Current Biology, Volume *23*
Supplemental Information

Evidence for Sex and Recombination in the Choanoflagellate *Salpingoeca rosetta*

Tera C. Levin and Nicole King

Figure S1, related to Figure 1. Analysis and stability of *S. rosetta* ploidy (A – D) Flow cytometry analysis to differentiate propidium iodide signals from *S. rosetta* nuclei, co-cultured bacteria, and yeast internal standard. *S. rosetta* samples were analyzed before (not shown) and (**A**) after adding a yeast internal standard. This before-and-after analysis allowed us to clearly identify the yeast signal and to draw the gates separating *S. rosetta* and yeast populations, based on forward and side scatter. As forward scatter separates particles roughly based on size, the small, isolated choanoflagellate nuclei are to the left of the larger, fixed, intact yeast cells. **(B)** Histogram of propidium iodide intensities from the yeast internal standard, gated in (**A**). These peaks provided the internal standard, allowing for comparison of DNA content among *S. rosetta* samples. **(C)** Clumps of nuclei can yield aberrantly large propidium iodide signals, but these clumps can be eliminated by doublet discrimination as follows. We plotted peak height by peak area in the propidium iodide channel. Data points off the diagonal represent clumps of nuclei, and these were excluded from the analysis. Bacterial signals may be discriminated from choanoflagellate signals due to their small DNA content, so we also eliminated the residual bacterial signal in the lower left corner during this step. **(D)** Final histogram of propidium iodide intensities from *S. rosetta* nuclei, after eliminating bacteria and clumps of nuclei in (**C**). Horizontal lines separate groups of samples that were run on different days, with slightly adjusted voltages and gates on each day. In particular, after the first four samples were obtained, the forward scatter parameter was later adjusted to bring the yeast data points off axis. (**E – H)** Splitting cultures into fresh growth media every 3 days for 3 months did not induce changes in ploidy in isolate A (**E**), isolate B (**F**), isolate C (**G**) or Px1 (**G**). Fig. 1B – D shows isolates A – C and Px1 three months earlier than the samples analyzed here. Isolate C was propagated in HN medium while the other cultures were propagated in CG medium. (**I**) An isolate B culture propagated daily in CG medium for 3 weeks had at least as many haploid cells as diploid cells, depending on the relative abundance of cells in G1 vs. G2.

