

1 **Title:**

2

3 **Cell differentiation controls iron assimilation in a choanoflagellate**

4

5 **Authors:**

6

7 Fredrick Leon¹, Jesus M. Espinoza-Esparza¹, Vicki Deng^{1,3}, Maxwell C. Coyle^{2,4}, Sarah
8 Espinoza², David S. Booth^{1*}

9

10 **Affiliations:**

11

12 ¹Chan Zuckerberg Biohub &
13 Department of Biochemistry and Biophysics
14 University of California, San Francisco School of Medicine
15 San Francisco, CA 94143

16

17 ²Howard Hughes Medical Institute &
18 Department of Molecular and Cell Biology
19 University of California, Berkeley
20 Berkeley, CA 94720

21

22 ³Current Address: Department of Molecular Biosciences
23 University of Texas, Austin
24 Austin, TX 78712

25

26 ⁴Current Address: Department of Molecular and Cellular Biology
27 Harvard University
28 Cambridge, MA 02138

29

30 *Corresponding author: David.Booth@ucsf.edu

1 Supplemental Figures:

2

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

4 **(A)** A schematic showing the separation of *S. rosetta* RNA from *E. pacifica* RNA.

5 **(B)** The preferential lysis procedure enriches RNA from *S. rosetta*. After performing preferential

6 lysis, RNA was purified from the supernatant using Trizol LS reagent. The RNA from the pellet

7 of preferential lysis, an unlysed sample of *S. rosetta* feeding on *E. pacifica*, or a sample of *E.*

8 *pacifica* was purified with Trizol LS. The purified RNA samples were combined with an equal

9 volume of formamide and incubated at 95°C for 5 minutes. After cooling to room temperature,

10 the samples were run on a 2% (w/v) agarose gel in TBE buffer at 100 V. Afterwards, the gel was

11 stained with SYBR Gold to detect RNA. In the gel, the supernatant (S) from whole lysates (W)

12 mostly contains *S. rosetta* rRNA (28S and 18S); the pellet (P) contains *E. pacifica* rRNA (23S

13 and 16S).

14 **(C)** Total RNA prepared for RNA-seq primarily contains rRNA from *S. rosetta*. Bioanalyzer

15 traces with predicted prokaryotic and eukaryotic rRNA subunit sizes indicated. Replicate RNA

16 samples used for sequencing demonstrate little to no carryover from bacterial contamination, as

17 shown by the low amount of 16S and 23S rRNA.

18

19 **Figure S2:** GO molecular functions and biological processes enriched in thecates.

20 Functional modules enriched in thecates mediate signal transduction, nutrient acquisition, and

21 gene regulation. Genes with transcript abundances that reliably ($q < 0.01$) changed more than

22 two-fold in thecates were analyzed for the enrichment of Gene Ontology (GO) categories,

23 focusing on GO molecular functions **(A)** and biological processes **(B)**. Genes that were

24 upregulated in thecates and associated with an enriched GO molecular function category were

25 counted (x-axis). Each bar on the graph is colored by the *P*-value, which reflects the probability

26 that a given category was represented by a random selection of genes from the *S. rosetta*

27 genome.

28

29 **Figure S3:** GO molecular functions and biological processes enriched in swimmers.

30 Functional modules enriched in swimmers support RNA processing, mitosis, and metabolism.

31 Genes with transcript abundances that reliably ($q < 0.01$) changed more than two-fold in

32 swimmers were analyzed for the enrichment of Gene Ontology (GO) categories, focusing on

33 GO molecular functions **(A)** and biological processes **(B)**. Genes that were upregulated in

34 swimmers and associated with an enriched GO molecular function category were counted

1 (x-axis). Each bar on the graph is colored by the *P*-value, which reflects the probability that a
2 given category was represented by a random selection of genes from the *S. rosetta* genome.

3

4 **Figure S4:** Validated qPCR primers confirm the disruption of *cytb561a* expression with the
5 *cytb561a*^{PTS} allele.

6 **(A)** qPCR primers for *cytb561a* and *cofilin* produce a single amplicon. Gel image of PCR
7 products amplified from qPCR primers designed for *cytb561a* and *cofilin*. Note the diminished
8 band from swimmer cDNA likely resulted from lower mRNA copies from slow swimmer cultures,
9 as shown in Fig. 2C. Shadow at 75 bp is a shadow of the loading dye.

10 **(B)** Standard curve for qPCR primers for *cytb561a*. To generate a standard curve, a serial
11 dilution of ssDNA standards was performed in triplicate. The concentration of the ssDNA
12 standard was determined by Qubit. Cycle thresholds (Ct) were converted to transcript copy
13 number by fitting a linear equation to the standard curve. This curve was used to determine the
14 abundance of transcripts from three independent experiments.

15 **(C)** Standard curve for qPCR primers for *cofilin*. The graph is laid out as for panel B.

16 **(D)** *Cytb561a*^{PTS} exhibits lowered *cytb561a* expression compared to wild-type thecates. We
17 compared the normalized expression of *cytb561a* from the strain bearing the *cytb561a*^{PTS} allele
18 to wild-type strains. The wild-type strains were slow swimmers cultured with ferric EDTA or
19 thecates cultured with ferric colloids, for we anticipated that these two conditions would result in
20 the greatest difference in *cytb561a* expression.

21

22 **Figure S5:** *S. rosetta* cell types, but not their feeder bacteria, grow differently with variable iron
23 sources.

24 **(A)** Thecate cell types exhibit differential growth with low concentrations of ferric colloids
25 compared to slow swimming cell types. *S. rosetta* cell density of slow swimmer and thecate
26 cultures when grown with titrations of ferric EDTA or ferric colloids.

27 **(B)** A variety of ferric chelates differentially impact *S. rosetta* growth. Slow swimmers and
28 thecates exhibit different growth characteristics, so to better compare their growth with iron
29 bound to different chelators, we used a ratio of the final cell density of cultures grown with
30 iron-chelator complexes to the final cell density of cultures grown without any supplemental iron.
31 With this metric, a ratio greater than one indicates that the cell type displays increased growth
32 with the iron-chelator complex; whereas, a ratio less than one indicates the converse. All
33 conditions were tested at 100 μM Fe³⁺, and pyoverdines and deferoxamine are bacterial
34 siderophores.

1 **(C)** *E. pacifica* feeder bacteria growth is easily rescued by low concentrations of iron. Optical
2 density at 600 nm (OD600) of *E. pacifica* measured approximately every 100 minutes for 48
3 hours. Note standard deviations overlap for all conditions except for a no iron control.
4 **(D)** *E. pacifica* grows similarly with ferric EDTA or ferric colloids. Because ferric colloids would
5 influence OD600 measurements in a standard growth curve experiment, *E. pacifica* cultures
6 were grown for 48 hours with 100 μM ferric EDTA or ferric colloids and then lightly centrifuged at
7 500xg for 2 minutes at room temperature to settle iron particulates. Afterwards, the supernatant
8 OD600 was measured.
9 **(E)** Ferric colloids release a small amount of labile iron. When media is prepared with 100 μM
10 ferric colloids, only $3.43 \pm 1.01 \mu\text{M}$ iron is liberated without biological activity. The amount of iron
11 in 4% PG media alone was $0.27 \pm 0.036 \mu\text{M}$ iron.

12

13 **Figure S6:** A screen for optimal nucleofection pulses to improve cargo delivery into *S. rosetta*.

14 **(A)** Nanoluc luminescence values from *S. rosetta* cultures transfected with different pulse codes.
15 Because of changes to the growth media for *S. rosetta*, we conducted a new screen for an
16 optimal nucleofection pulse. Cells were transfected with a plasmid that drives nanoluc
17 expression with a promoter cloned from elongation factor L. Pulse codes correspond to Lonza
18 4D-Nucleofector codes. The screen was performed in four independent trials. Dot indicates prior
19 pulse code used, asterisk denotes optimized pulse code.

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

27

28 **Figure S7:** Genotypes of engineered strains.

29 For each strain that was engineered, the targeted locus (unique identifiers and allele names
30 shown on the left) was amplified by PCR and then sequenced with the Sanger method. The
31 sequencing traces for locus are displayed with the predicted and confirmed sequences. Two
32 strains (*heph^{alfa}* and *cytb561c^{alfa}*) were heterozygous with a wild type and edited allele, as seen
33 in superimposed peaks at each nucleotide position. The sequences of heterozygotes were
34 further analyzed by DECODR¹¹² to confirm the presence of both sequences.

1

2 **Figure S8:** Thecates grow slight faster and to a larger population size with ferric colloids
3 **(A-C)** Growth curves of wild type and *cytb561a^{PTS}* thecates show wild type thecates grow to a
4 larger population size when grown with ferric colloids. Growth curves of cultures grown in (A)
5 nutrient replete conditions (25% RA), (B) 4% PG with ferric EDTA, and (C) 4% PG with ferric
6 colloids. Cell densities were counted every 6 hours for 48 hours.
7 **(D-E)** Comparisons of growth curve parameters highlight the increase in carrying capacity as
8 the main factor for higher wild type thecate cell type growth with ferric colloids. (D) Carrying
9 capacity, and (E) doubling time of the growth curves shown in figures (B) and (C).

10

11 **Figure S9:** Thecates phagocytose ferric colloids.

12 **(A)** Fluorescently labeled ferric colloids were embedded with fluorescent dextrans. During the
13 precipitation of FeCl_3 to make ferric colloids, fluorescently labeled dextrans were added. The
14 resulting colloids were incubated with dextranase, washed, and then resuspended in ASW
15 before adding to thecate cultures for time-course microscopy.

16 **(B)** Embedded fluorescent dextrans highlight particles of ferric colloids. The images show
17 colloids of similar size and density, with only the co-precipitated ferric colloids and dextran
18 exhibiting fluorescence.

19 **(C)** Corroborating examples of ferric colloid ingestion. Time courses of thecates feeding on ferric
20 colloid particles (magenta) are displayed in the fluorescence channel (below) and in a merge
21 with brightfield and fluorescence channel. 0 seconds denotes the initial point of contact between
22 the cell and the tracked colloid particle.

23

24 **Figure S10:** Simplified iron acquisition pathways in model eukaryotes.

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

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

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

3 **(D)** *Chlamydomonas reinhardtii*. Ferric cations are first reduced, then oxidized, and finally
4 imported in its ferric form for increased substrate specificity. Iron can also be directly bound by
5 phytoferritin, which is thought to be released into the periplasm.

6

7 **Figure S11:** Phylogenetic trees of iron acquisition proteins

8 Unrooted maximum likelihood phylogenies of the iron acquisition proteins shown in Fig. 3A.

9 Alignment, trimming, and tree building details are listed in table S3.

10

11 **Figure S12:** Animal transferrin receptors evolved from an ancestral M28 peptidase.

12 **(A)** Animal transferrin receptors emerged from ancestral animal M28 peptidase. Maximum likelihood
13 phylogeny built from a search using a new HMM model of the domain architecture shown in (B).

14 Animal transferrin receptors are highlighted by a gray outline, which was determined by the loss of
15 peptidase enzymatic residues shown in (C) and (D).

16 **(B)** The transferrin receptor protein domain architecture. Domain labels are for the protease
17 associated (PA) domain, M28 peptidase (M28 Pept.) domain, and the transferrin receptor specific
18 dimerization (Dimer) domain. Pfam accession IDs are listed below the associated domain.

19 **(C)** Transferrin - transferrin receptor complex with highlighted catalytic residues. The structure of
20 transferrin - transferrin receptor complex as determined by cryo-electron microscopy (PDB 1SUV).
21 The catalytic residues of the transferrin receptor are highlighted and colored.

22 **(D)** The loss of key catalytic residues marks the evolutionary transition from peptidase to transferrin.
23 Alignments of transferrin receptors across animals show a remarkable loss of key M28 peptidase
24 catalytic residues. This loss likely demarks the transition from peptidase to transferrin receptor, yet
25 the binding of transferrin outside the catalytic site may have allow some related proteins to have dual
26 functions as peptidases and transferrin receptors.

27

28 **Figure S13:** Thecates detectably express Cytb561a, Fpn, and Heph in western blots.

29 **(A)** Western blot of thecates cultures of ALFA-tagged strains, grown in nutrient rich media (25%
30 RA) and iron depleted media (4% PG). Strains are named after the ALFA-tagged gene.

31 **(B)** Western blot of swimmer cultures of alfa-tagged strains, grown in nutrient rich media (25%
32 RA) and iron depleted media (4% PG). Arrow indicates faint Fpn band visible only in the 25%
33 RA condition. The expression of Fpn in media with high iron concentrations and the lack of
34 expression in low iron media is consistent with Fpn regulating homeostatic iron levels through
35 post-transcriptional regulation.

1

2 **Figure S14:** Thecates stably express Cytb561a independent of iron availability.

3 Western blot of *cytb561a*^{ALFA} thecates and swimmers grown in nutrient rich media (25% RA) and
4 iron depleted media (4% PG) show minimal change in response to iron availability, indicating
5 little to no post-transcriptional regulation of Cytb561a.

6

7 **Figure S15:** Additional immunofluorescence of thecates and swimmers

8 **(A)** The localization of Cytb561a, Fpn, and Heph in thecates, visualized through
9 immunofluorescence of ALFA tagged proteins. Images are not deconvolved, and only processed in
10 FIJI according to the methods.

11 **(B)** The localization of Fpn and Heph in swimmers, visualized through immunofluorescence of ALFA
12 tagged proteins. Note, Cytb561a is not detectable in swimmers, matching the protein levels seen in
13 western blots (Fig. S13) and by transcript levels shown by RNAseq and qPCR (Fig. 2B-C). Images
14 are not deconvolved, and only processed in FIJI according to the methods.

15

16 **Figure S16:** Cytochrome b561 paralogs have retained ascorbate binding residues.

