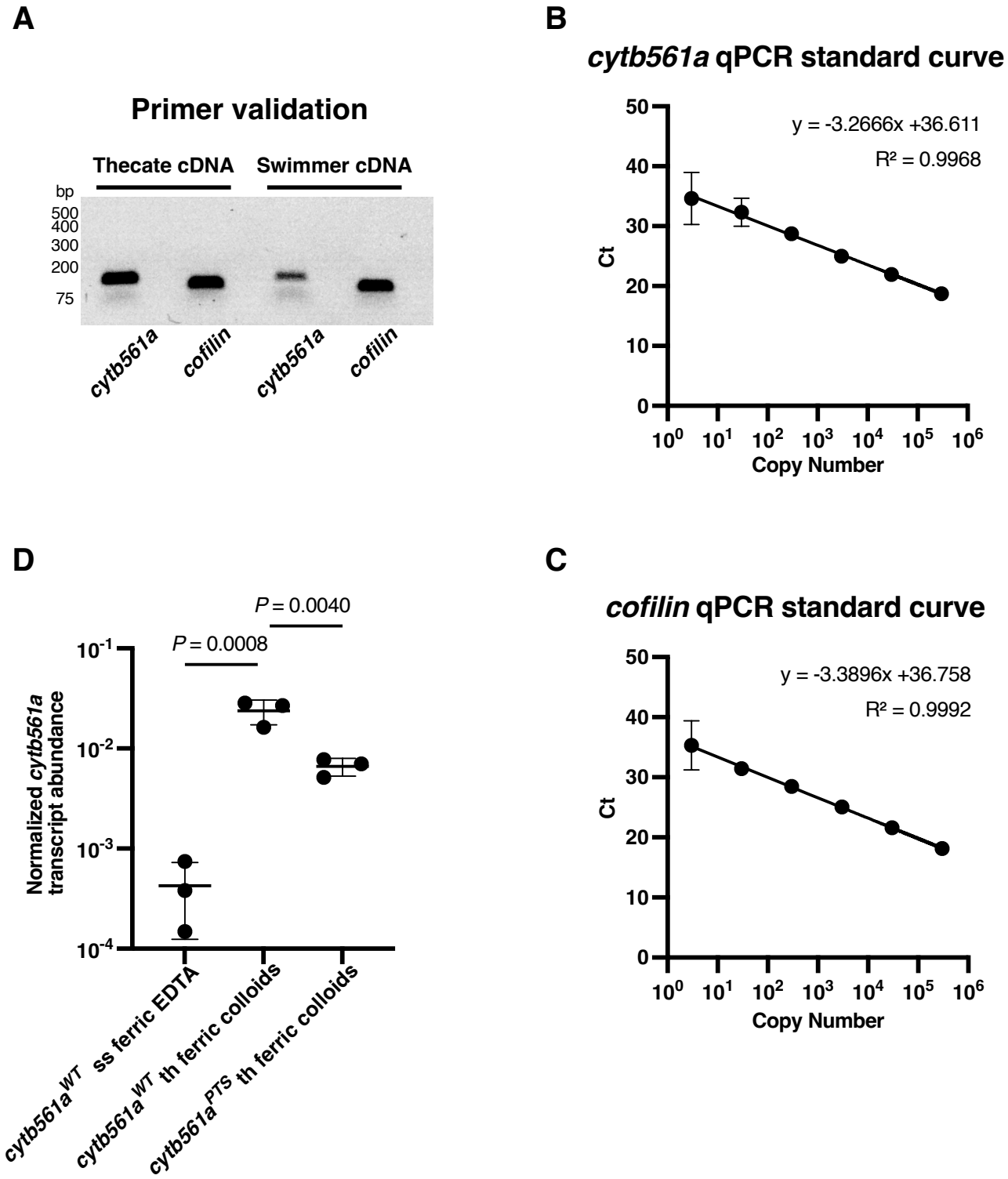


Figure S2: Validated qPCR primers confirm the disruption of *cytb561a* expression with the *cytb561a*^{PTS} allele.