Figure S1

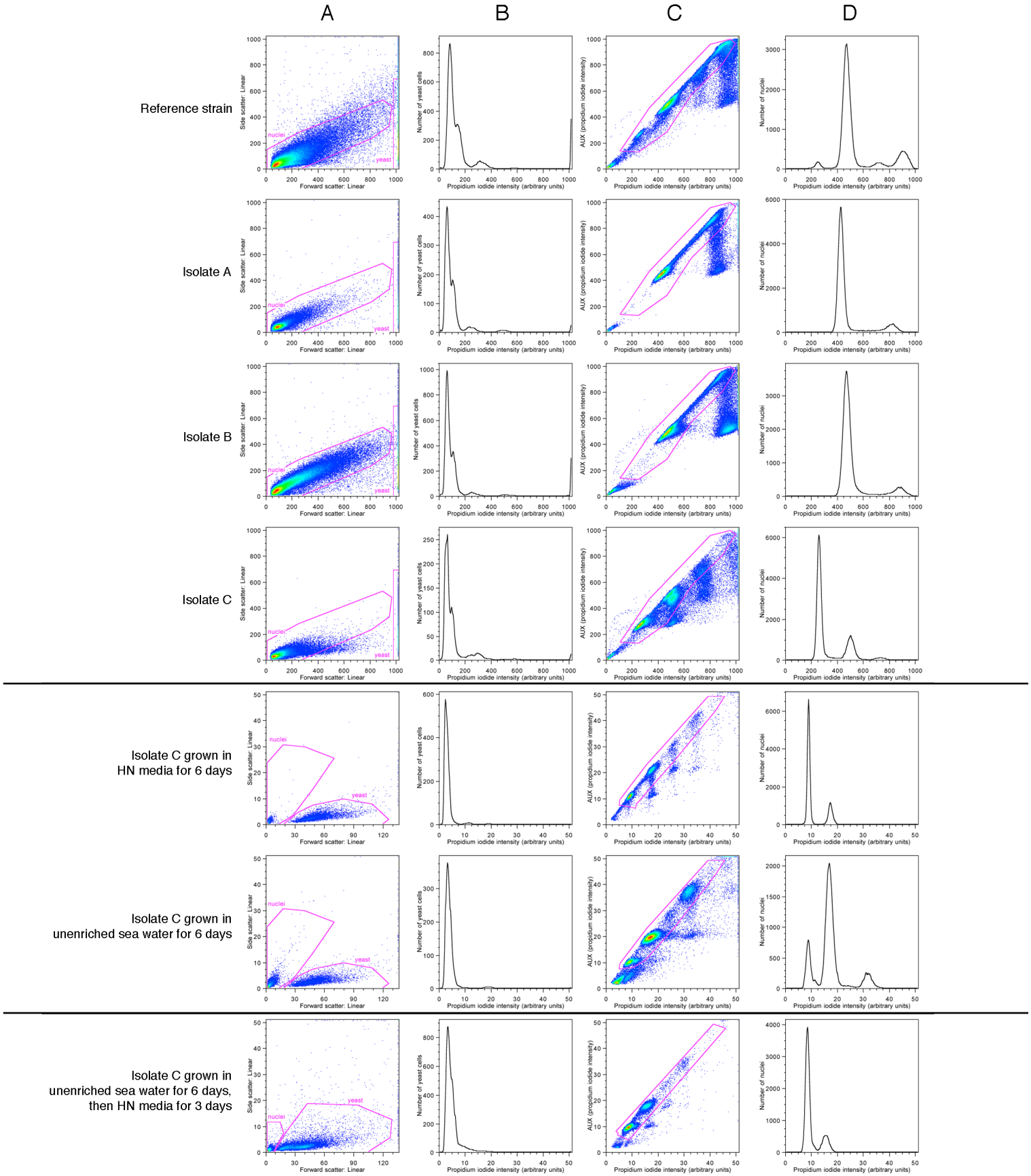

A
B
C
D
Reference strain
Isolate A
Isolate B
Isolate C
Isolate C grown in HN media for 6 days
Isolate C grown in unenriched sea water for 6 days
Isolate C grown in unenriched sea water for 6 days, then HN media for 3 days
Side scatter: Linear
Forward scatter: Linear
Number of yeast cells
Propidium iodide intensity (arbitrary units)
AUX (propidium iodide intensity)
Number of nuclei
nuclei
yeast

Figure S1 ctd.

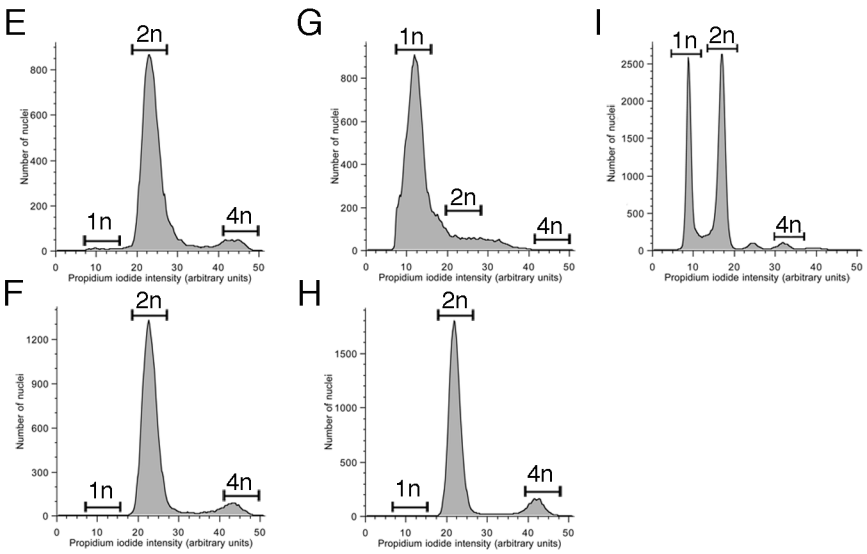


Figure S2, related to Figure 2. Detection and analysis of *S. rosetta* SNPs. (A) Homozygosity and heterozygosity of sequenced isolates. Because all of the sequenced isolates were clonally derived, we expect there to be only one or two copies of each homologous chromosome in each population. To determine for each isolate whether it was heterozygous or homozygous, we analyzed the frequency of alleles in SNP-containing reads generated by short-read genome sequencing. For heterozygous loci in a diploid organism, we expect on average 50% of the reads to contain a SNP (with the other 50% carrying the same allele as the reference sequence), whereas homozygous loci (either in haploids or in diploids with the same allele on both homologous chromosomes) should have a SNP in 100% of the reads. The majority of SNPs in isolate A were heterozygous whereas the majority of SNPs in isolates B and C were homozygous. As isolate C is haploid, we expected to detect only homozygous SNPs, yet there were a small number of SNPs called as heterozygous. Read mis-mapping and/or read mapping in mis-assembled regions of the genome may explain these infrequent "heterozygous" SNP calls. **(B)** Full alignment of a portion of supercontig 8 that was cloned and sequenced from isolates A, B, and C. The two different clones from isolate A are indicated as A1 and A2. Of the multiple clones analyzed from isolate B, all sequences were identical, consistent with the predicted homozygosity of SNPs in this sample. Similarly, all clones from haploid isolate C were identical. The alignment is colored according to percent sequence identity; positions that are identical in all samples are dark blue, regions with shared insertions or deletions are light blue, and SNPs and unique insertions are in white. Dots indicate a section of Isolate A2 that we were unable to sequence due to the long GT repeat region, which was excluded from Fig. 2C as we could not determine the genotype of all four clones in this region. Fig. 2C in the main text shows the SNPs present from position 684 to 2070 (numbered according to Isolate A1) in this alignment.