17 **(A-B)** The ascorbate binding pocket of human cytochrome b561 protein DCYTB. The crystal
18 structure (PDB 5ZLG) of human duodenal cytochrome b561 (DCYTB) features a positively charged
19 binding pocket on the cytosolic face of each polypeptide of the homodimeric complex **(A)**. Each
20 pocket binds one ascorbate molecule (orange) that is coordinated by specific electrostatic and
21 hydrophobic contacts **(B, cyan residues)**

22 **(C)** Cytb561 homologs have retained ascorbate binding residues. An alignment of cytochrome B561
23 protein sequences shows conservation in positions that correspond to ascorbate-binding residues in
24 human DCYTB.

25

26 **Figure S17:** Cytochrome b561 dimerization is a feature of Group A proteins.

27 **(A-B)** The crystal structure of human DCYTB (PDB 5ZLG) forms a homodimeric complex **(A)**.
28 Residues that mediate homodimerization **(B, pink residues)**. Heme groups are depicted in black
29 with coordinated iron atoms in red.

30 **(C)** Cytochrome b561 homologs from Group A, including human DCYTB and *S. rosetta* Cytb561a,
31 have key residues for dimerization, while groups B and C do not. An alignment of cytochrome b561
32 protein sequences reveals that Group A homologs conserve similar residues at positions that
33 correspond to the dimerization residues in DCYTB. Sequences in Group B do not align at ~7
34 positions corresponding to the dimeric interface, and Group C sequences do not align at ~10
35 positions.

1

2 **Figure S18:** Cytb561a is predicted to dimerize similarly to human DCYTB.

3 **(A)** Expected position errors of dimerization predictions indicate that Cytb561a robustly forms a
4 dimeric complex. Dimeric structures of each Cytb561 paralog bound to heme cofactors were
5 predicted with AlphaFold v3. The expected position error graphs from those predictions depict
6 the distance between the same residue across predicted dimer models. The smaller the
7 expected distance between residues indicates greater consistency between different models
8 generated during the prediction.

9 **(B)** Predicted dimers of Cytb561a consistently align with human DCYTB dimers. The top three
10 predicted models of Cytb561 paralogs from *S. rosetta* (grey structures) were aligned to the
11 crystal structure of human DCYTB. Alignments were restricted to only one chain from each dimeric
12 complex, allowing the second chain to be positioned based on the predicted structure. To better
13 visualize the position of each dimer, the amino-terminal residue is colored in green and the
14 carboxy-terminal residue in magenta. In these alignments, all three models of Cytb561a from *S.*
15 *rosetta* consistently align with the DCYTB dimer. In contrast, the second polypeptide chain from
16 Cytb561b and Cytb561c display variable positions across all three models.

17

18 **Figure S19:** The luminal surfaces of Cytochrome b561 feature electrostatic properties that are
19 general features of Cytochrome b561 subgroups.

20 **(A)** The luminal surface of the DCYTB homodimer (PDB 5ZLG) has negatively and positively
21 charged patches that bind iron by itself or iron complexed with ligands.

22 **(B)** Loops (yellow) that bridge transmembrane helices form the luminal surface of DCYTB.

23 **(C)** Cytb561 subgroups have distinctive electrostatic properties on their luminal surfaces. The
24 sequences of loops that comprise the luminal surface of Cytochrome b561 were concatenated to
25 calculate the isoelectric point of the luminal surface for each paralog. Individual paralogs from each
26 subgroup are displayed as black dots, and the distribution of charges is shown for each subgroup as
27 a grey violin plot. A Kolmogorov-Smirnov test compared the distributions of isoelectric points
28 between Cytochrome b561 subgroups.

29 **(D)** A character map of luminal surface charge suggests that electrostatic properties can evolve
30 within Cytochrome b561 subgroups. A maximum likelihood phylogeny of cytochrome B561 proteins
31 (as in Fig. 4A) with the isoelectric points of the luminal surface for each protein shows that surface
32 properties are variable even in well-resolved clades.

33

34 **Figure S20:** Maps of marine upwelling velocities show increased upwelling near coastlines and
35 the equator.

1 **(A)** Global upwelling velocities at 100m depth averaged over a three year span, measured every
2 10° latitude and longitude. Missing values are due to shallow locations or landmasses.

3 **(B)** Upwelling velocities at 100m depth averaged over a three year span at every location
4 sampled during the Tara Oceans Expeditions.

5

6

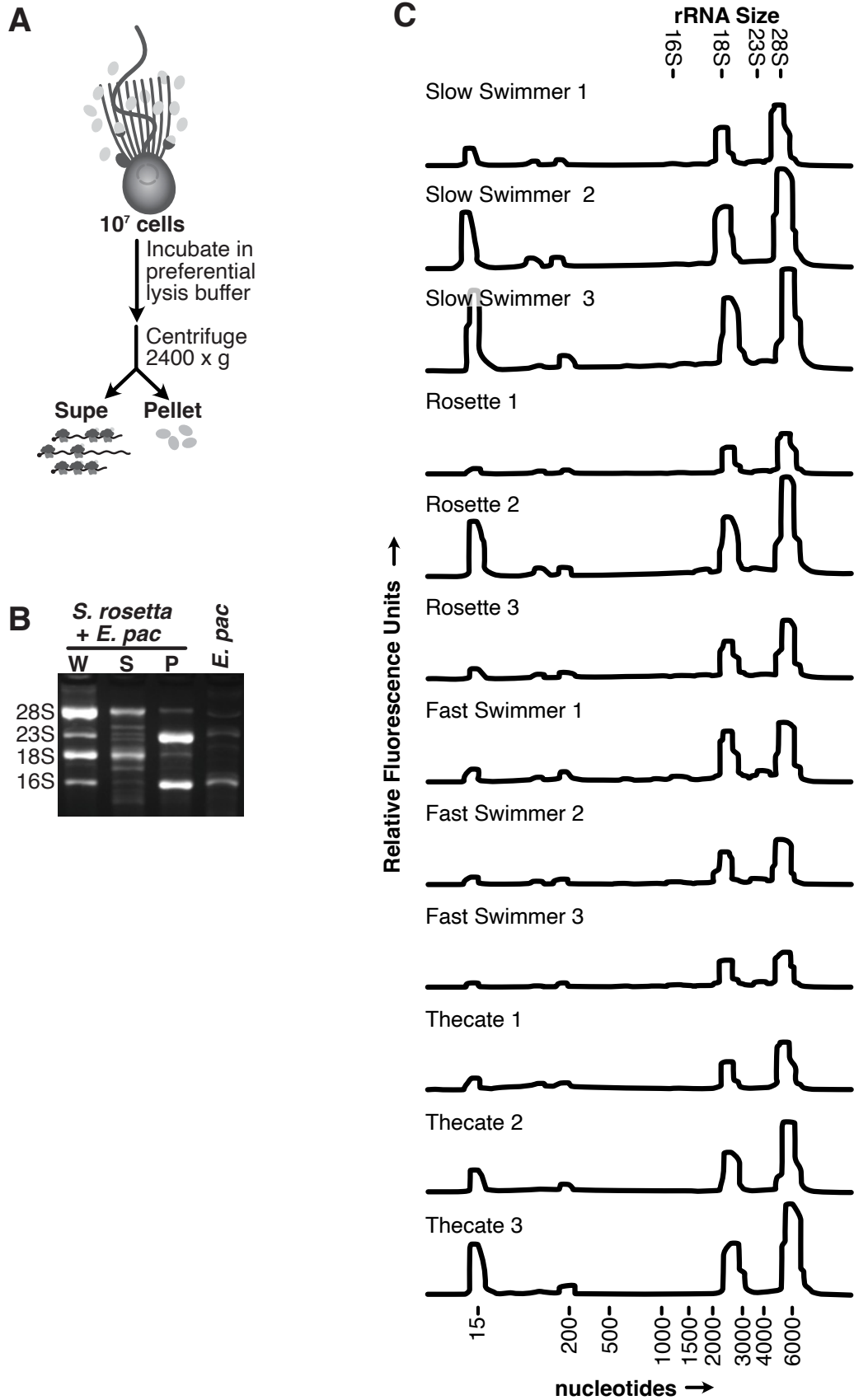


Figure S1. A preferential lysis buffer enriches RNA from *S. rosetta* to improve transcriptomes.

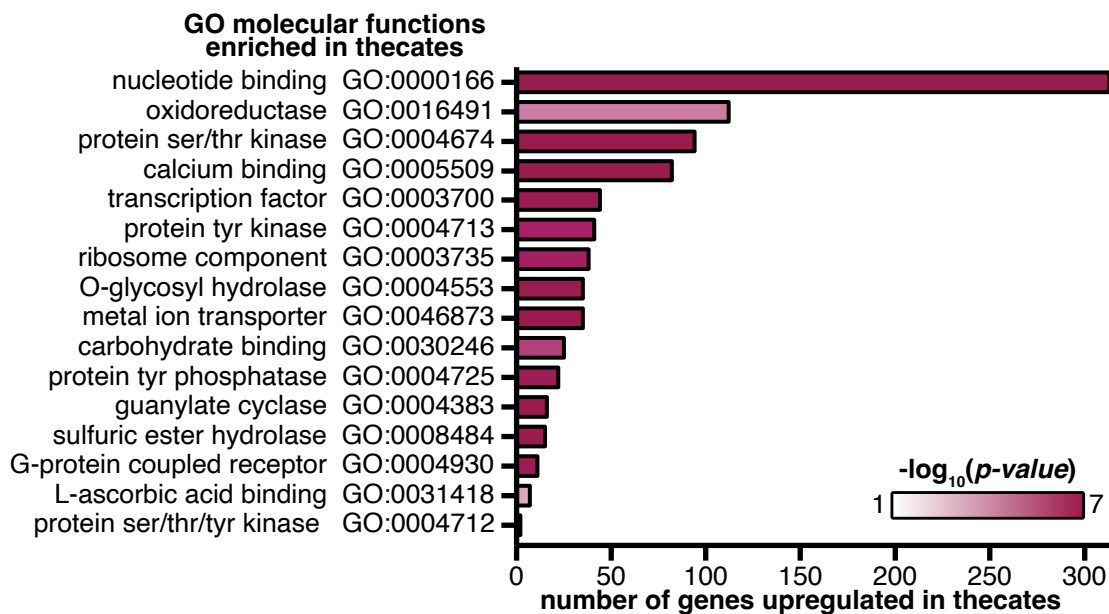
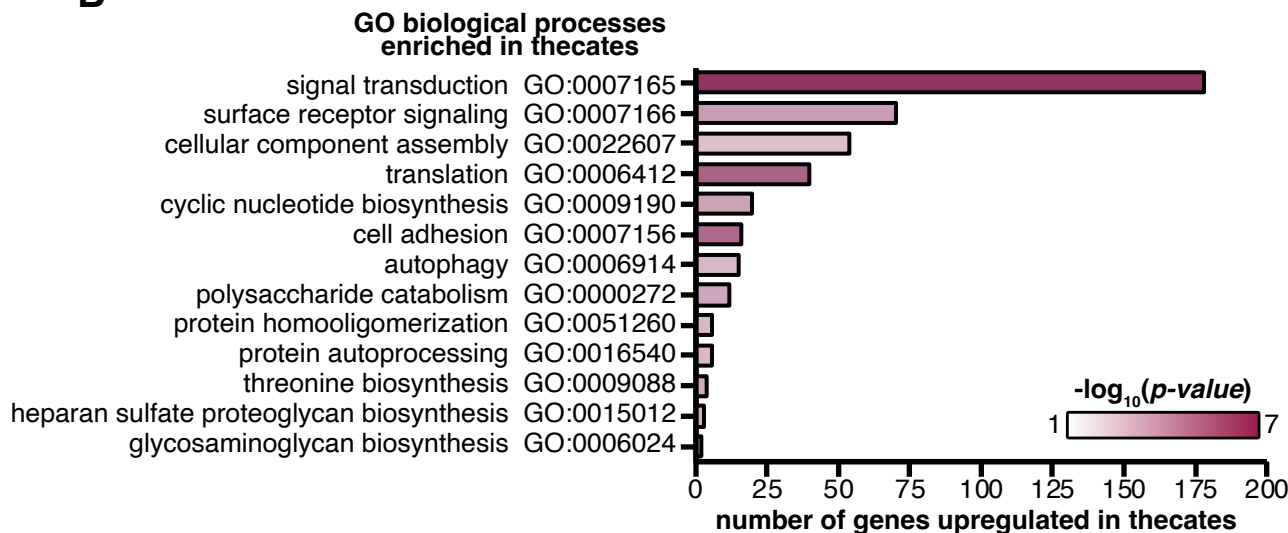
A**B**

Figure S2: GO molecular functions and biological processes enriched in thecates.

