The following supplement accompanies the article

Differential gene expression is tied to photochemical efficiency reduction in virally infected *Emiliania huxleyi*

Ilana C. Gilg, Stephen D. Archer, Sheri A. Floge, David M. Fields, Alex I. Vermont, Anna H. Leavitt, William H. Wilson, Joaquín Martínez Martínez*

*Corresponding author: jmartinez@bigelow.org

Marine Ecology Progress Series 555: 13–27 (2016)

SUPPLEMENTARY MATERIALS

Quantitative RT-PCR assay design, multiplex optimization, and analysis for digital PCR.

Quantitative RT-PCR assay design

Genes implicated in the assembly and/or repair of PSII were selected for analysis from the *E. huxleyi* pan genome (http://genome.jgi-psf.org/Emihu1/Emihu1.home.html) (Table 1) based on their presence in an *E. huxleyi* CCMP 374 transcriptome dataset (MMETSP1006-MMETSP1009, http://data.imicrobe.us/project/view/104). We detected 2 out of the 3 annotated paralogs of *ftsH* from the pan genome (protein IDs 427625 and 416950) in the *E. huxleyi* CCMP 374 transcriptome. Because the majority of these hits were to 427625, we chose to design an assay to specifically target this paralog. We could not find evidence for 466116 sequence in the transcriptome, nor in a recent proteome from *E. huxleyi* CCMP 1516 (McKew *et al.,* 2013), which calls into question its core protein designation (Read et al*,* 2013). Two superoxide dismutase (*sod2*) paralogs were also selected for investigation (Table 1 in the main article; see Materials and Methods in the main article for justification for the selection of these 2 paralogs). MMETSP1006 – 1009 transcripts of each gene of interest were aligned along with the most likely gene model from the *E. huxleyi* pan genome and the complete gene using MEGA 6 (Tamura et al. 2013). Exon boundaries, when present, were cross-checked and in some cases, corrected relative to the transcript sequences. Probes and primers were then designed using the Integrated DNA Technologies (IDT) web-based PrimerQuest tool (http://www.idtdna.com/Primerquest/Home/Index). Whenever possible, forward or reverse primers were designed across transcript-validated exon junctions, resulting in a cDNA-specific assay. When genes lacked introns or when design across exon junctions was non-optimal, assays were designed within exons and a genomic correction was applied to remove amplification from contaminating DNA. In this study, the *ftsH, psbA, apx* and *sod2B* assays required a genomic correction. All probe and primer sequences were blasted against GenBank and against the *E. huxleyi* pan genome to confirm target specificity. Potential probes and primers were also examined for hairpins, self-dimers and hetero-dimers using IDT's OligoAnalyzer 3.1 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/). Probes were labeled with either FAM or TET fluorophores (Table 1) for use in multiplex reactions.

Quantitative RT-PCR optimization and multiplexing

Quantitative PCR was optimized for the RainDrop digital PCR system (RainDance Technologies, Billerica, MA, USA). To optimize the reactions, positive control amplicons for each assay were initially generated using 3 µl mixed cDNA from *E. huxleyi* CCMP374 infected and non-infected samples, 450 nM each forward and reverse primer (per assay), 1X FailSafe buffer™ E, and 1U FailSafe™ Enzyme Mix (Epicentre, Madison, WI, USA). Thermal cycling conditions were: 95ºC for 3 min, followed by 35

rounds of 95ºC denaturation for 30 s, 60ºC annealing for 30 s, and 72ºC extension for 30 s with a final extension at 72ºC for 2 min. Amplicons were visualized on a 2.2% agarose FlashGel™ (Lonza, Basel, Switzerland) to ensure that each PCR reaction produced only a single product of the correct size. Amplicons were purified with a GenElute™ PCR Clean-Up Kit (Sigma-Aldrich, St. Louis, MO, USA) and serial diluted to 10^{-8} in low EDTA molecular-grade TE buffer (Agilent). PCR reactions were optimized for two sets of 4-plex reactions. Initial real-time PCR optimization was conducted using a CFX96 Real-Time PCR Detection System (Bio-Rad) and the serial dilutions of positive control amplicons to determine the approximate target copy number in each dilution. Each real-time PCR reaction contained 1X TaqMan® Genotyping Master Mix (Life Technologies), 900 nM total target primer set, 200 nM target probe, and 2.5 µl diluted positive control template. Thermal cycling conditions included an initial activation/denaturation step at 95ºC for 10 min, followed by 55 cycles at 95ºC for 15 sec, then 60ºC for 1 min. Dilutions that produced Cq values between 25 and 30 were used for further optimization since they contained approximately 1,000 target copies μl^{-1}. Next, to test for potential inhibition from multiple targets and/or multiple assays, real-time PCR reactions were run for each permutation of 1) single assay and positive control template, 2) single assay and template mixture containing equal volumes of all 4 positive controls intended for the multiplex and 3) pairwise duplex assays (i.e. one TET and one FAM) with the 4-template mix. Reaction conditions proceeded as above, except in those samples receiving the template mix, where 10 µl total (2.5 µl of each) template was used. Any observed inhibition in fluorescence resulted in the removal of the assay from the planned multiplex.