Figure S2

A

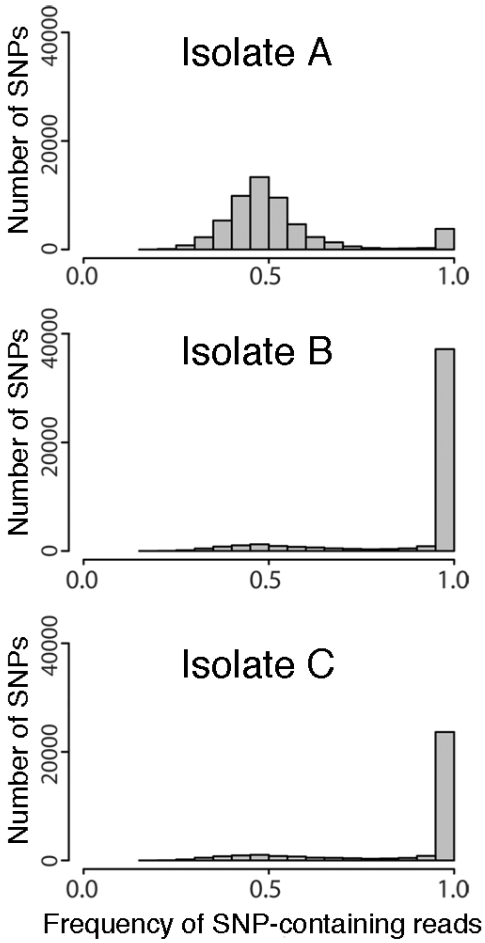

Isolate A
Isolate B
Isolate C
Number of SNPs
0
20000
40000
0.0
0.5
1.0
Frequency of SNP-containing reads

Figure S2 ctd.