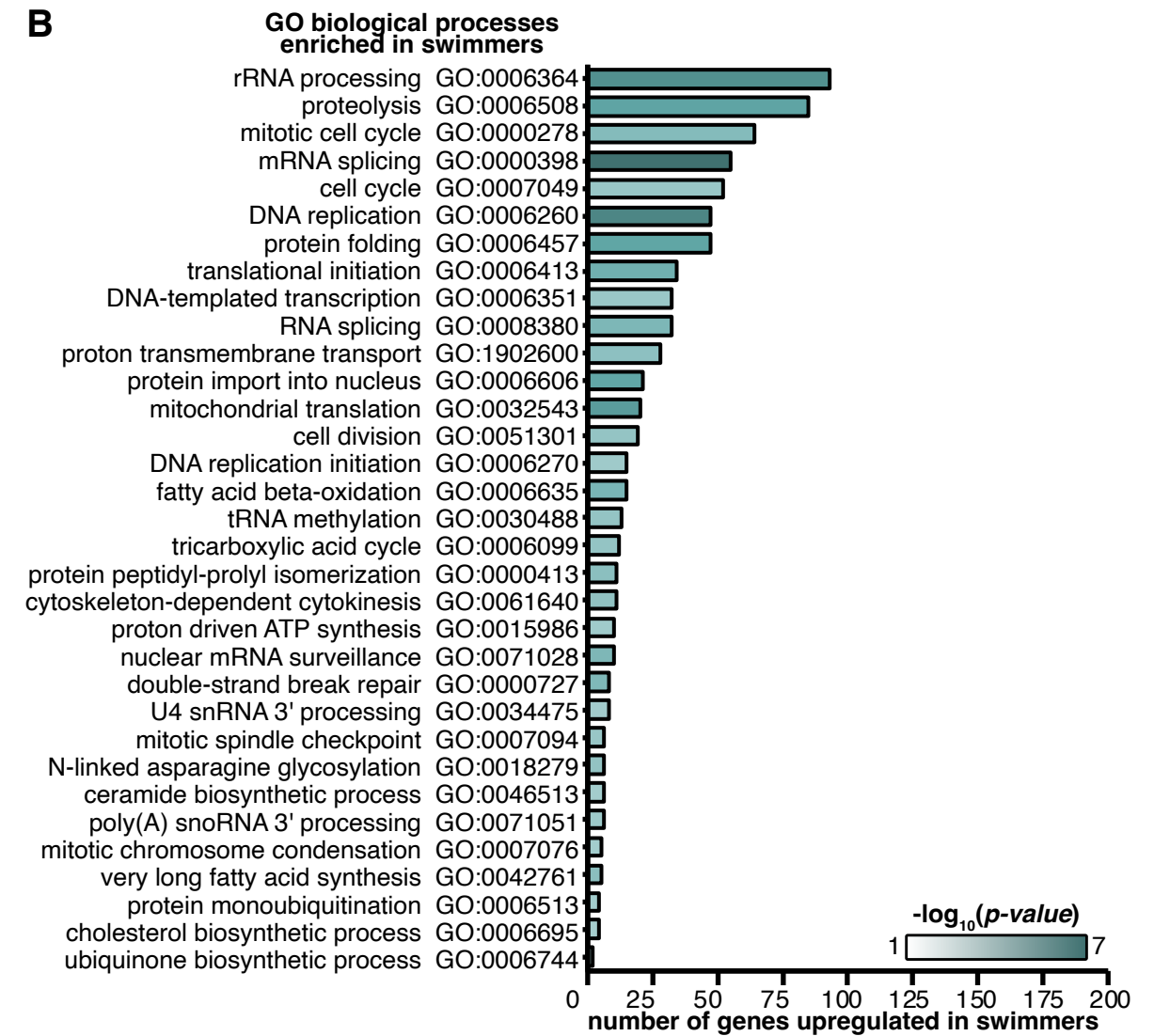
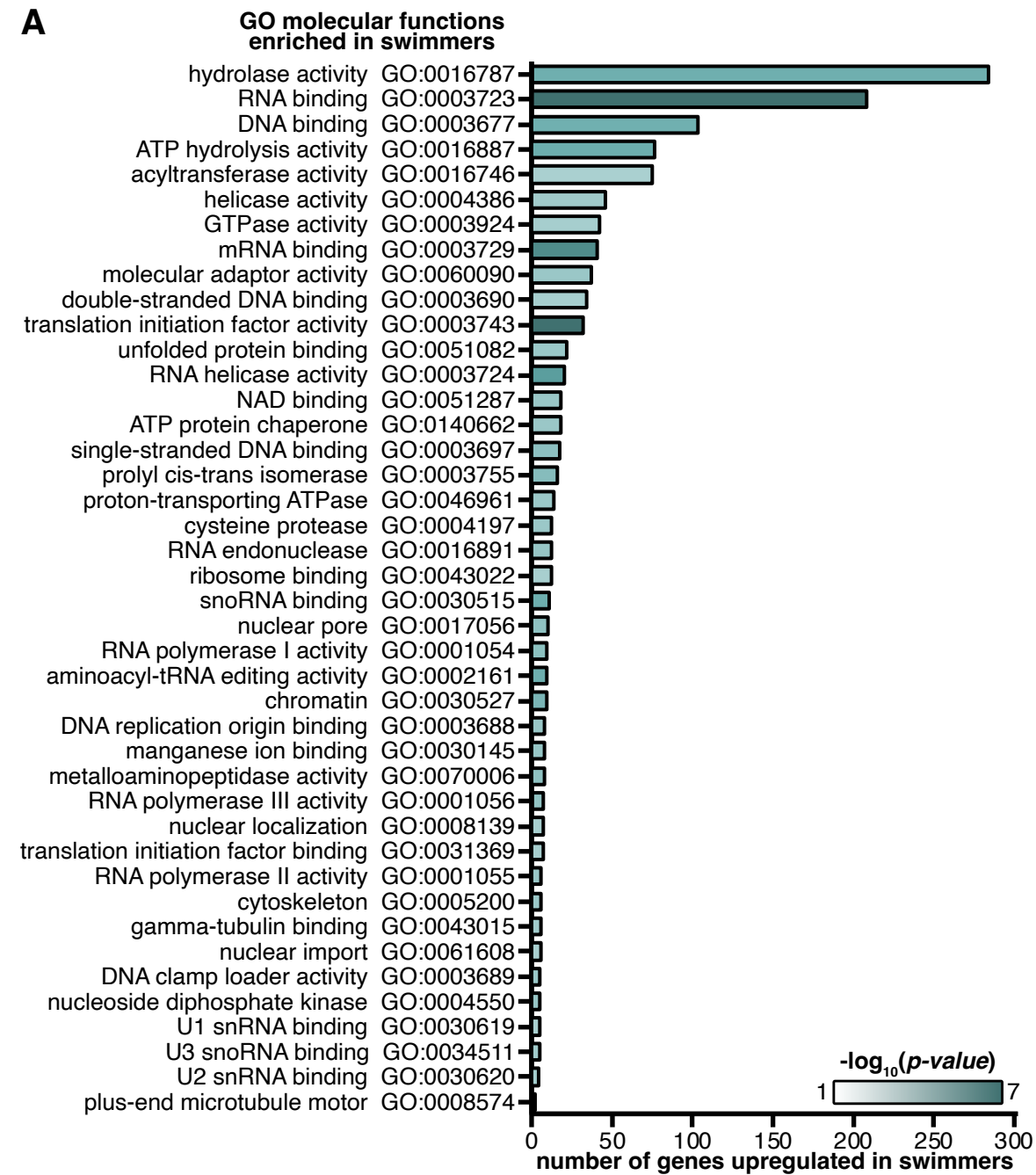


Figure S3: GO molecular functions and biological processes enriched in swimmers.

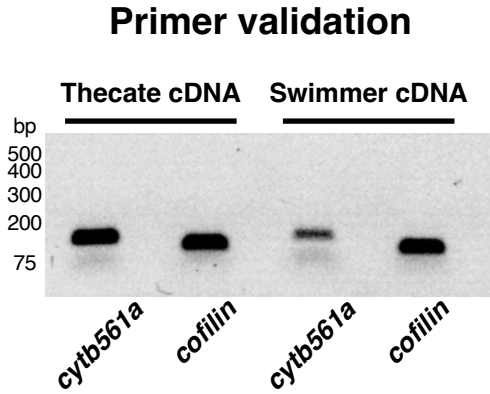
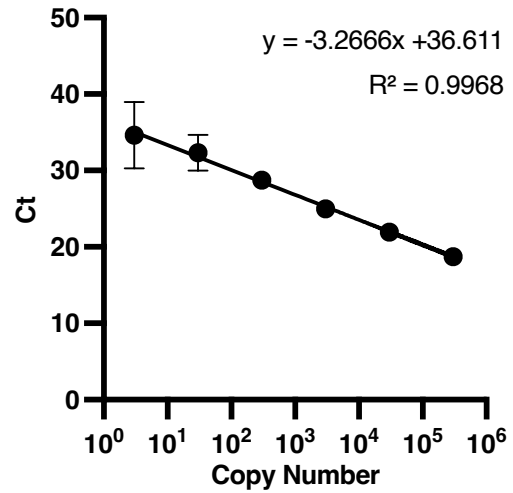
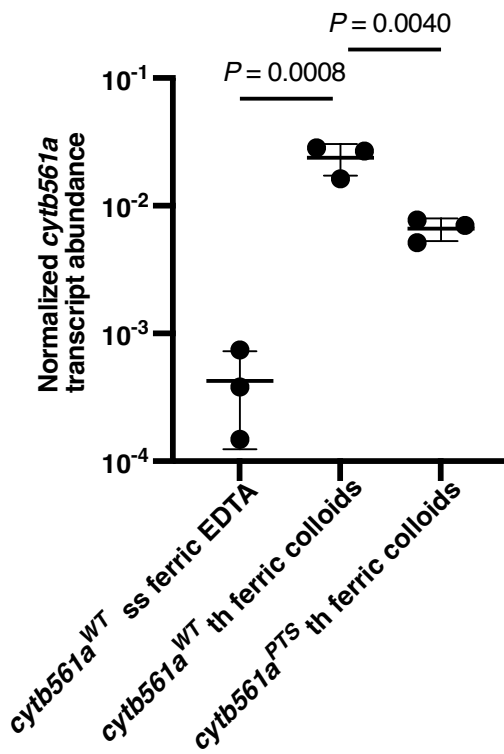
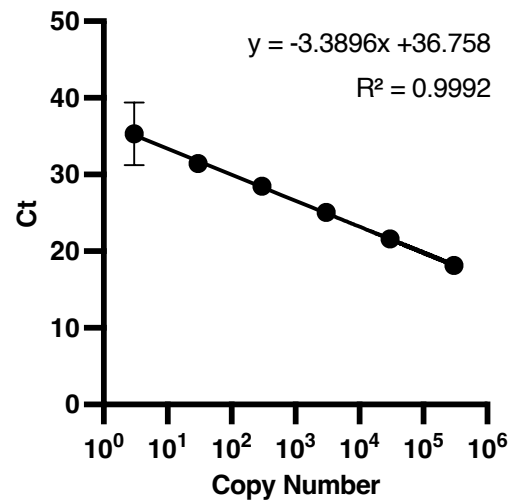
A**B*****cytb561a* qPCR standard curve****D****C*****cofilin* qPCR standard curve**

Figure S4: Validated qPCR primers confirm the disruption of *cytb561a* expression with the *cytb561a*^{PTS} allele.

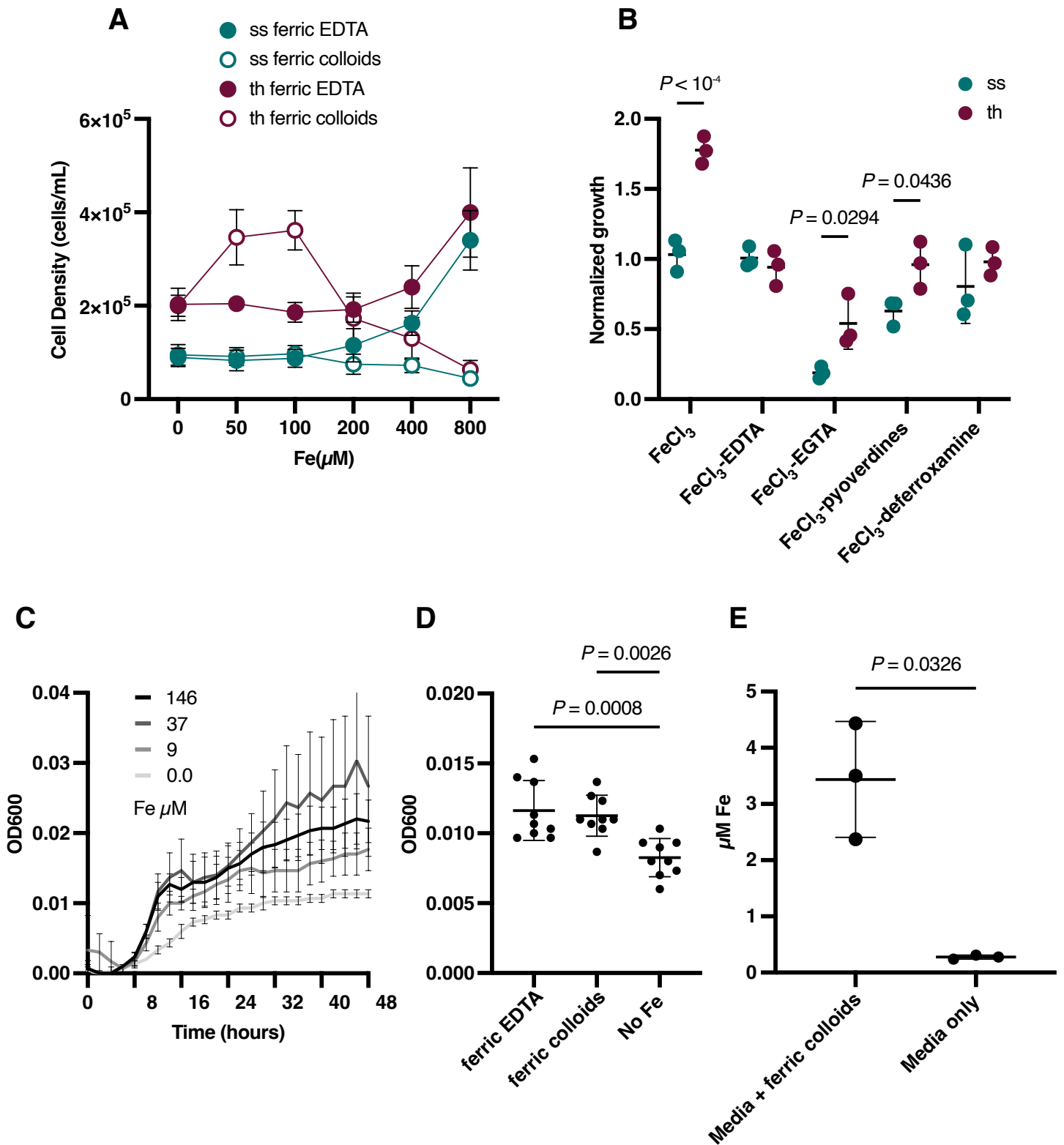


Figure S5: *S. rosetta* cell types, but not their feeder bacteria, grow differently with variable iron sources.