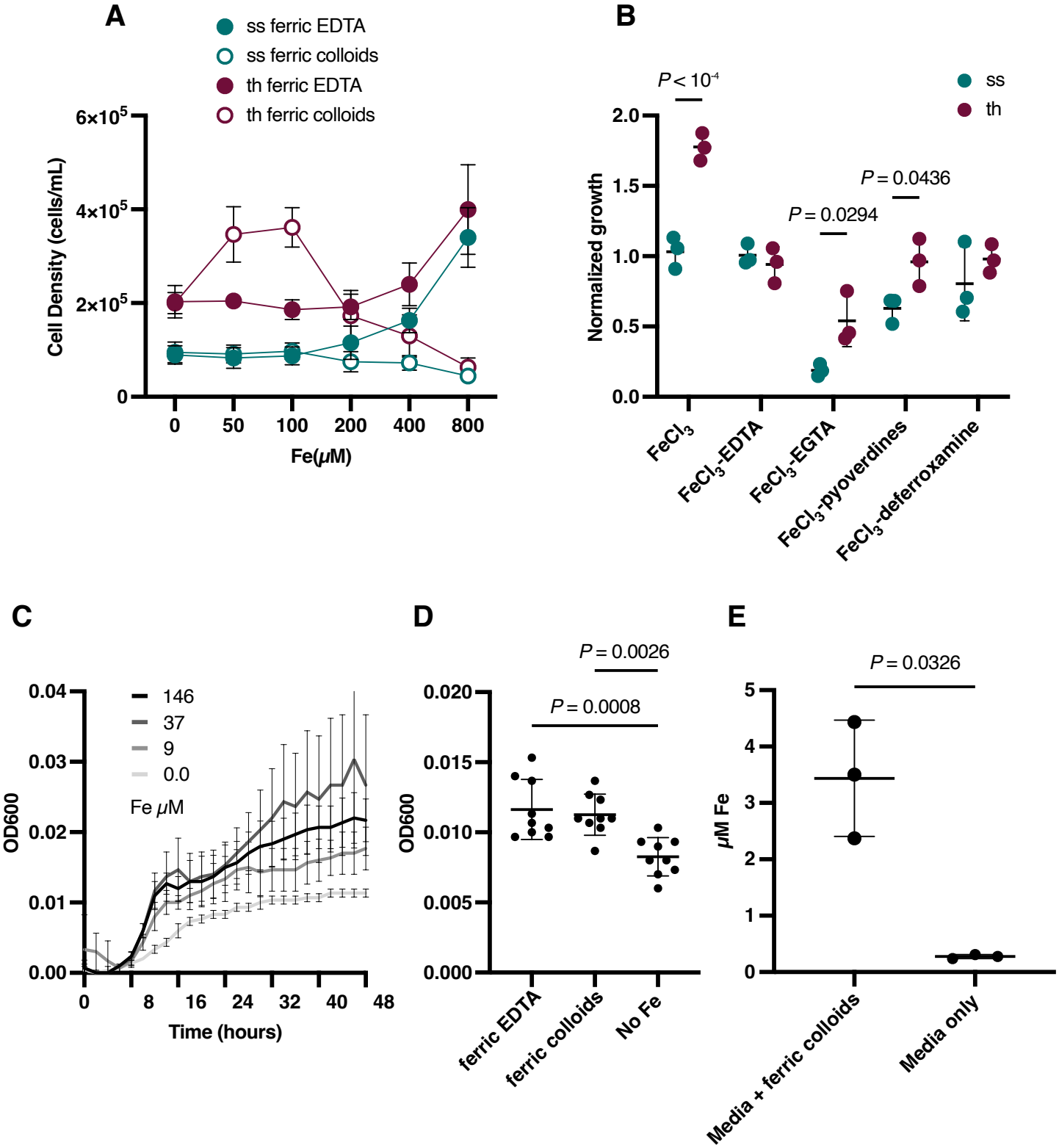


Figure S3: *S. rosetta* cell types, but not their feeder bacteria, grow differently with variable iron sources.