B

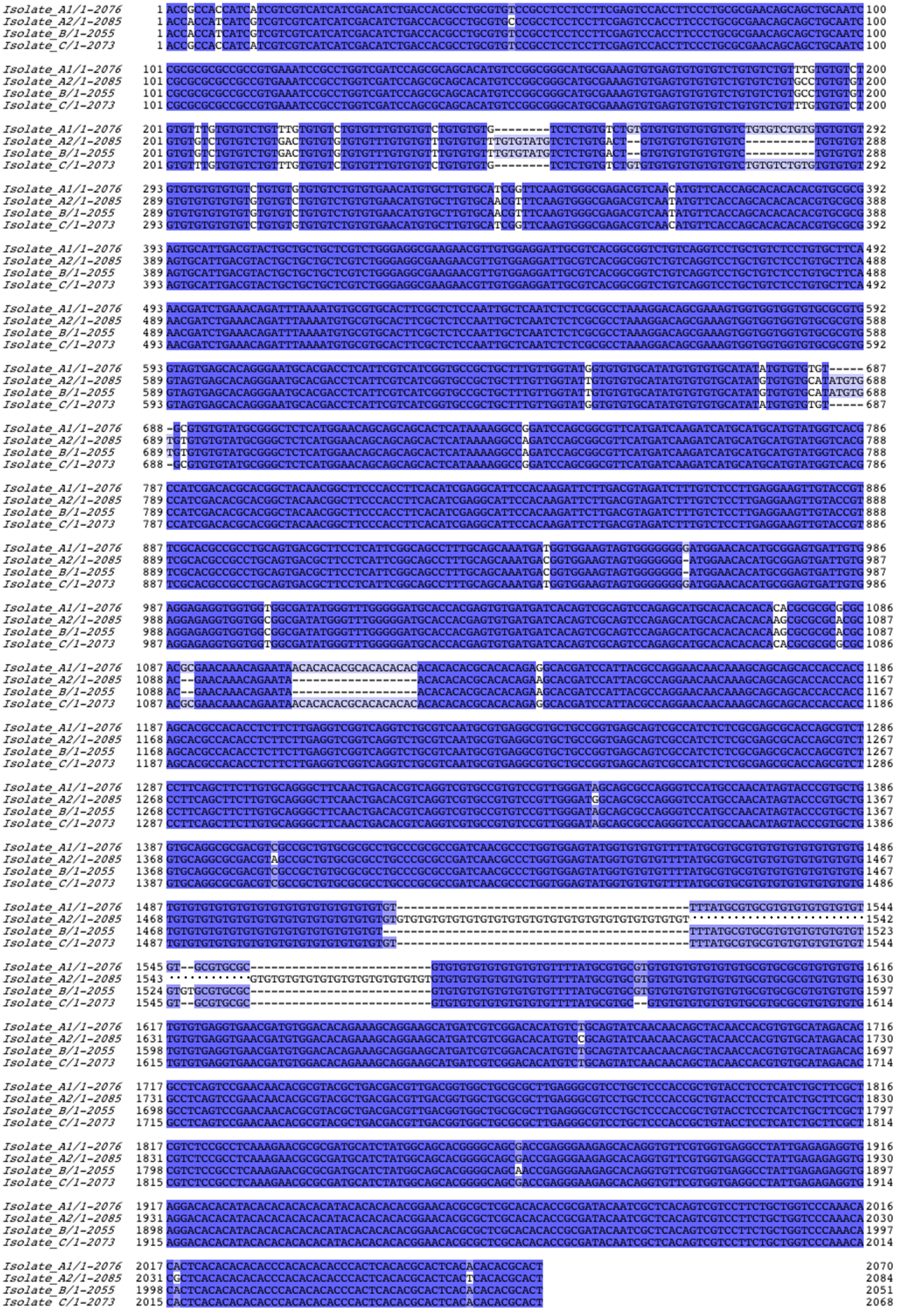
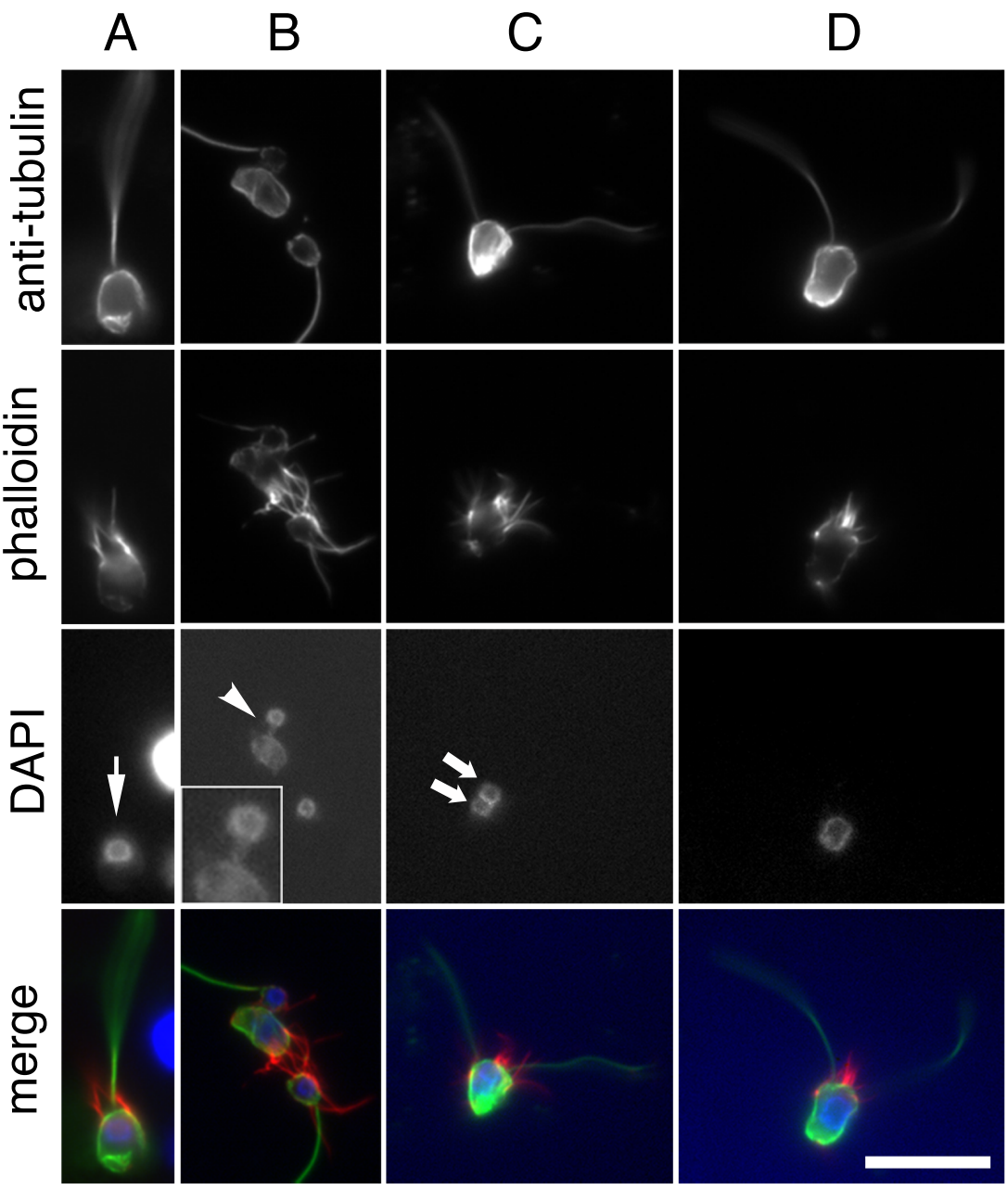