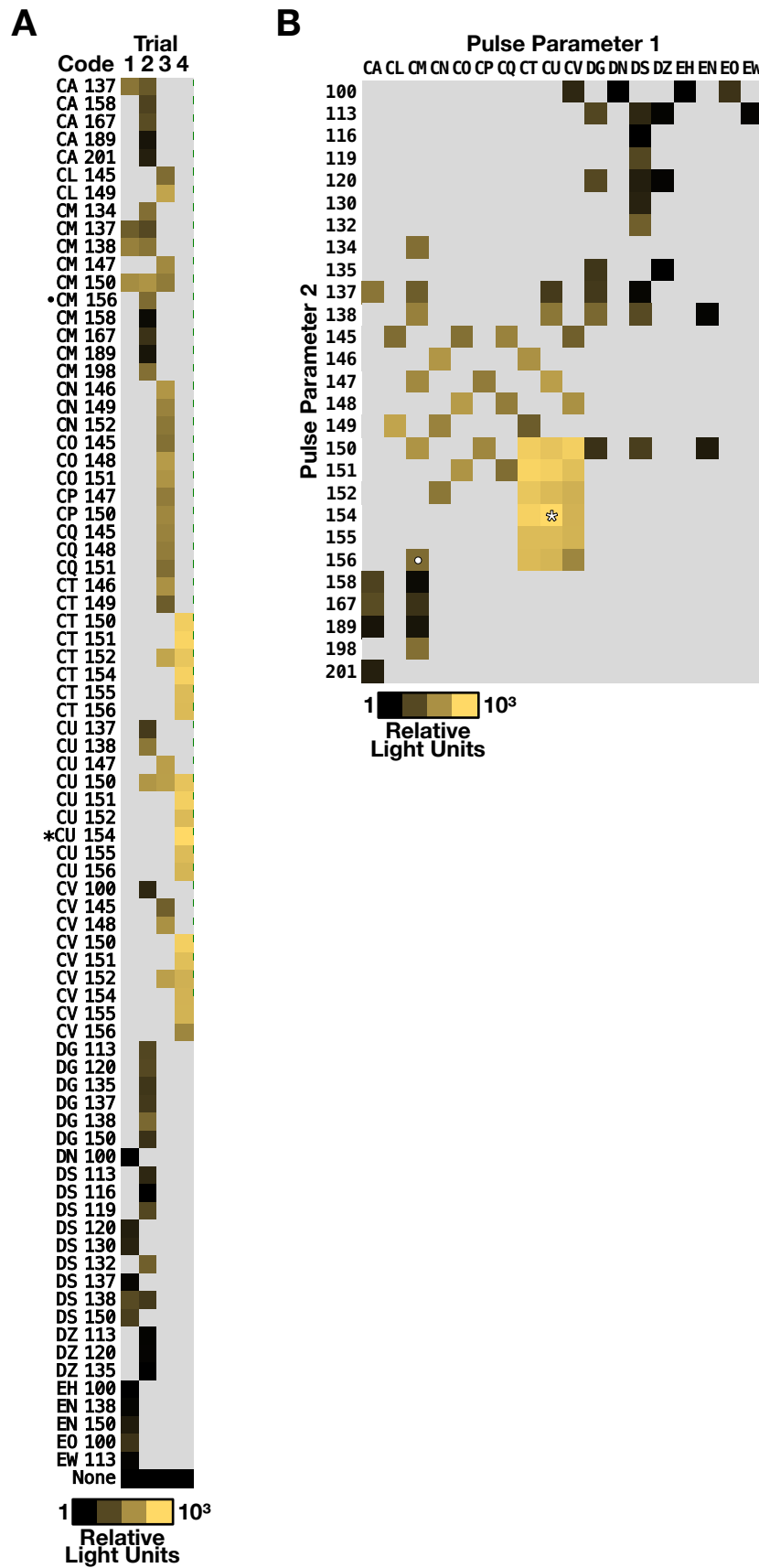


Figure S6: A screen for optimal nucleofection pulses to improve cargo delivery into *S. rosetta*.

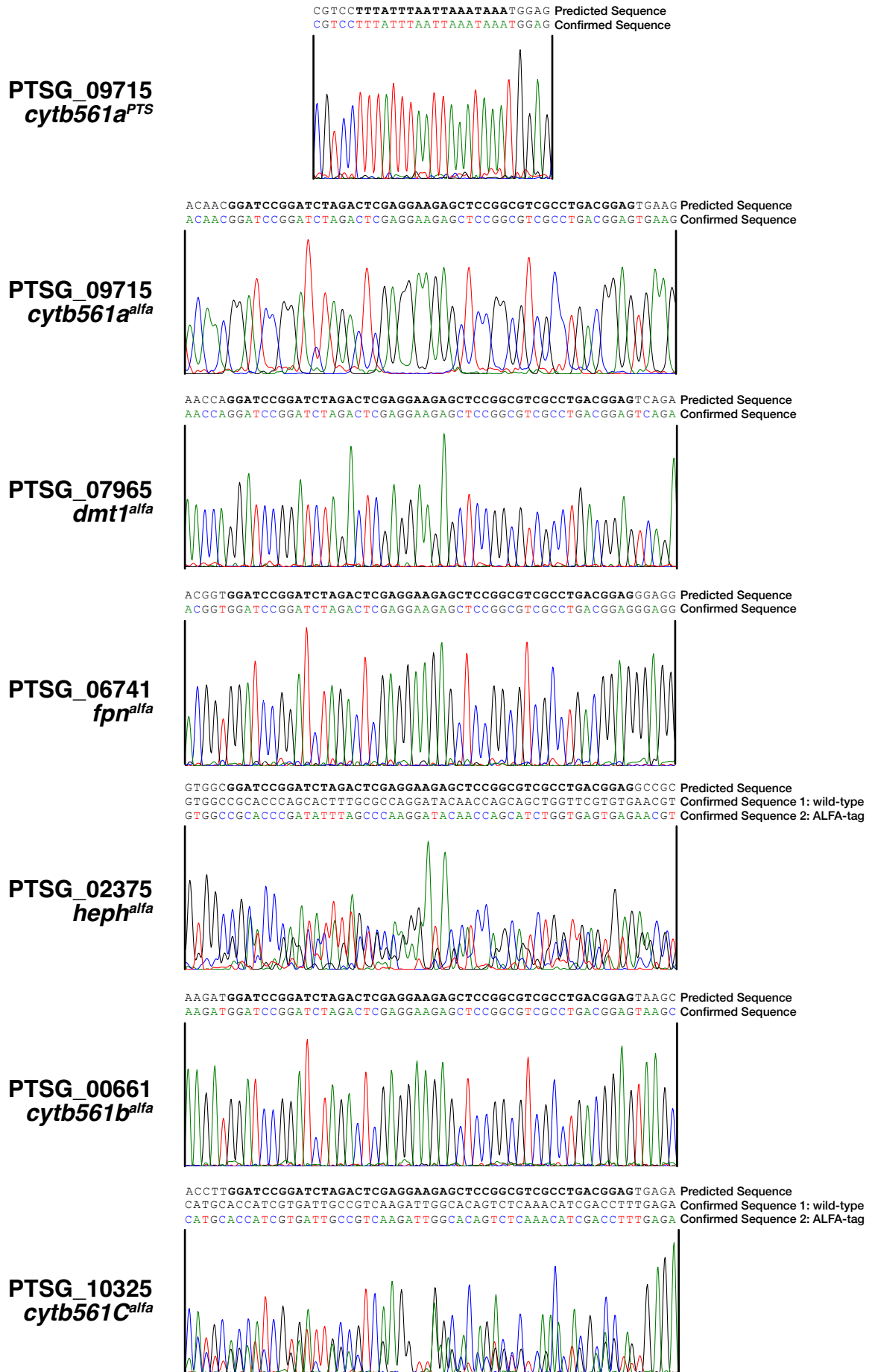


Figure S7: Genotypes of engineered strains.

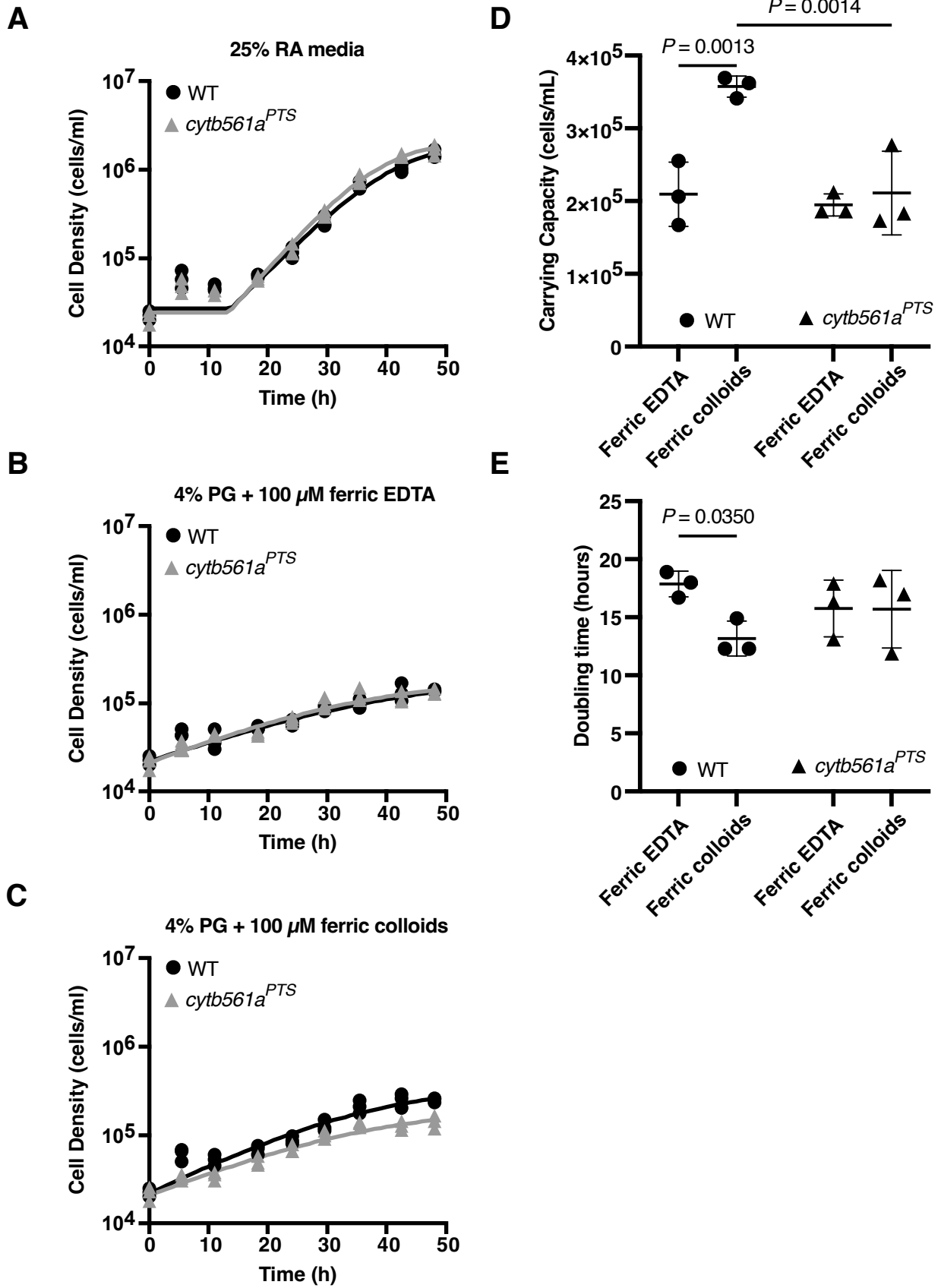


Figure S8: Thecates grow slight faster and to a larger population size with ferric colloids.