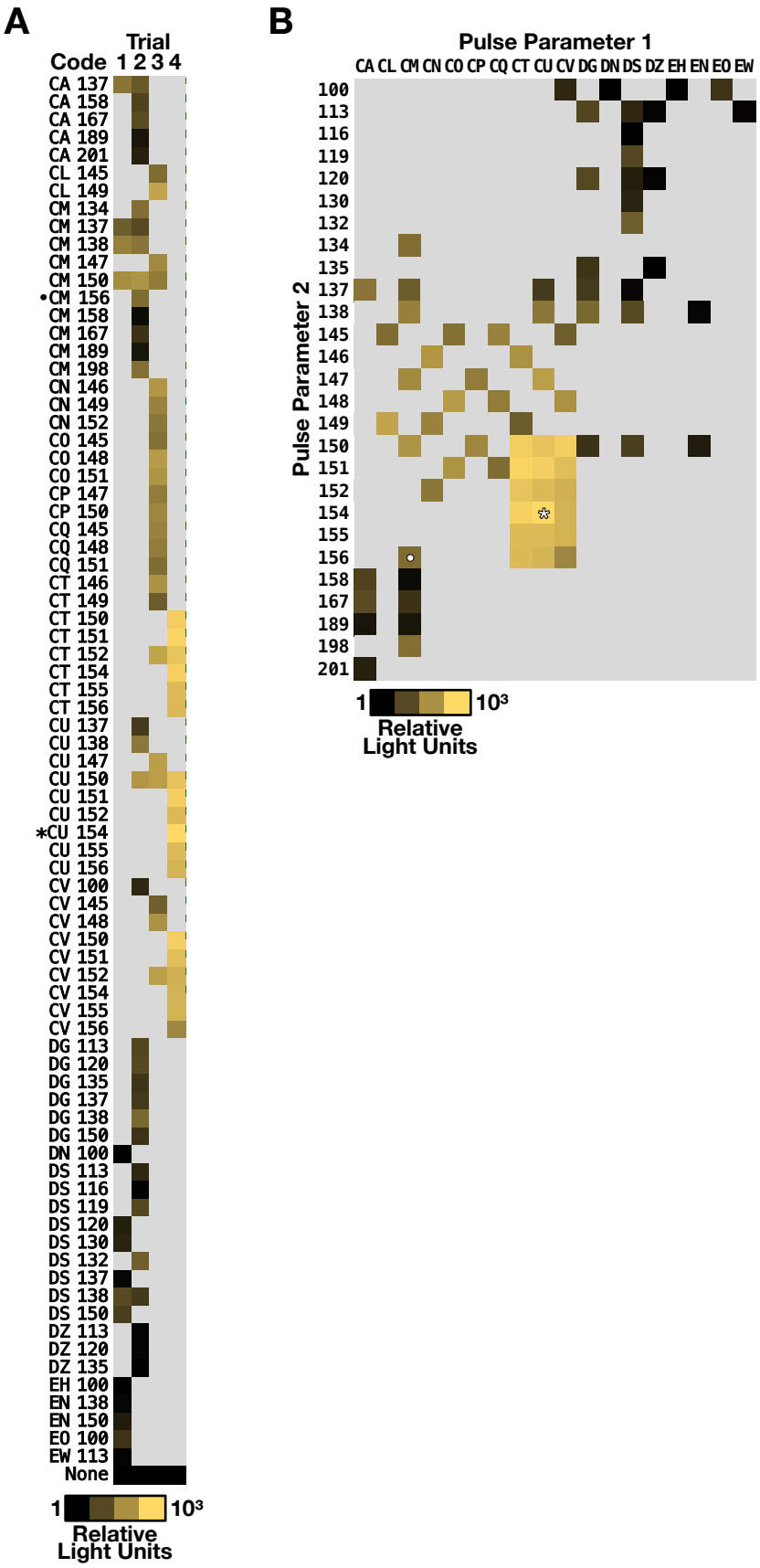


Figure S4: A screen for optimal nucleofection pulses to improve cargo