Figure S3, related to Figure 3. Immunofluorescence of cells during haploid-to-diploid transitions. (A) Haploid isolate C cells grown in HN media have one flagellum and one doughnut-shaped nucleoplasm (indicated by an arrow), as revealed by staining with beta-tubulin antibodies and DAPI. **(B)** Nutrient deprived cells differentiate into male and female gamete morphologies and often adhere to form loose clumps. The arrowhead in (**B**) indicates DNA that bridges between the nuclei of two cells, possibly genetic transfer. **(C – D)** Nutrient deprived cultures also contain cells with multiple flagella and either one (**D**) or multiple (**C**) nuclei. These cells may represent intermediate stages of the cell and nuclear fusion involved in sex. Scale bar shows 10 μm.

Figure S3



Movie S1, related to Figure 3. Time-lapse video of cell fusion. Isolate C cells in unenriched sea water were imaged as the population was becoming diploid. A small, round, male gamete encounters a larger, more elongated female gamete and contacts the collar of the female gamete for 20 minutes. The male gamete orients to the base of the collar of the female gamete and the cells fuse to form a larger cell, which swims away. The movie is displayed at 420x real time.

Supplemental Experimental Procedures

Px1 and the *S. rosetta* reference genome sequence

All strains of *S. rosetta* currently in culture were derived from an environmental isolate that was started from a single, rosette colony (ATCC 50818) [S1]. Px1 (ATCC PRA-366) was derived from the environmental isolate through a combination of antibiotic treatments and serial dilution, and includes a single feeder bacterium, *Algoriphagus machipongonensis* [S1-S3]. Px1 has been sequenced and its genome is publicly available [S2].

At the time that the genome was sequenced, the ploidy of *S. rosetta* and the ability of this organism to undergo sexual reproduction in the lab were unknown. However, although we do not have information about the ploidy of the culture that was sequenced, we can reconstruct its ploidy in retrospect based on the data presented here.

Given that isolates A, B, and C were all clonally derived from Px1, we infer that Px1 was a mixed population of haploid and heterozygous diploid cells. This fact initially seemed paradoxical, as Px1 was itself clonally derived from the environmental isolate. However, our finding that rapidly passaged isolate B cells became largely haploid (Fig. S1I) suggests the following scenario: the single cell that founded Px1 was a heterozygous diploid, and during the rapid passaging to expand the culture, a subset of the clonal population underwent meiosis. The resulting culture was no longer clonal, and was instead a mixture of heterozygous diploid cells (similar to isolate A) and the haploid, meiotic progeny of these cells. We note that our analysis of the ploidy of a Px1 culture yielded a similar mixture of haploid and diploid cells (Fig. 1B).

During preparation of the cells for genome sequencing, the mixed-ploidy Px1 culture was passaged extremely rapidly to propagate a large number of cells. Based on our finding that rapid passaging induces *S. rosetta* cultures to become more haploid, we infer that at the time of harvesting, the sequenced culture was dominated by haploids. As *S. rosetta* grows exponentially in high nutrient conditions, if a single haploid clone was slightly more prevalent than its sister cells early on, then a single genotype could later predominate the culture and the resulting genome sequence. Any polymorphisms that might have existed in the non-dominant haploids and diploids in the culture would have represented a small proportion of reads, and may have been dismissed as sequencing error. Supporting this scenario, upon reexamination of SNP data generated during the genome sequencing project [S2], we in fact detected some of the same segregating polymorphisms shown in Fig. 2A.