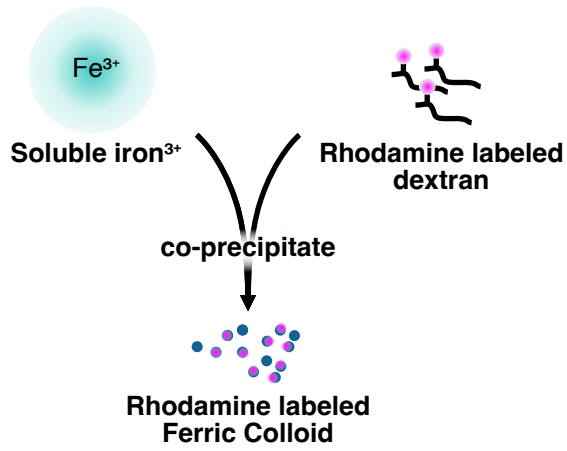
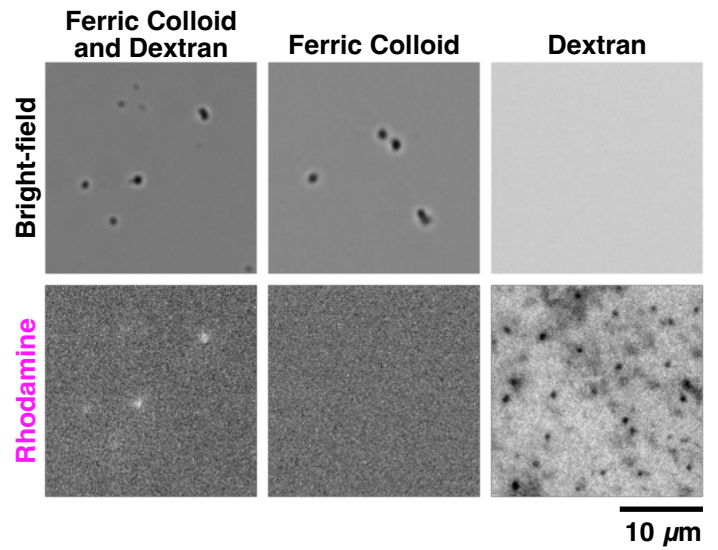
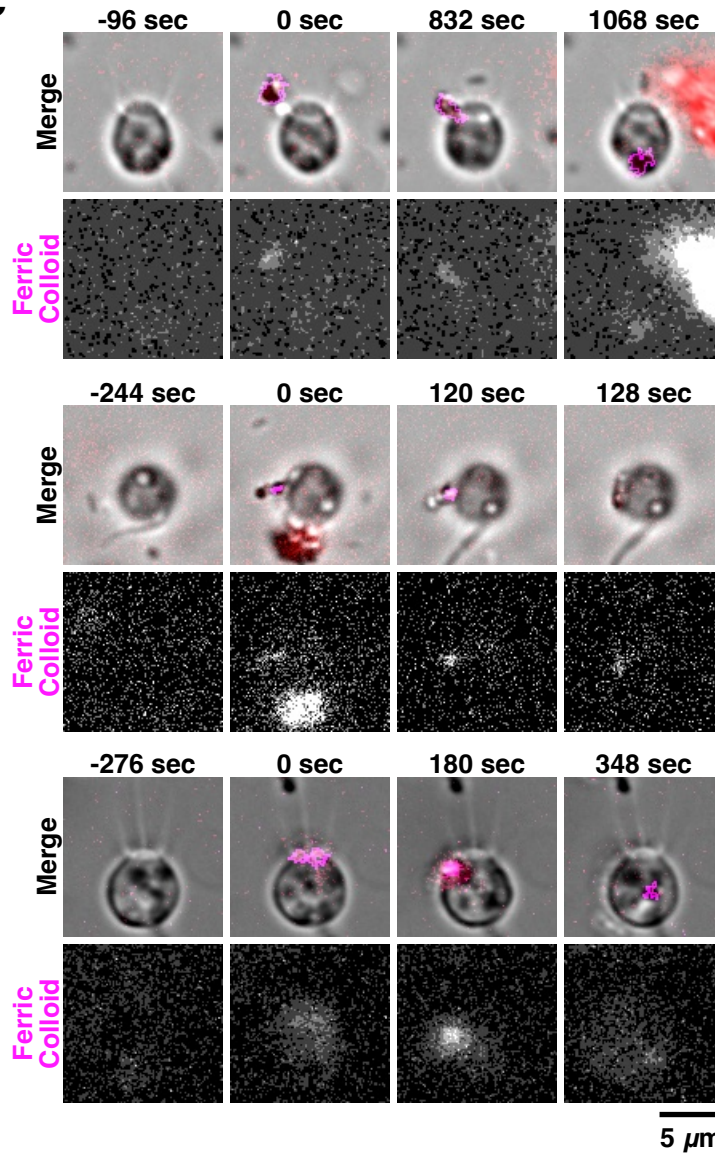
A**B****C**

Figure S9: Thecates phagocytose ferric colloids.

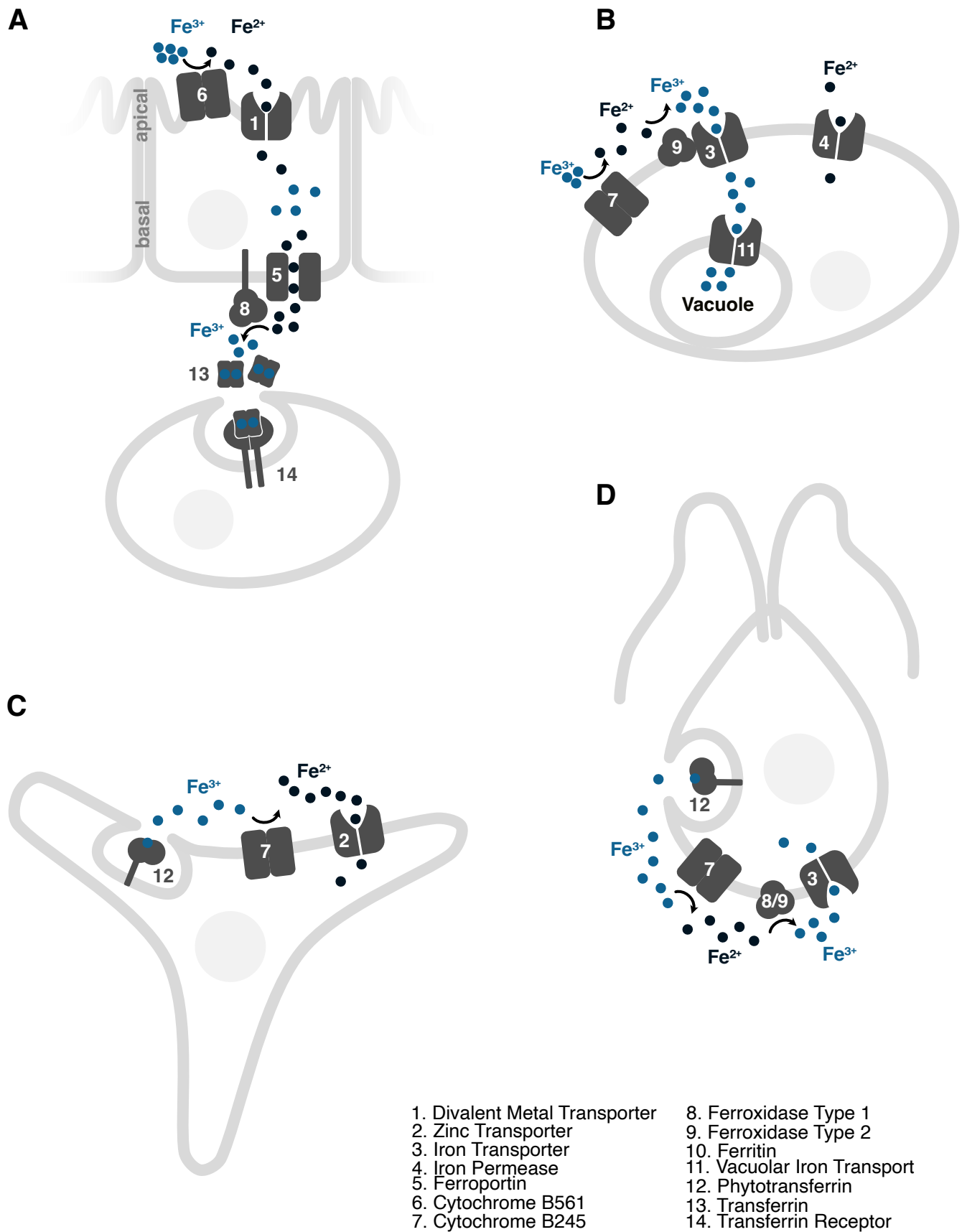
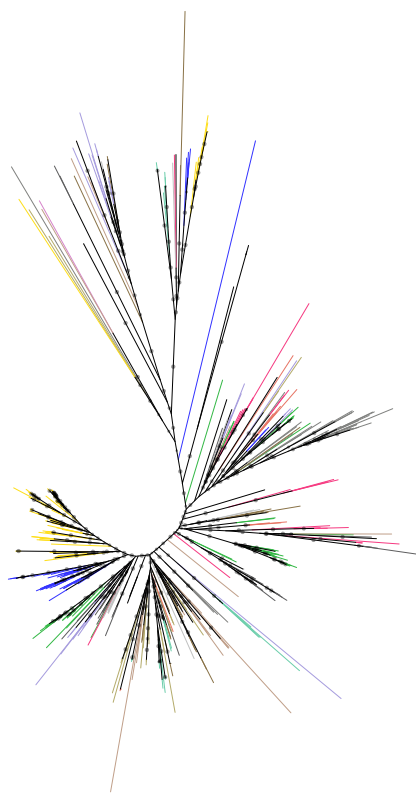
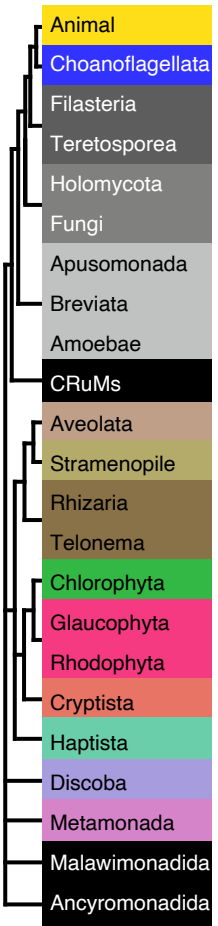


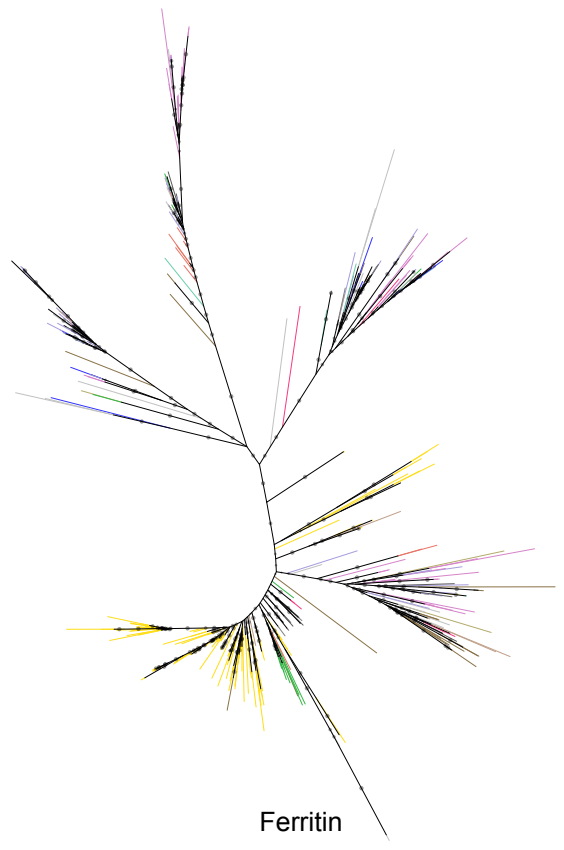
Figure S10: Simplified iron acquisition pathways in model eukaryotes.

Scale: 1
bootstrap

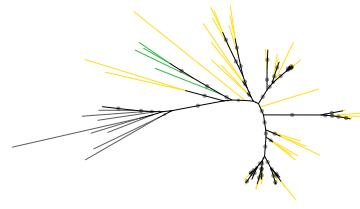
- 70
- 85
- 100



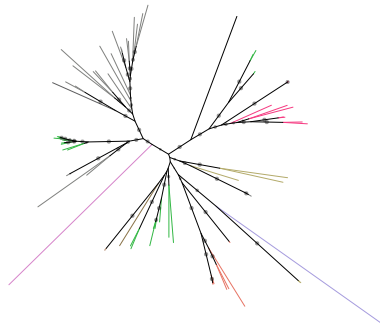
Divalent Metal Transporter



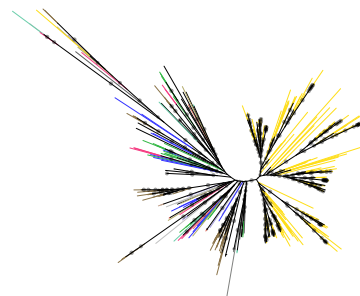
Ferritin



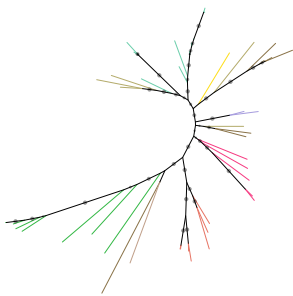
Transferrin



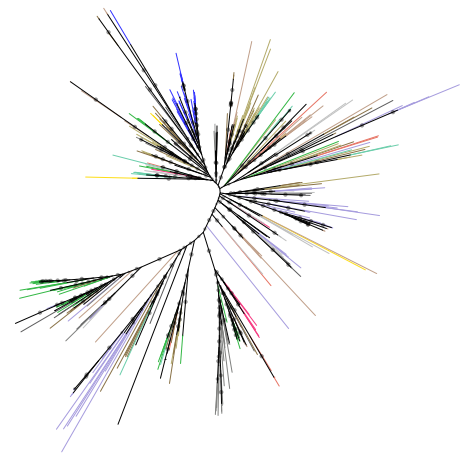
Iron Transporter



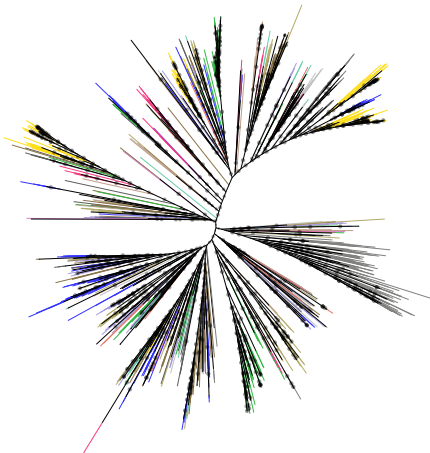
Ferrooxidase type 1
(Hephaestin)



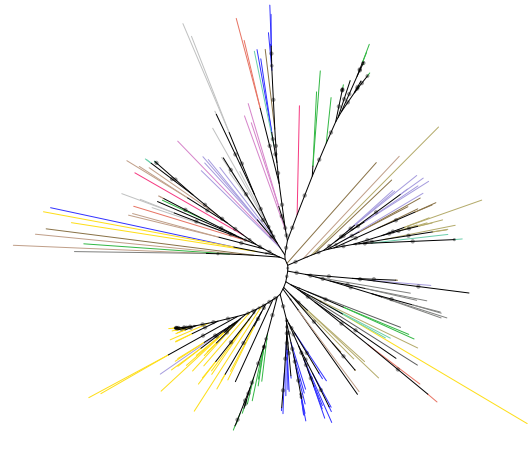
Phytotransferrin



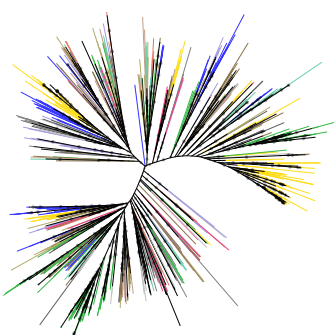
Vacuolar Iron Transporter



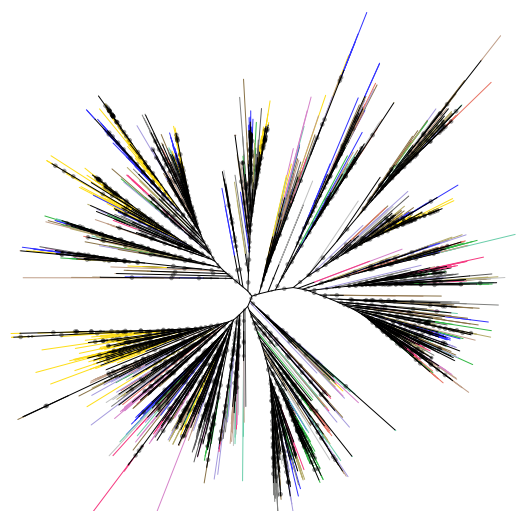
Cytochrome B245



Ferroportin



Cytochrome B561



Zinc Transporter

Figure S11: Phylogenetic trees of iron acquisition proteins.