delivery into *S. rosetta*.

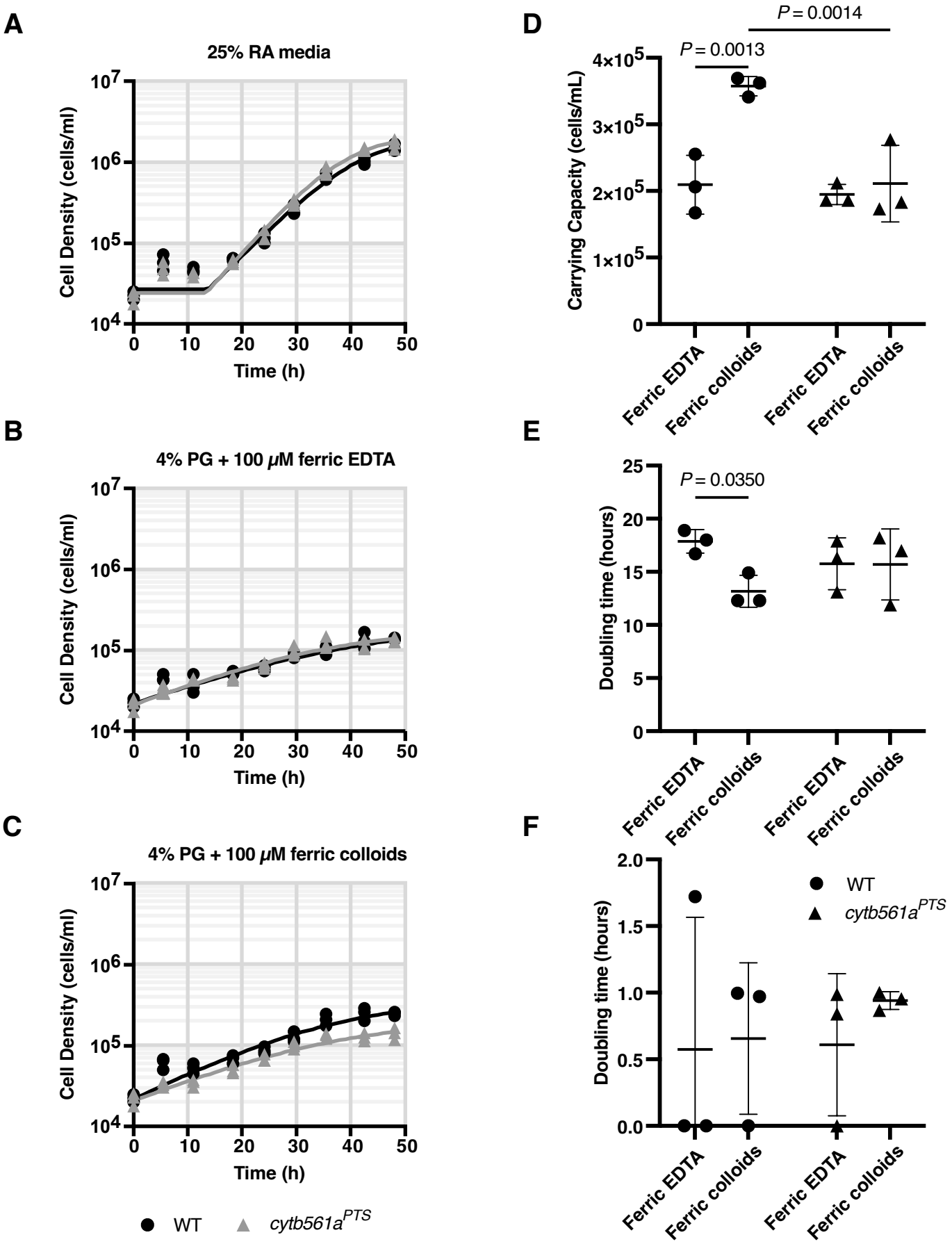
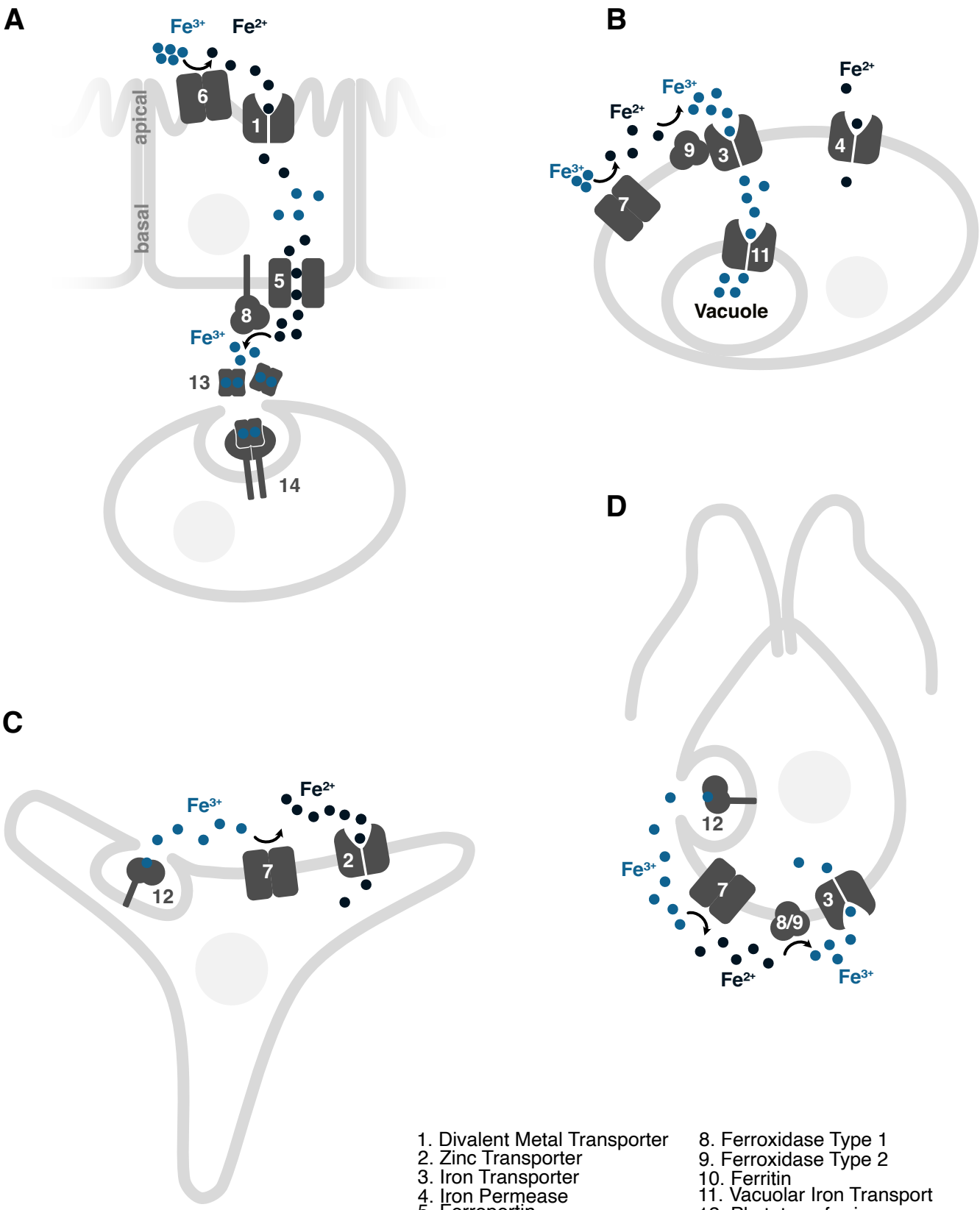