Establishment of *S. rosetta* isolates

Isolates A, B, and C were derived from Px1 through multiple clonal isolation steps (Fig. 1A). To generate isolates A and B, a culture of Px1was passaged in CG medium every 3-4 days for approximately 9 months. Using an approximation of 1 generation per 10 hours (derived from the growth curve in [S4]), we estimate that this corresponds to 500-1000 generations. Subsequently, this long-term culture was mutagenized with EMS prior to clonal isolation. A sample of 5 x 10^5 cells/mL sea water was exposed to 0.3% (vol/vol) EMS (ethyl methanesulfonate, Sigma) at room temperature for 1hr. Following mutagenesis, the EMS was deactivated and diluted by washing the cells 3 times in 5% (wt/vol) sodium thiosulfate (Sigma) in unenriched sea water. Each wash included a spin at 1000 x g for 10min to pellet the cells. Cells were resuspended in 2 mL 10% (vol/vol) CG medium diluted in unenriched sea water and incubated at room temperature for 24 hours to recover. All of the above steps included a control sample that experienced the same washes and incubations, but was not exposed to EMS. Isolate A was derived from the control sample and isolate B was derived from the mutagenized sample. Clonal isolation was achieved using 2 serial dilution-to-extinction steps, similar to [S5]; cells were diluted to approximately 1 cell per well in 10%

(vol/vol) CG medium in sea water and aliquoted into 96-well plates. After 5-7 days, plates were examined visually to find isolated clones, which were then expanded and frozen [S6].

By counting the number of empty wells in the 96-well plates, we can estimate the probability of clonal isolation using the Poisson distribution. Estimates of lambda may be obtained by -ln (proportion of wells with no cells), and the probability of obtaining a clonal isolate is calculated as the (probability of getting exactly 1 cell per well)/(the probability of getting 1 or more cells per well). Using this method, we estimate that there is a >99% probability that isolates A and B each underwent a single cell bottleneck at least once during the two clonal isolation steps.

In isolate C, *A. machipongonensis* was replaced with a new feeder species, *Echinicola pacifica* (DSM 19836) [S7] through repeated antibiotic treatment and limiting dilution. As *E. pacifica* and *A. machipongonensis* have very similar antibiotic sensitivities, we first generated a strain of erythromycin-resistant *E. pacifica*. Briefly, *E. pacifica* cultures were mixed with *E. coli* carrying the pNJR5 vector [S8] and allowed to mate for 48 hours on polycarbonate mating filters (Millipore; 25 mm diameter, 0.45 μm pore). Bacteria were then plated on Sea water Complete Media plates [S9] with 10 μg/mL erythromycin and 200 μg/ml gentamycin to select for transformed *E. pacifica*. An individual *E. pacifica* colony was isolated and retested to confirm erythromycin resistance. Meanwhile, a culture of Px1 *S. rosetta* with *A. machipongonensis* was grown for 5 days in unenriched sea water with 20 μg/mL erythromycin to reduce the amount of bacteria in the culture. A single colony of erythromycin-resistant *E. pacifica* was added to this culture, and choanoflagellates were isolated by limiting dilution as above. These antibiotic treatment and limiting dilution procedures were repeated until *E. pacifica* was the only remaining bacterial species in the culture, as monitored by plating and 16S RFLP analysis [S1]. After this time, antibiotic treatment was discontinued. There were a total of 5 serial clonal isolation steps, yielding a >99.9% probability that isolate C experienced a clonal bottleneck. This strain is publicly available as ATCC PRA-390 under the name SrEpac, for *S. rosetta* grown with *E. pacifica* bacteria.

Quantification of DNA content

We used flow cytometry to measure the DNA content of isolated, propidium iodide-stained *S. rosetta* nuclei. Previously, this type of analysis was not practical in choanoflagellates, as the large quantity of DNA from the co-cultured bacterial prey obscured the signal from choanoflagellate DNA. Therefore, we developed a Percoll density gradient technique (similar to [S10]) to separate and eliminate a large fraction of the bacteria from the culture. Three to four million choanoflagellate cells were concentrated to 250 μL and loaded on top of 2 mL Percoll/sorbitol/sea water mixture [500 mM sorbitol, 50 mM Tris pH 8, 15 mM $MgCl_2$, 0.4% (vol/vol) artificial sea water, 10% (vol/vol) Percoll, PVP-coated (GE Healthcare Biosciences)] in a 15mL conical tube. This solution was centrifuged at 1000 x g for 10 minutes on a low brake setting in an accuSpin 1R centrifuge (Fisher) to allow for density-based equilibration. Choanoflagellates were retained in the pellet, while the bacteria-enriched supernatant was discarded. Cells were then washed in unenriched sea water.