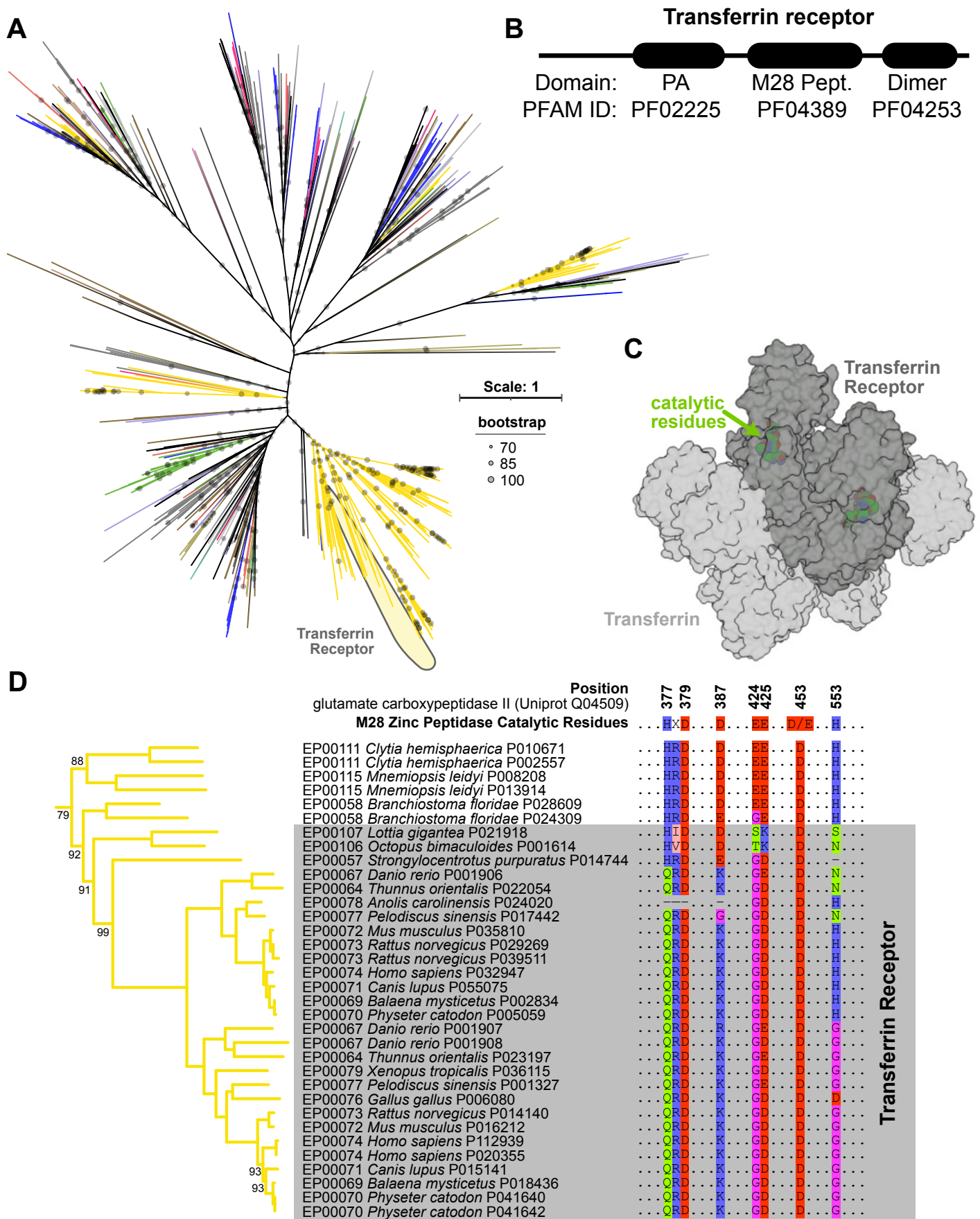


Figure S12: Animal transferrin receptors evolved from an ancestral M28 peptidase.

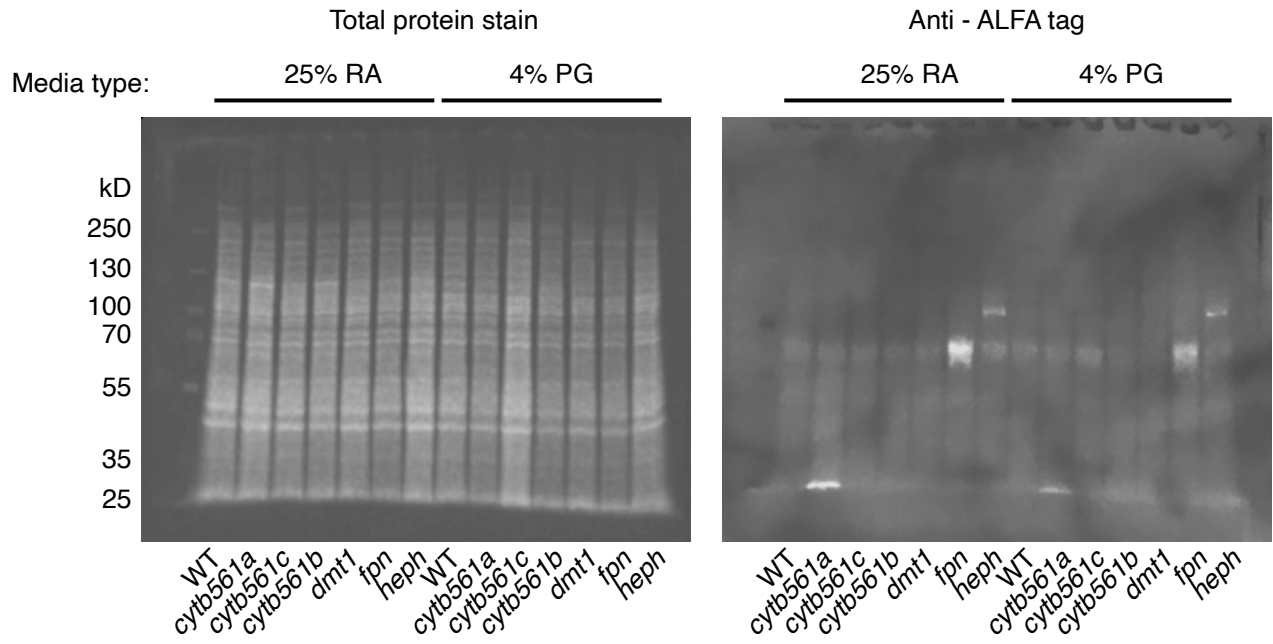
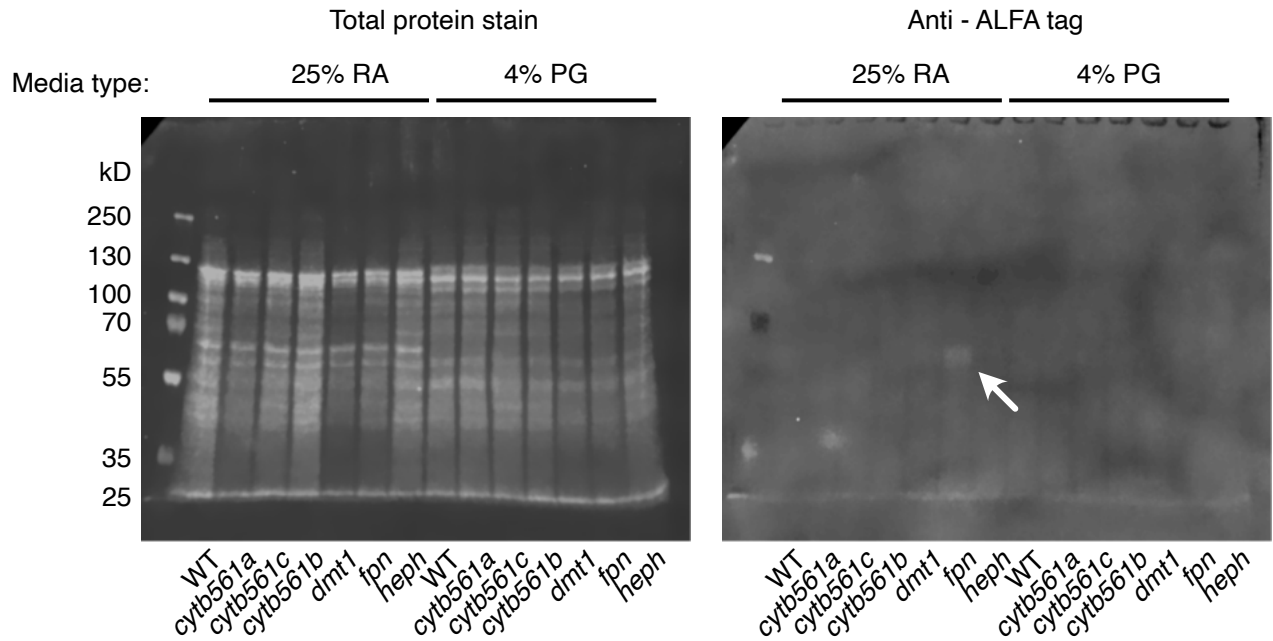
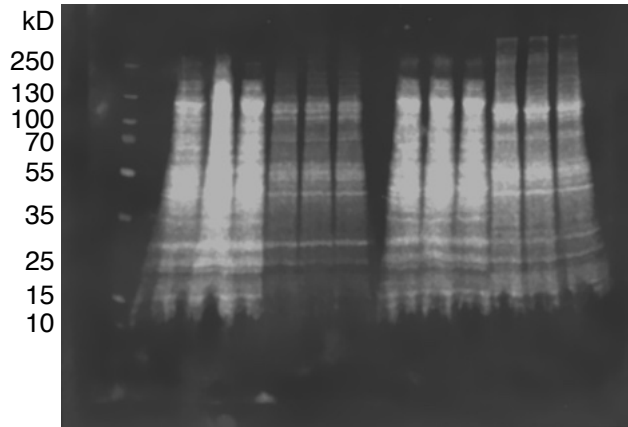
A**Thecates****B****Swimmers**

Figure S13: Thecates detectably express Cytb561a, Fpn, and Heph in western blots.

A

Total protein stain

Thecates Swimmers
25%RA 4%PG 25%RA 4%PG



B

Anti - ALFA tag

Thecates Swimmers
25%RA 4%PG 25%RA 4%PG

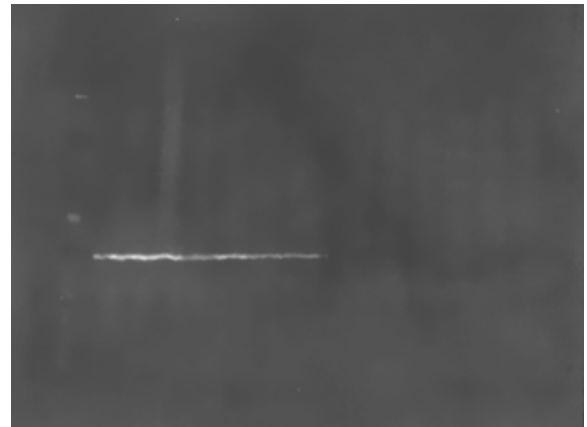


Figure S14: Thecates stably express Cytb561a independent of iron availability.