Figure S5: Thecates grow slight faster and to a larger population size with ferric colloids



1. Divalent Metal Transporter
2. Zinc Transporter
3. Iron Transporter
4. Iron Permease
5. Ferroportin
6. Cytochrome B561
7. Cytochrome B245
8. Ferroxidase Type 1
9. Ferroxidase Type 2
10. Ferritin
11. Vacuolar Iron Transport
12. Phytotransferrin
13. Transferrin
14. Transferrin Receptor

Figure S6: Simplified iron acquisition pathways in model eukaryotes.

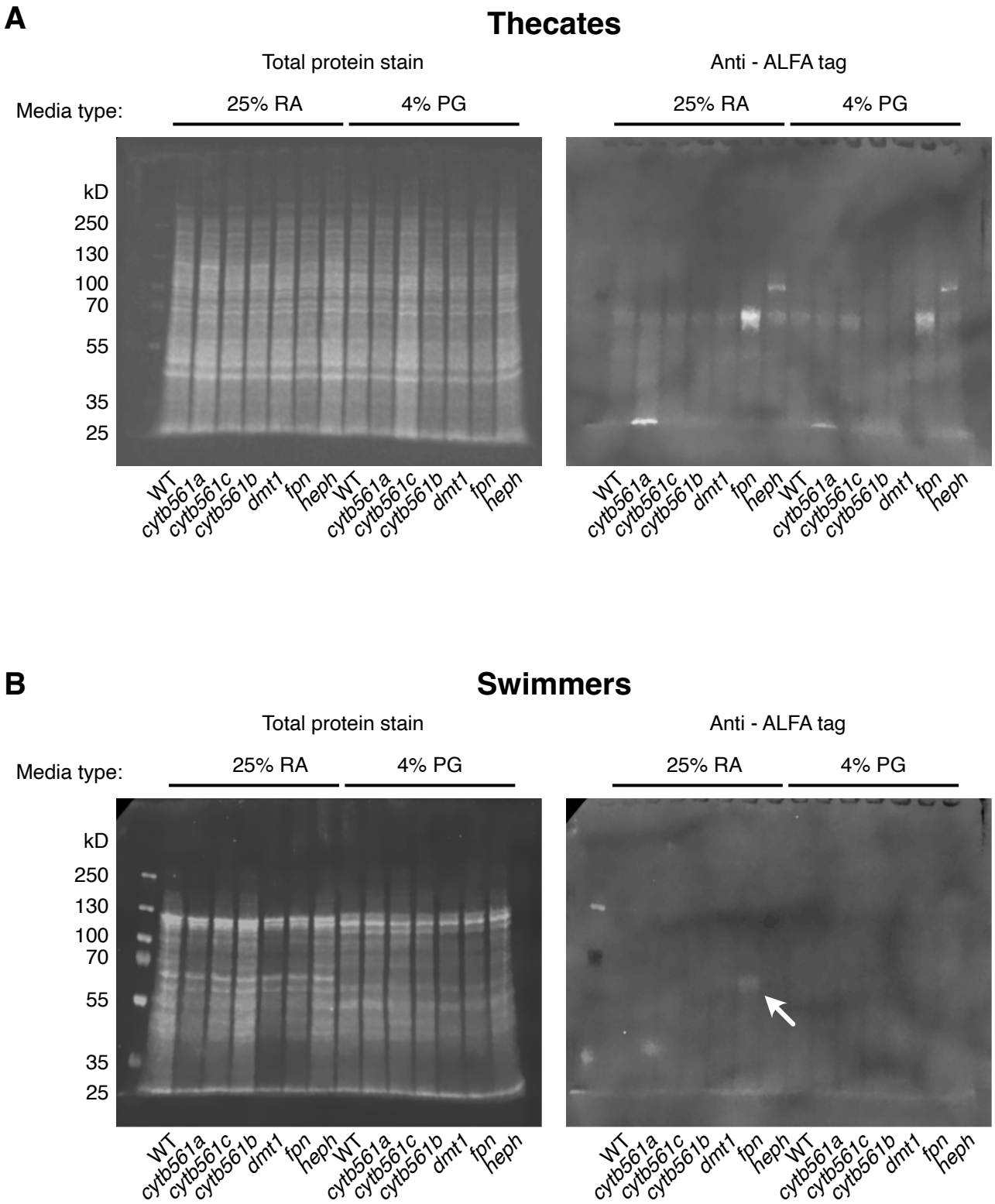


Figure S7: Thecates detectably express Cytb561a, Fpn, and Heph in western blots.

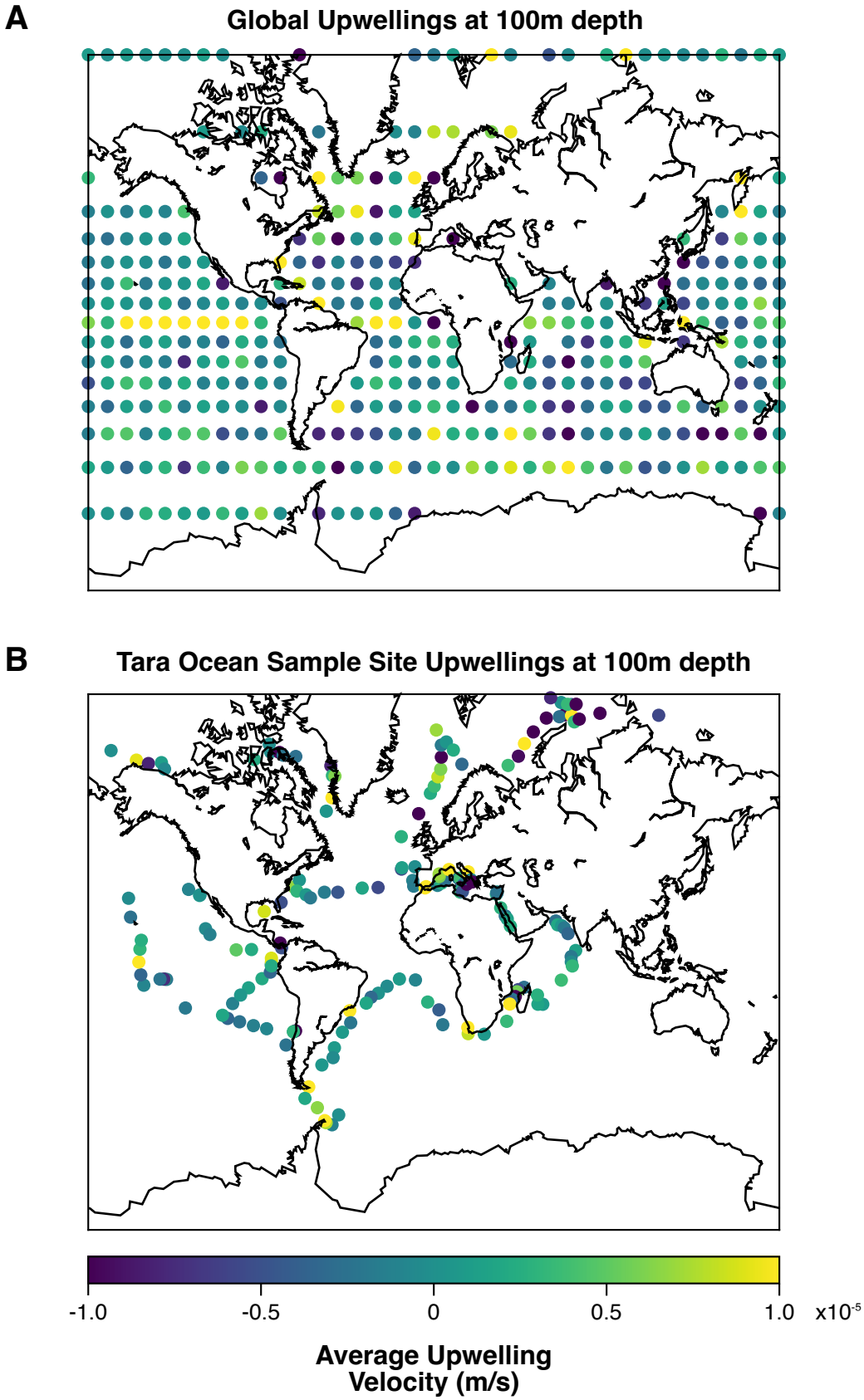


Figure S8: Maps of marine upwelling velocities show increased upwelling near coastlines and the equator