While the density gradient greatly reduced the bacteria present in the water column, it did not eliminate the bacteria that were trapped on or inside of choanoflagellates that had been feeding. To physically separate *S. rosetta* nuclei from their prey bacteria, we used a gentle lysis protocol (similar to [S11]) and only measured the DNA in intact *S. rosetta* nuclei. Briefly, pelleted *S. rosetta* cells were resuspended in a low-salt buffer (50mM Tris pH 8, 140 mM NaCl, 1.5 mM $MgCl_2$) to induce swelling and mild osmotic shock. *S. rosetta* cells were spun down again and the pellet was resuspended in ice-cold low-salt buffer with NP-40 and RNAse A [50 mM Tris pH 8, 140 mM NaCl, 1.5mM $MgCl_2$, 0.9 mg/mL RNAse A (Qiagen),

0.25% (vol/vol) NP-40]. After a 5 minute incubation at 4 °C, the cell lysate was passed through a 40 μm filter and stained with 20 μg/ml propidium iodide (Sigma). Samples were analyzed on a Beckman-Coulter FC-500 cytometer, with gating based on forward and side scatter to differentiate stained *S. rosetta* nuclei from contaminating bacteria and the yeast internal standard (see below and Fig. S1A—D). The propidium iodide fluorescence was used as the trigger signal, and approximately 50,000 events were measured per sample. Doublet discrimination was used to eliminate aggregates of nuclei from the analysis [S12]. Flow cytometry data were analyzed using FlowJo from Tree Star Inc. (Ashland, OR).

To enable comparison of the DNA content among different *S. rosetta* cell lines, we used a sample of haploid DDY904 *Saccharomyces cerevisiae* [S13] as an internal standard in all experiments. A single yeast colony was inoculated into 2 mL YPD medium and grown shaking overnight at 200 rpm at room temperature. Yeast cells were prepared similar to [S14], by fixation in 70% ethanol, treatment with 9 mg/mL RNAse A (Qiagen) at 37 °C for 2 hours, treatment with 0.5 mg/mL proteinase K (Fisher) at 50C for 1 hour, and resuspension in 50mM Tris pH 7.5. A small sample of 2 to 5 μL of prepared yeast was spiked into each *S. rosetta* lysate sample. This internal standard allowed us to verify which peaks in different *S. rosetta* samples represented nuclei with equivalent DNA content. We gained additional information about *S. rosetta* ploidy by comparing the sizes of the yeast and *S. rosetta* peaks to each other. The *S. rosetta* '1n' peaks were 3 to 5.4 times larger than the yeast internal standard peaks, corresponding to the fact that the *S. rosetta* genome is approximately 4.4 times larger than the *Saccharomyces cerevisiae* genome and suggesting that the presumed '1n' *S. rosetta* peaks were indeed haploid. Differences in the fixation and staining methods for the yeast and *S. rosetta* can introduce variability into propidium iodide staining of these different species, but even so the size of the yeast peaks corresponds better to the '1n' peak being from haploid *S. rosetta* (4.4 times larger) than to a diploid *S. rosetta* (8.8 times larger). We therefore conclude that *S. rosetta* alternates between haploid and diploid states as opposed to diploid and tetraploid states. Finally, as an additional control, we also processed a sample of isolate A in all experiments, as isolate A was consistently diploid under standard growth conditions.

Estimating the percentage of haploid and diploid cells in mixed cultures

We used the relative sizes of the 1n, 2n, and 4n peaks to estimate the percentage of cells in the culture that were diploid. As the size of the 2n peak is the sum of the haploid G2 peak and the diploid G1 peak, it is not possible to use the height of the 2n peak as a proxy for the number of diploids. Instead, we used the following approach to approximate the number of diploid and haploid cells. First, we assumed that at a given time in an unsynchronized culture, the haploid and diploid cells will have a similar proportion of cells in the G1 and G2 phases of the cell cycle. Thus:

Proportion of cells that are haploid = $x + ax$

Proportion of cells that are diploid = $y + ay$

where x is the proportion of G1 haploids, y is the proportion of G1 diploids, and a is the ratio of cells that are in G1/G2. From the flow cytometry data, we can measure the proportion of cells in the 1n peak (= x), the 2n peak (= $ax + y$), and the 4n peak (= ay), and solve for a using the quadratic equation. This gives two solutions; a large G1/G2 ratio (with 2n consisting mostly of diploids in G1) or a small G1/G2 ratio (with 2n consisting mostly of haploids in G2). Cultures that have spent several days in unenriched seawater are not rapidly dividing, and so we use the solution with a large G1/G2 ratio to estimate the proportion of cells that are diploid.