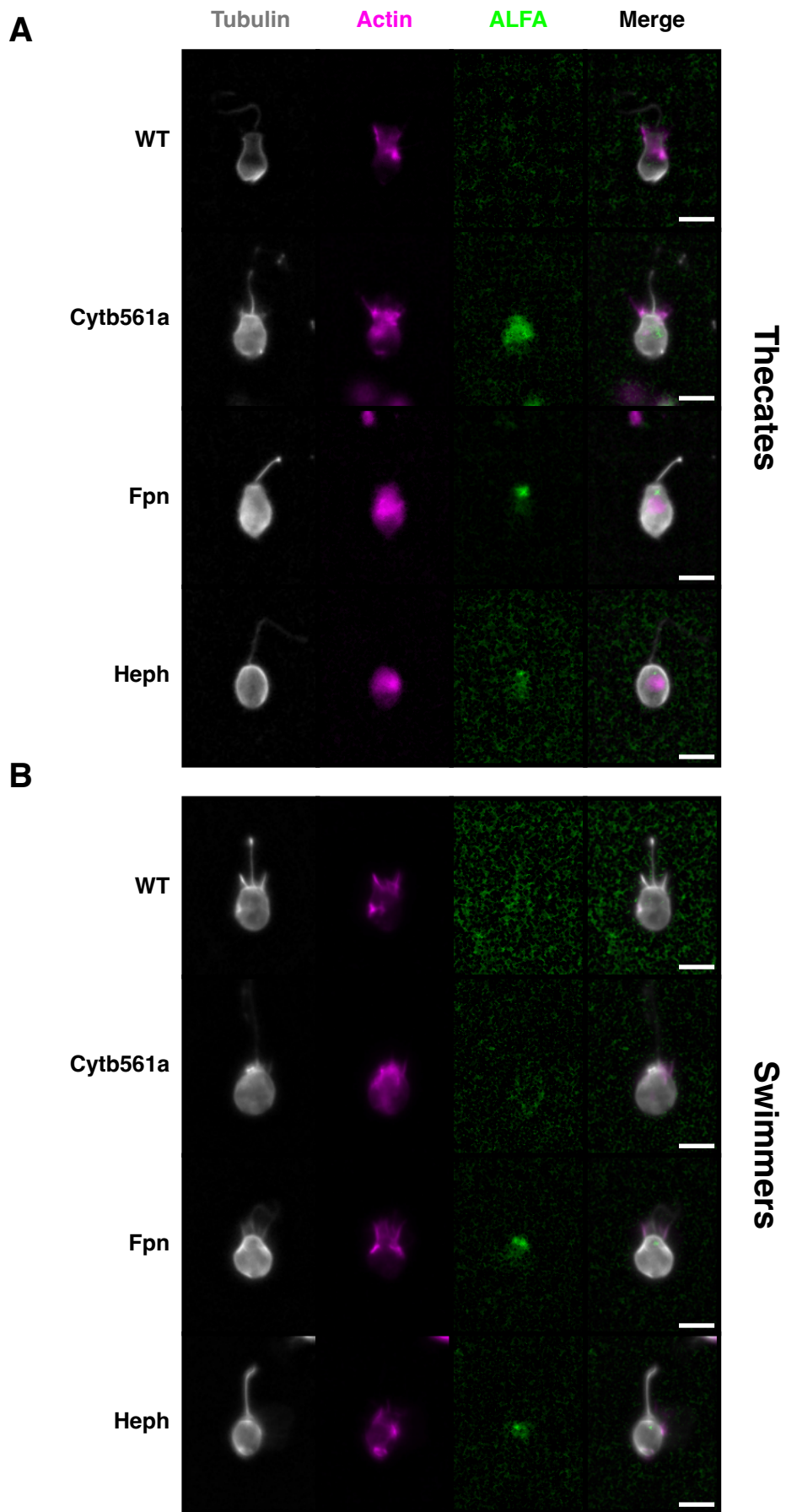


Figure S15: Additional immunofluorescence of thecates and swimmers.

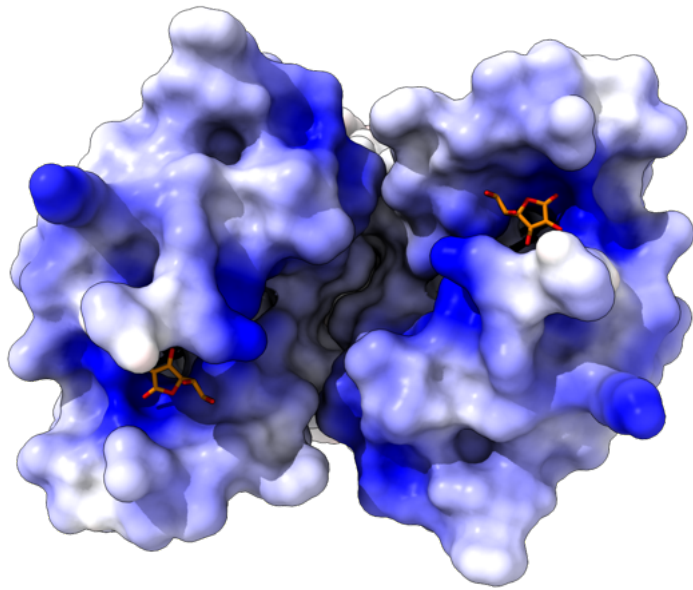
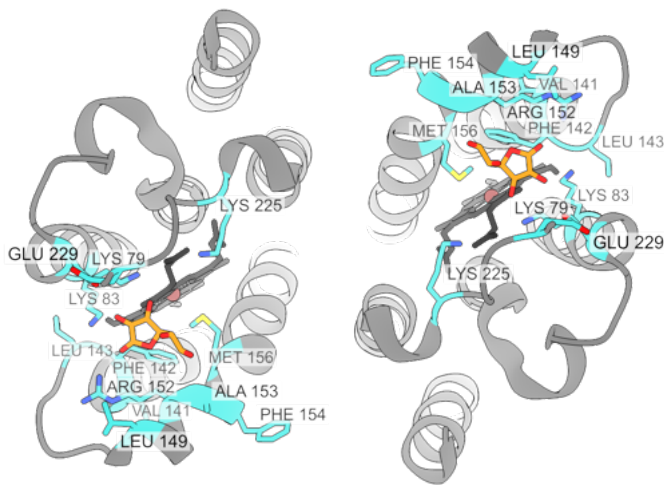
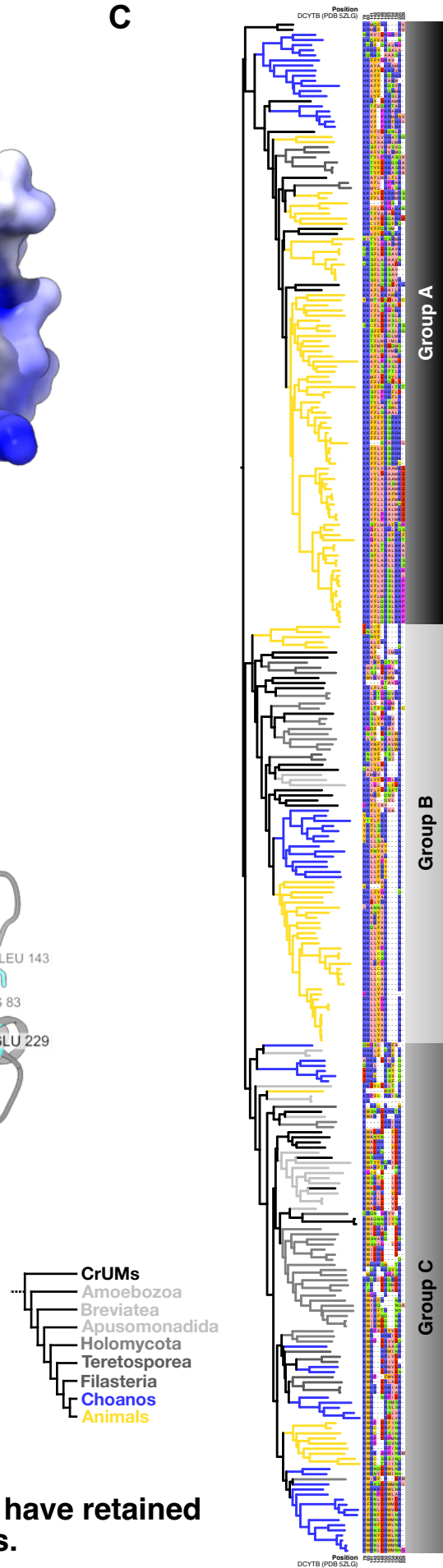
A**B****C**

Figure S16: Cytochrome b561 paralogs have retained ascorbate binding residues.

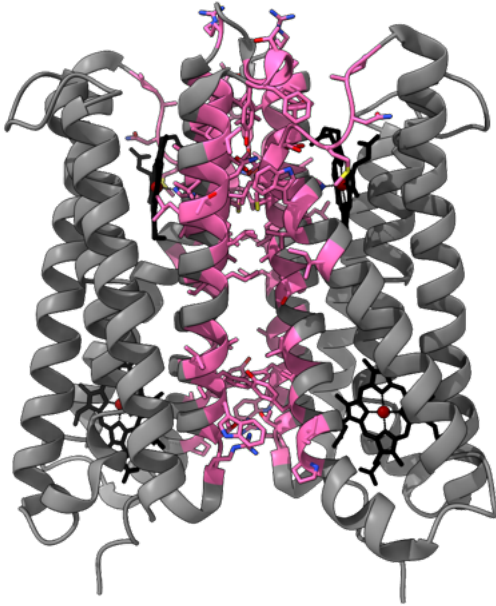
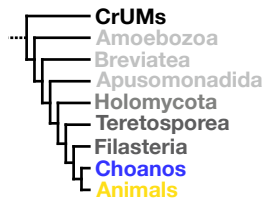
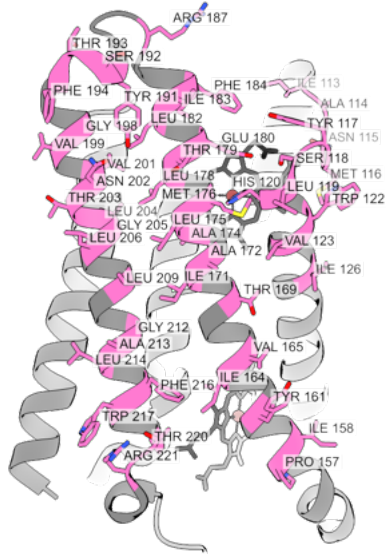
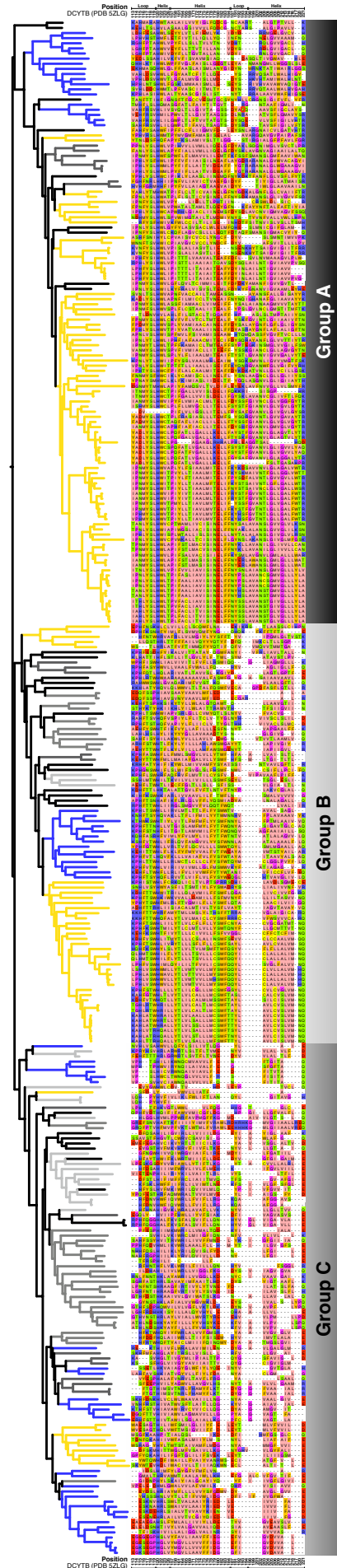
A**B****C**

Figure S17: Cytochrome b561 dimerization is a feature of Group A proteins.