References:

1. Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. *Genome Biol.* **14**, R15 (2013).
2. Levin, T. C. & King, N. Evidence for Sex and Recombination in the Choanoflagellate *Salpingoeca rosetta*. *Curr. Biol.* **23**, 2176–2180 (2013).
3. Levin, T. C., Greaney, A. J., Wetzel, L. & King, N. The Rosetteless gene controls development in the choanoflagellate S. rosetta. *eLife* **3**, (2014).
4. Nedashkovskaya, O. I. *et al.* Echinicola pacifica gen. nov., sp. nov., a novel flexibacterium isolated from the sea urchin Strongylocentrotus intermedius. *Int. J. Syst. Evol. Microbiol.* **56**, 953–958 (2006).
5. Dayel, M. J. *et al.* Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. *Dev. Biol.* **357**, 73–82 (2011).
6. Alegado, R. A. *et al.* A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *eLife* **1**, e00013 (2012).
7. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7894–7899 (2016).
8. Peng, C.-C., Dormanns, N., Regestein, L. & Beemelmanns, C. Isolation of sulfonosphingolipids from the rosette-inducing bacterium Zobellia uliginosa and evaluation of their rosette-inducing activity. *RSC Adv.* **13**, 27520–27524 (2023).
9. Perotti, O., Esparza, G. V. & Booth, D. S. A red algal polysaccharide influences the multicellular development of the choanoflagellate Salpingoeca rosetta. 2024.05.14.594265 Preprint at https://doi.org/10.1101/2024.05.14.594265 (2024).
10. Fairclough, S. R., Dayel, M. J. & King, N. Multicellular development in a choanoflagellate. *Curr. Biol.* **20**, R875–R876 (2010).
11. Richter, D. J., Fozouni, P., Eisen, M. B. & King, N. Gene family innovation, conservation and loss on the animal stem lineage. *eLife* **7**, e34226 (2018).
12. Fumagalli, M. R., Zapperi, S. & Porta, C. A. L. Gene regulatory programs in the life history of Salpingoeca rosetta. 2023.06.12.544615 Preprint at https://doi.org/10.1101/2023.06.12.544615 (2023).
13. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
14. Huntley, R. P. *et al.* The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res.* **43**, D1057–D1063 (2015).
15. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
16. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
17. Gaudet, P. & Dessimoz, C. Gene Ontology: Pitfalls, Biases, and Remedies. in *The Gene Ontology Handbook* (eds. Dessimoz, C. & Škunca, N.) 189–205 (Springer, New York, NY, 2017). doi:10.1007/978-1-4939-3743-1_14.
18. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22 (2022).
19. Sutak, R., Camadro, J.-M. & Lesuisse, E. Iron Uptake Mechanisms in Marine Phytoplankton. *Front. Microbiol.* **11**, (2020).
20. Caza, M. & Kronstad, J. W. Shared and distinct mechanisms of iron acquisition by bacterial and fungal pathogens of humans. *Front. Cell. Infect. Microbiol.* **3**, 80 (2013).
21. Allen, M. D., del Campo, J. A., Kropat, J. & Merchant, S. S. FEA1, FEA2, and FRE1, Encoding Two Homologous Secreted Proteins and a Candidate Ferrireductase, Are Expressed Coordinately with FOX1 and FTR1 in Iron-Deficient Chlamydomonas reinhardtii. *Eukaryot. Cell* **6**, 1841–1852 (2007).