Determination of SNP linkage

To investigate the linkage between SNPs, we cloned and sequenced a 2.3 kb region of supercontig 8 from isolates A, B, and C. We used PCR to amplify the region with a mixture of Taq (New England Biosciences) and Pfu (Finnzymes) polymerases and these primers: TGGAACAGTGCAGTCCTGTG and AGCTCAAACAGGTACGGACA. The region was cloned into the TOPO TA vector (Invitrogen) and transformed into TOP10 cells (Invitrogen). Multiple clones for each isolate were analyzed by Sanger sequencing to determine the linkage among SNPs. This region of supercontig 8 contains long stretches of GT dinucleotide repeats that hindered Sanger sequencing, but this problem was ameliorated by using elevated concentrations of dGTP in the Sanger sequencing reaction. However, the isolate A copy 2 clone contains an expansion of a GT repeat region that we were unable to sequence across fully; this region was excluded from Fig. 2B, but the full alignment (made with FSA [S15] and Jalview [S16]) is shown in Fig. S2A.

Supplemental References

S1. Dayel, M. J., Alegado, R. A., Fairclough, S. R., Levin, T. C., Nichols, S. A., McDonald, K., and King, N. (2011). Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. Dev Biol *357*, 73–82.

S2. Fairclough, S. R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H. M., Begovic, E., Richter, D. J., Russ, C., Westbrook, M. J., et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. Genome Biol *14*, R15.

S3. Alegado, R. A., Grabenstatter, J. D., Zuzow, R., Morris, A., Huang, S. Y., Summons, R. E., and King, N. (2013). Algoriphagus machipongonensis sp. nov., co-isolated with a colonial choanoflagellate. International Journal of Systematic and Evolutionary Microbiology *63*, 163–168.

S4. Fairclough, S. R., Dayel, M. J., and King, N. (2010). Multicellular development in a choanoflagellate. Curr Biol *20*, R875–6.

S5. King, N., Young, S. L., Abedin, M., Carr, M., and Leadbeater, B. S. C. (2009). Isolation of Single Choanoflagellate Cells from Field Samples and Establishment of Clonal Cultures. Cold Spring Harbor Protocols *2009*, pdb.prot5147–pdb.prot5147.

S6. King, N., Young, S. L., Abedin, M., Carr, M., and Leadbeater, B. S. C. (2009). Long-Term Frozen Storage of Choanoflagellate Cultures. Cold Spring Harbor Protocols *2009*, pdb.prot5149–pdb.prot5149.

S7. Nedashkovskaya, O. I., Kim, S. B., Vancanneyt, M., Lysenko, A. M., Shin, D. S., Park, M. S., Lee, K. H., Jung, W. J., Kalinovskaya, N. I., Mikhailov, V. V., et al. (2006). Echinicola pacifica gen. nov., sp. nov., a novel flexibacterium isolated from the sea urchin Strongylocentrotus intermedius. International Journal of Systematic and Evolutionary Microbiology *56*, 953–958.

S8. Shoemaker, N. B., Barber, R. D., and Salyers, A. A. (1989). Cloning and characterization of a Bacteroides conjugal tetracycline-erythromycin resistance element by using a shuttle cosmid vector. J Bacteriol *171*, 1294–1302.

S9. Atlas, R. M. (2004). Handbook of Microbiological Media 3rd ed. (CRC Press).

S10. Price, C., Reardon, E., and Guillard, R. (1978). Collection of Dinoflagellates and Other Marine Microalgae by Centrifugation in Density Gradients of a Modified Silica

Sol. Limnology and Oceanography *23*, 548–553.

S11. Woo, J. L., and Berk, A. J. (2007). Adenovirus ubiquitin-protein ligase stimulates viral late mRNA nuclear export. J. Virol. *81*, 575–587.

S12. Wersto, R. P., Chrest, F. J., Leary, J. F., Morris, C., Stetler-Stevenson, M., and Gabrielson, E. (2001). Doublet discrimination in DNA cell-cycle analysis. Cytometry *46*, 296–306.

S13. Uchida, M., Sun, Y., McDermott, G., Knoechel, C., Le Gros, M. A., Parkinson, D., Drubin, D. G., and Larabell, C. A. (2011). Quantitative analysis of yeast internal architecture using soft X-ray tomography. Yeast *28*, 227–236.

S14. Hutter, K. J., and Eipel, H. E. (1978). Flow cytometric determinations of cellular substances in algae, bacteria, moulds and yeasts. Antonie Van Leeuwenhoek *44*, 269–282.

S15. Bradley, R. K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., and Pachter, L. (2009). Fast statistical alignment. PLoS Comput Biol *5*, e1000392.

S16. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.