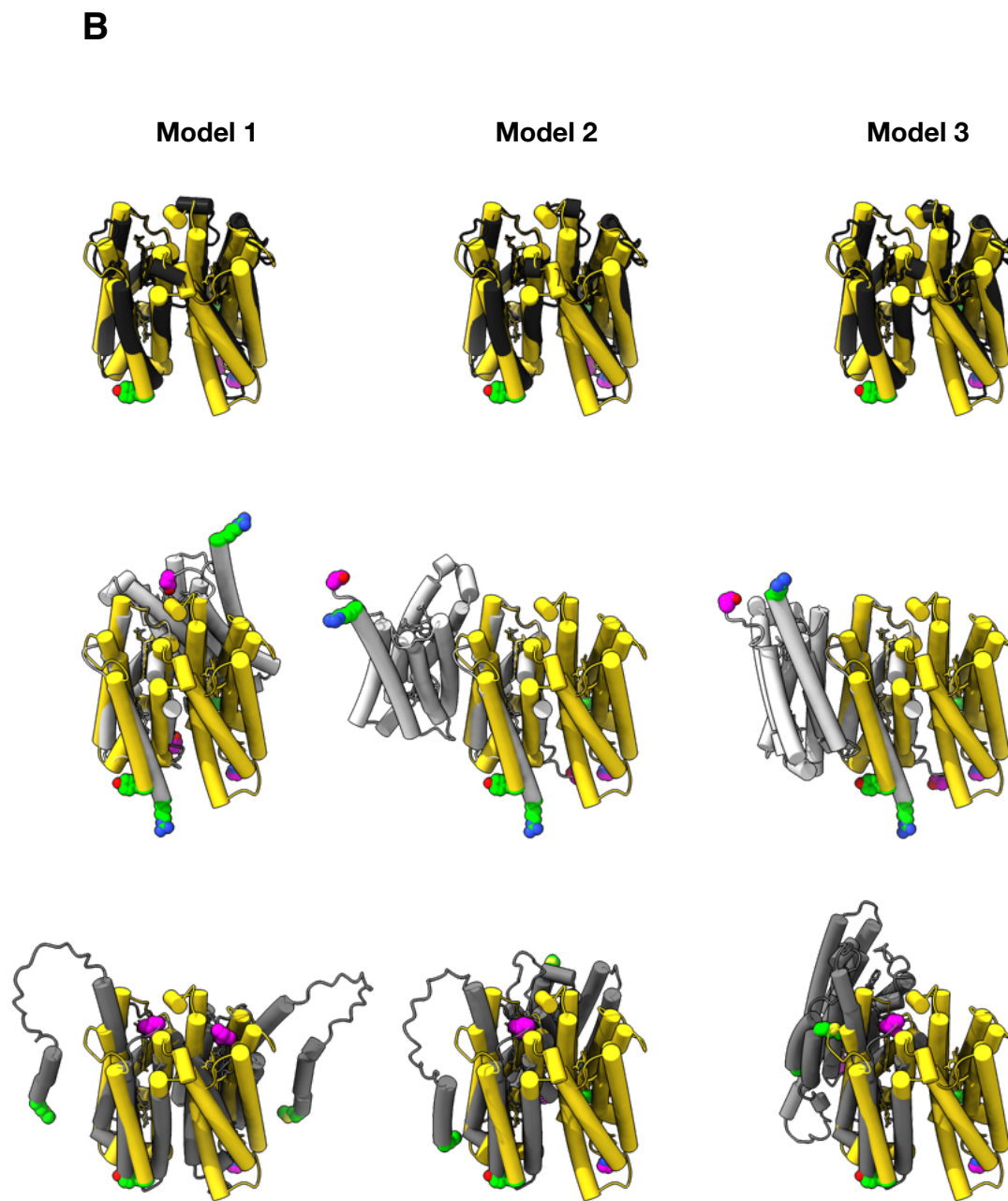
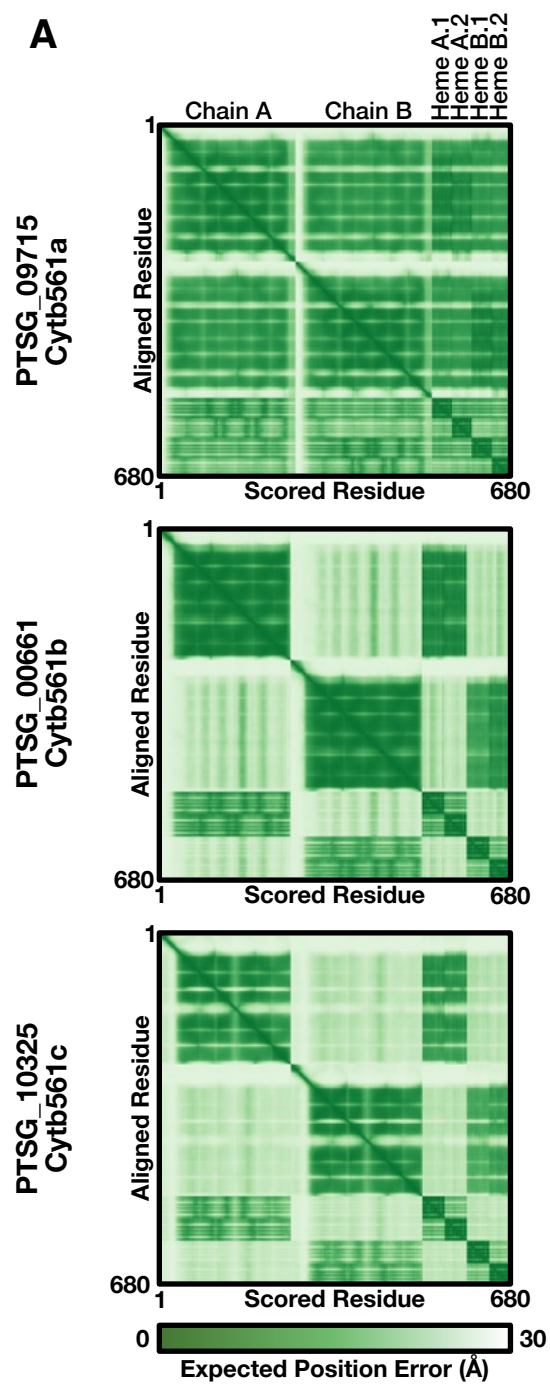


Figure S18: Cytb561a is predicted to dimerize similarly to human DCYTB.

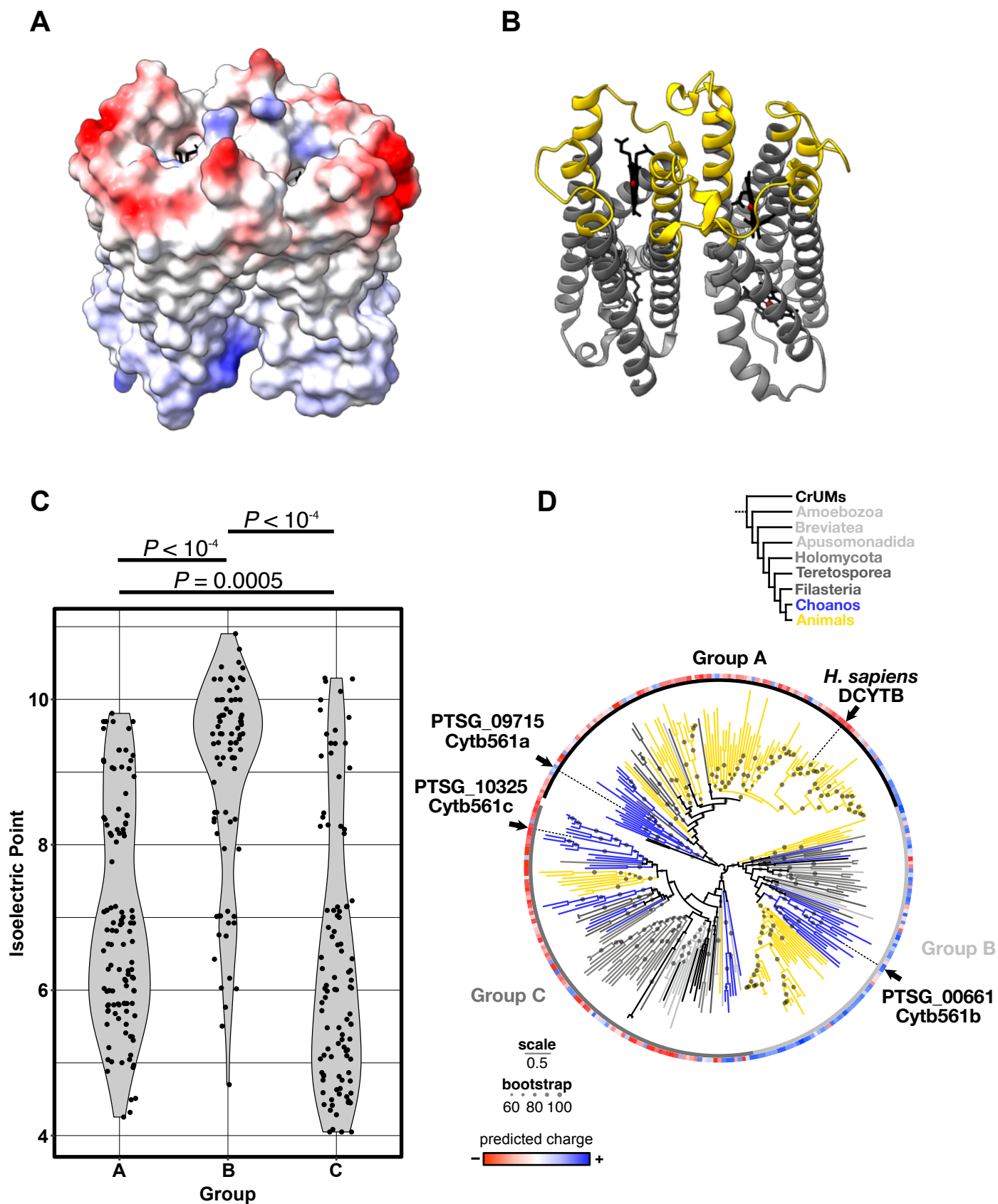


Figure S19: The lumenal surfaces of Cytochrome b561 feature electrostatic properties that are general features of Cytochrome b561 subgroups

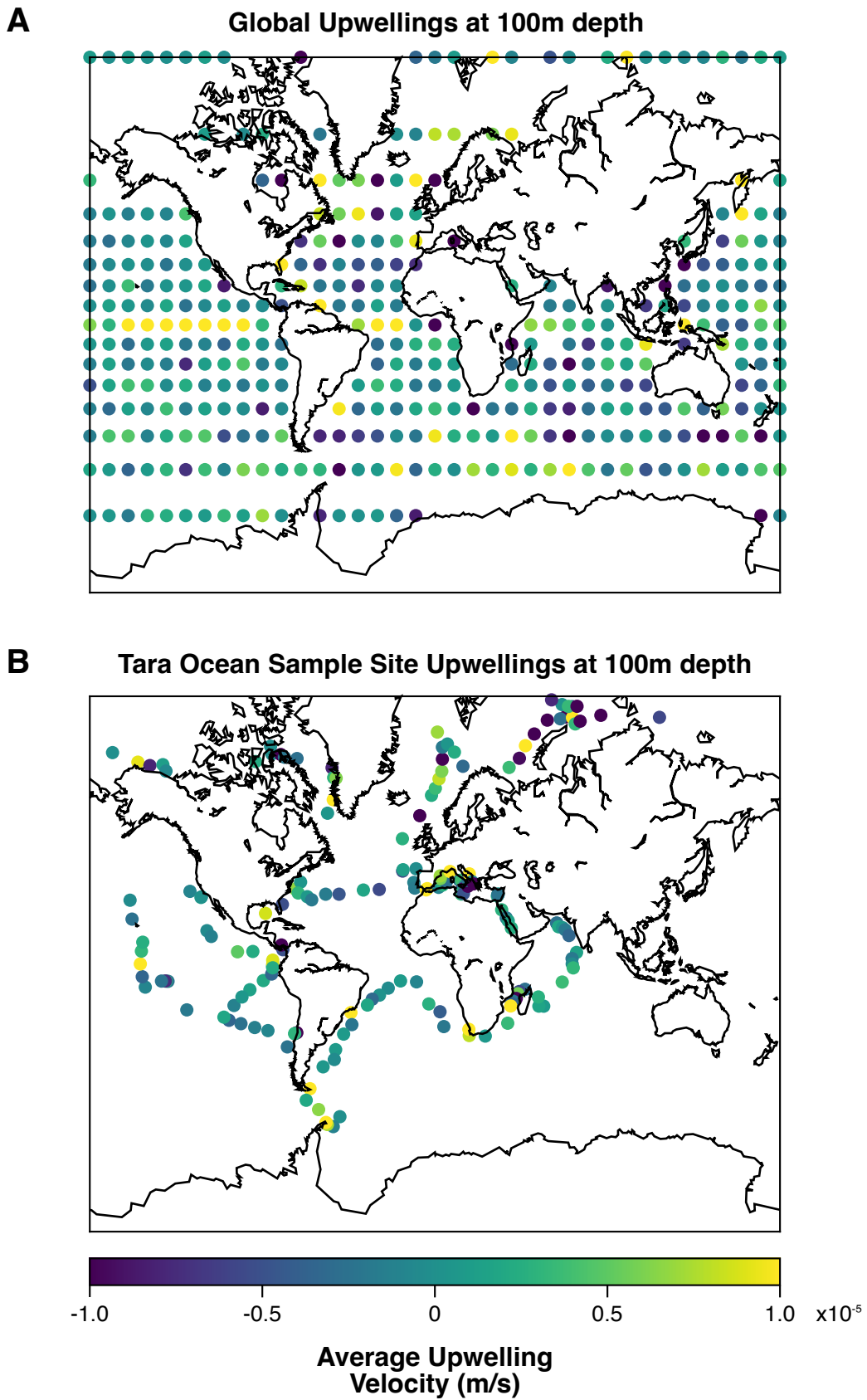


Figure S20: Maps of marine upwelling velocities show increased upwelling near coastlines and the equator.

1 **Tables:**

2

3 **Table S1:** Media table.

4 Components and recipes for all media used: 5% seawater complete (SWC0, 10% cereal grass
5 media (CGM3), 10% red algae (RA), 25% red algae (RA), 15% red algae, 2% peptone, yeast
6 extract, glycerol (15/2), and 4% peptone, glycerol (PG).

7

8 **Table S2:** RNA sequencing metadata.

9 Tabulated strain IDs, culturing conditions, NCBI accession IDs, and sequencing metadata for
10 RNA sequencing.

11

12 **Table S3:** Table of differential gene expression between cell types.

13 Pairwise comparisons of gene expression between cell types, measured by RNA sequencing in
14 transcripts per million (TPM) and scaled reads per base. Each gene's expression is presented
15 as individual triplicate values, along with the mean and standard deviation. Comparisons are
16 made between any two cell types and shown as fold change, standard error, and associated
17 *P*-value and *q*-value. Each gene is also annotated with accession IDs for uniprot, pfam, interpro,
18 orthoDB, unipathway, and gene ontology (GO).

19

20 **Table S4:** Table of oligonucleotides and gRNAs for mutant generation and screening.

21 Tabulated guide RNAs, repair template oligonucleotides, and primers for generating, screening,
22 and isolating mutants.

23

24 **Table S5:** Strain table.

25 The strains of *S. rosetta* used and created during this study, along with the crRNAs, repair
26 templates, Cas12a screening gRNAs, and screening primers for generating each strain.

27

28 **Table S6:** Iron acquisition paralog search and phylogenetic tree metadata.

29 Table of the proteins, HMM models, and PFAM IDs used to search for paralogs, along with the
30 associated protein names used in Fig. 3A. Metadata for HMMer searches, multiple sequence
31 alignments, trimming, and phylogenetic tree are included. The second sheet contains paralog
32 hits filtered for uniqueness and length, which is the raw data used in Fig. 3A. The third sheet is
33 the same paralog heatmap, but without the additional length filtering, resulting in more hits per
34 proteome. Colors of species match Fig. 3A, as does the protein order.