22. Kaplan, J. Mechanisms of Cellular Iron Acquisition: Another Iron in the Fire. *Cell* **111**, 603–606 (2002).
23. Wells, M. L., Mayer, L. M. & Guillard, R. R. L. A chemical method for estimating the availability of iron to phytoplankton in seawater. *Mar. Chem.* **33**, 23–40 (1991).
24. Tagliabue, A. *et al.* Authigenic mineral phases as a driver of the upper-ocean iron cycle. *Nature* **620**, 104–109 (2023).
25. Booth, D. S. & King, N. Genome editing enables reverse genetics of multicellular development in the choanoflagellate Salpingoeca rosetta. *eLife* **9**, e56193 (2020).
26. Ramos-Alonso, L., Romero, A. M., Martínez-Pastor, M. T. & Puig, S. Iron Regulatory Mechanisms in Saccharomyces cerevisiae. *Front. Microbiol.* **11**, (2020).
27. Richter, D. J. *et al.* EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community J.* **2**, (2022).
28. Wunderlich, J. *et al.* Iron transport pathways in the human malaria parasite Plasmodium falciparum revealed by RNA-sequencing. 2024.04.18.590068 Preprint at https://doi.org/10.1101/2024.04.18.590068 (2024).
29. Stearman, R., Yuan, D. S., Yamaguchi-Iwai, Y., Klausner, R. D. & Dancis, A. A Permease-Oxidase Complex Involved in High-Affinity Iron Uptake in Yeast. *Science* **271**, 1552–1557 (1996).
30. Dix, D. R., Bridgham, J. T., Broderius, M. A., Byersdorfer, C. A. & Eide, D. J. The FET4 gene encodes the low affinity Fe(II) transport protein of Saccharomyces cerevisiae. *J. Biol. Chem.* **269**, 26092–26099 (1994).
31. McQuaid, J. B. *et al.* Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* **555**, 534–537 (2018).
32. Pantopoulos, K., Porwal, S. K., Tartakoff, A. & Devireddy, L. Mechanisms of Mammalian Iron Homeostasis. *Biochemistry* **51**, 5705–5724 (2012).
33. Richter, D. J., Fozouni, P., Eisen, M. B. & King, N. Gene family innovation, conservation and loss on the animal stem lineage. *eLife* **7**, e34226 (2018).
34. Bode, H.-P., Dumschat, M., Garotti, S. & Fuhrmann, G. F. Iron Sequestration by the Yeast Vacuole. *Eur. J. Biochem.* **228**, 337–342 (1995).
35. Philpott, C. C. & Protchenko, O. Response to Iron Deprivation in Saccharomyces cerevisiae. *Eukaryot. Cell* **7**, 20–27 (2008).
36. Götzke, H. *et al.* The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nat. Commun.* **10**, 4403 (2019).
37. Laundon, D., Larson, B. T., McDonald, K., King, N. & Burkhardt, P. The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLOS Biol.* **17**, e3000226 (2019).
38. Brunet, T. & Booth, D. S. Chapter One - Cell polarity in the protist-to-animal transition. in *Current Topics in Developmental Biology* (ed. Tepass, U.) vol. 154 1–36 (Academic Press, 2023).
39. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
40. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* **47**, W5–W10 (2019).
41. Torruella, G., Galindo, L. J., Moreira, D. & López-García, P. Phylogenomics of neglected flagellated protists supports a revised eukaryotic tree of life. 2024.05.15.594285 Preprint at https://doi.org/10.1101/2024.05.15.594285 (2024).
42. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
43. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci. Publ. Protein Soc.* **30**, 70–82 (2021).

44. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
45. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
46. Ganasen, M. *et al.* Structural basis for promotion of duodenal iron absorption by enteric ferric reductase with ascorbate. *Commun. Biol.* **1**, 1–12 (2018).
47. Carradec, Q. *et al.* A global ocean atlas of eukaryotic genes. *Nat. Commun.* **9**, 373 (2018).
48. Vernette, C. *et al.* The Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton genes. *Nucleic Acids Res.* **50**, W516–W526 (2022).
49. Marsh, R. & van Sebille, E. Chapter 3 - Surface drift, gyres, and the fate of plastic. in *Ocean Currents* (eds. Marsh, R. & van Sebille, E.) 63–102 (Elsevier, 2021). doi:10.1016/B978-0-12-816059-6.00012-7.
50. Fitzwater, S. E. *et al.* Iron, nutrient and phytoplankton biomass relationships in upwelled waters of the California coastal system. *Cont. Shelf Res.* **23**, 1523–1544 (2003).
51. von der Heyden, B. P. & Roychoudhury, A. N. A review of colloidal iron partitioning and distribution in the open ocean. *Mar. Chem.* **177**, 9–19 (2015).
52. Capone, D. G. & Hutchins, D. A. Microbial biogeochemistry of coastal upwelling regimes in a changing ocean. *Nat. Geosci.* **6**, 711–717 (2013).
53. Behrenfeld, M. J., Bale, A. J., Kolber, Z. S., Aiken, J. & Falkowski, P. G. Confirmation of iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature* **383**, 508–511 (1996).
54. Wells, M. L., Mayer, L. M., Donard, O. F. X., de Souza Sierra, M. M. & Ackelson, S. G. The photolysis of colloidal iron in the oceans. *Nature* **353**, 248–250 (1991).
55. Barbeau, K. & Moffett, J. W. Laboratory and field studies of colloidal iron oxide dissolution as mediated by phagotrophy and photolysis. *Limnol. Oceanogr.* **45**, 827–835 (2000).
56. Bundy, R. M., Jiang, M., Carter, M. & Barbeau, K. A. Iron-Binding Ligands in the Southern California Current System: Mechanistic Studies. *Front. Mar. Sci.* **3**, (2016).
57. Manck, L. E. *et al.* Petrobactin, a siderophore produced by Alteromonas, mediates community iron acquisition in the global ocean. *ISME J.* **16**, 358–369 (2022).
58. Barbeau, K., Moffett, J. W., Caron, D. A., Croot, P. L. & Erdner, D. L. Role of protozoan grazing in relieving iron limitation of phytoplankton. *Nature* **380**, 61–64 (1996).
59. Barbeau, K., Kujawinski, E. B. & Moffett, J. W. Remineralization and recycling of iron, thorium and organic carbon by heterotrophic marine protists in culture. *Aquat. Microb. Ecol.* **24**, 69–81 (2001).
60. Tranvik, L. J., Sherr, E. B. & Sherr, B. F. Uptake and utilization of ‘colloidal DOM’ by heterotrophic flagellates in seawater. *Mar. Ecol. Prog. Ser.* **92**, 301–309 (1993).
61. Gonzalez, J. M. & Suttle, C. A. Grazing by marine nanoflagellates on viruses and virus-sized particles: ingestion and digestion. *Mar. Ecol. Prog. Ser.* **94**, 1–10 (1993).
62. Leadbeater, B. S. C. *The Choanoflagellates: Evolution, Biology, and Ecology*. (Cambridge University Press, Cambridge, United Kingdom, 2015).
63. Wilken, S. *et al.* Choanoflagellates alongside diverse uncultured predatory protists consume the abundant open-ocean cyanobacterium Prochlorococcus. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2302388120 (2023).
64. Worden, A. Z. *et al.* Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**, (2015).
65. Keeling, P. J. Combining morphology, behaviour and genomics to understand the evolution and ecology of microbial eukaryotes. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20190085 (2019).

66. Dupont, C. L., Butcher, A., Valas, R. E., Bourne, P. E. & Caetano-Anollés, G. History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc. Natl. Acad. Sci.* **107**, 10567–10572 (2010).
67. Knoll, A. H. & Nowak, M. A. The timetable of evolution. *Sci. Adv.* **3**, e1603076 (2017).
68. Wade, J., Byrne, D. J., Ballentine, C. J. & Drakesmith, H. Temporal variation of planetary iron as a driver of evolution. *Proc. Natl. Acad. Sci.* **118**, e2109865118 (2021).
69. Sperling, E. A., Knoll, A. H. & Girguis, P. R. The Ecological Physiology of Earth's Second Oxygen Revolution. *Annu. Rev. Ecol. Evol. Syst.* **46**, 215–235 (2015).
70. Mikhailov, K. V. *et al.* The origin of Metazoa: a transition from temporal to spatial cell differentiation. *BioEssays* **31**, 758–768 (2009).
71. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
72. Sebé-Pedrós, A., Degnan, B. M. & Ruiz-Trillo, I. The origin of Metazoa: a unicellular perspective. *Nat. Rev. Genet.* **18**, 498–512 (2017).
73. Sebé-Pedrós, A. *et al.* Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188 (2018).
74. Najle, S. R. & Ruiz-Trillo, I. The Protistan Origins of Animal Cell Differentiation. in *Origin and Evolution of Metazoan Cell Types* (CRC Press, 2021).
75. Márquez-Zacarías, P. *et al.* Evolution of Cellular Differentiation: From Hypotheses to Models. *Trends Ecol. Evol.* **36**, 49–60 (2021).
76. Allen, A. E. *et al.* Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc. Natl. Acad. Sci.* **105**, 10438–10443 (2008).
77. Glaesener, A. G., Merchant, S. S. & Blaby-Haas, C. E. Iron economy in Chlamydomonas reinhardtii. *Front. Plant Sci.* **4**, (2013).
78. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
79. Tanay, A. & Sebé-Pedrós, A. Evolutionary cell type mapping with single-cell genomics. *Trends Genet.* **37**, 919–932 (2021).
80. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
81. Brunet, T. & King, N. The Single-Celled Ancestors of Animals: A History of Hypotheses. (2020) doi:10.20944/preprints202011.0302.v1.
82. Brunet, T. & King, N. The Single-Celled Ancestors of Animals: A History of Hypotheses. in *The Evolution of Multicellularity* (CRC Press, 2022).
83. Knoll, A. H. The Multiple Origins of Complex Multicellularity. *Annu. Rev. Earth Planet. Sci.* **39**, 217–239 (2011).
84. Richter, D. J. & King, N. The Genomic and Cellular Foundations of Animal Origins. *Annu. Rev. Genet.* **47**, 509–537 (2013).
85. Brunet, T. & King, N. The Origin of Animal Multicellularity and Cell Differentiation. *Dev. Cell* **43**, 124–140 (2017).
86. Booth, D. S., Szmidt-Middleton, H. & King, N. Transfection of choanoflagellates illuminates their cell biology and the ancestry of animal septins. *Mol. Biol. Cell* **29**, 3026–3038 (2018).
87. Perotti, O., Viramontes, G. & Booth, D. A red algal polysaccharide influences the multicellular development of the choanoflagellate S. rosetta. (2024).
88. Woznica, A. *et al.* STING mediates immune responses in the closest living relatives of animals. *eLife* **10**, e70436 (2021).
89. Coyle, M. C. *et al.* An RFX transcription factor regulates ciliogenesis in the closest living relatives of animals. *Curr. Biol.* **33**, 3747-3758.e9 (2023).
90. Brunet, T. *et al.* A flagellate-to-amoeboid switch in the closest living relatives of animals. *eLife* **10**, e61037 (2021).
91. Rutaganira, F. U. N., Scopton, A. P., Dar, A. C. & King, N. A small molecule screen

reveals the essentiality of p38 kinase for choanoflagellate cell proliferation. 2022.08.26.505350 Preprint at https://doi.org/10.1101/2022.08.26.505350 (2022).

92. Harayama, T. & Riezman, H. Understanding the diversity of membrane lipid composition. *Nat. Rev. Mol. Cell Biol.* **19**, 281–296 (2018).

93. Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol. Rev.* **40**, 133–159 (2016).

94. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

95. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2017).

96. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).

97. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).

98. Tabata, R. *et al.* Systemic Regulation of Iron Acquisition by Arabidopsis in Environments with Heterogeneous Iron Distributions. *Plant Cell Physiol.* **63**, 842–854 (2022).

99. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).

100. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).

101. Richter, D. J. *et al.* EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. 2020.06.30.180687 Preprint at https://doi.org/10.1101/2020.06.30.180687 (2022).

102. Steenwyk, J. L., Iii, T. J. B., Li, Y., Shen, X.-X. & Rokas, A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biol.* **18**, e3001007 (2020).

103. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

104. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

105. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* gkae268 (2024) doi:10.1093/nar/gkae268.

106. Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2**, 150023 (2015).