Next, PCR reactions were optimized for the digital PCR (dPCR) using a RainDrop® digital PCR system (RainDance Technologies, Billerica, MA, USA). First, reactions were run on the dPCR system in duplex with TET and FAM assays to confirm no inhibition was present and to check that the template copy number determined through real-time PCR was consistent with the quantification provided in the digital platform. Each 25 µl reaction contained 1X TaqMan® Genotyping Master Mix (Life Technologies); 900 nM total target primer set and 200 nM target probe for each respective TET and FAM assay; 1X droplet stabilizer; and 10 µl positive control template mixture. PCR reactions were first processed by the RainDrop Source (RainDance Technologies), which generated 5 million droplets per sample. Amplification proceeded in a C1000 Touch deep-well thermal cycler (Bio-Rad) with the following thermal protocol: 95ºC for 10 minutes; 55 cycles of 95ºC for 15 s then 60ºC for 60 s with a ramping rate of 0.5ºC s^{-1}; 98ºC for 10 minutes; and hold at 8ºC. The samples were then processed on the RainDrop Sense (RainDance Technologies), which detects the fluorescence in each droplet and generates FACS-like dotplots containing populations of PCR positive and PCR negative droplets, respectively (Supplemental Fig. S1). PCR-positive populations were gated, spectrally corrected and analyzed using RainDropAnalyst 14.0.0.0 software (RainDance Technologies). A correction was applied for intact droplets, and counts for each assay were normalized to total droplets produced. No inhibition from duplex reactions was detected and positive control quantification was consistent with the real-time results. Next, duplex PCR reactions were repeated as described above but with pairwise FAM and TET assays under varying probe concentrations (100 and 200 nM respective probe, all permutations). This was done to determine the inherent brightness to dim ratio for each assay and allowed us to determine which FAM assay should be lowered to 100 nM and which should be maintained at 200 nM in the multiplex. Finally, reactions were multiplexed (high/ low FAM and high/ low TET, respectively) to confirm no inhibition under these conditions. Each 25 µl 4-plex reaction contained 1X TaqMan® Genotyping Master Mix (Life Technologies); 900 nM total target primer set and 200 nM target probe for high TET and FAM assays; 450 nM total target primer set and 100 nM target probe for low TET and FAM assays, 1X droplet stabilizer and 10 µl positive control template mixture. Multiplex setup was as follows (see Table 1): Multiplex A; HCF136 (high TET), *ctpA*(low TET), *sod2A* (low FAM), and a high FAM assay (data not presented here); Multiplex B; *sod2B* (high TET), *ftsH* (high FAM), and low TET and FAM assays (data not presented here). Some assays in the 4-plex setup investigated the expression of genes not pertinent to this study, as noted above. No significant reduction in PCR-positive droplet populations was observed when 4-plexed as compared to the duplex reactions. Multiplex reactions were rerun with a variety of

PCR enhancers (1-2 mM additional $MgCl_2$,1-2 M betaine, and 5% DMSO) to see if population resolution would improve, but ultimately the no-addition multiplex reaction conditions yielded the clearest resolution of populations and provided optimal conditions for the experimental templates. As the final step in the optimization process, 8-16 no-RT RNA templates were run in Multiplex A and B reactions to validate the template specificity of the assay. Multiplex B was designed to include all assays that were not mRNA specific and all resulted in positive amplification from no-RT RNA template. Although the *ftsH* assay was designed around a putative exon junction, no-RT RNA templates nonetheless produced amplification under the multiplex reaction conditions, so it was ultimately multiplexed with others requiring a genomic correction (Table 1).

Digital PCR Data Analysis

The data were analyzed in RainDropAnalyst (v 14.0.0.0; RDA) and displayed as 2-dimensional scatter plots. Positive reactions for each assay were seen as discreet clusters within the plots (Supplemental Fig. S1). PCR-positive populations were gated and analyzed using RainDropAnalyst 14.0.0.0 software as recommended by RainDance Technologies. The respective spectral correction derived from each multiplex optimization was applied to each set of reactions and intact droplets were enumerated based on gates derived from multiplexing optimization. Intact droplets for each assay were normalized to 5 million and then normalized to 1 ng input RNA based on Experion quantification (see RNA Purification in Materials and Methods). Relative expression (i.e. transcript copy number) was ultimately plotted for every 10 ng RNA to better visualize genes expressed at very low levels. This correction factor was used because 1) normalization to a housekeeping gene was non-optimal as expression of these genes does not remain constant during infection and 2) normalization to cell abundance was also problematic because cell loss after pelleting samples (prior to RNA extraction) was not consistent between samples. Infected cells produced a tighter, stickier pellet while the control pellet was much less stable. Genomic contamination and correction for Multiplex B was performed using 2 µl no-RT RNA template under the same reaction conditions and spectral correction as cDNA templates. Intact droplets for each assay were normalized to 5 million droplets, then to equivalent cDNA inputs. These values were directly subtracted from the cDNA populations at each given time point. In the rare instances where a genomic DNA subtraction produced a negative number (n=6), the value of expression for those replicates was set to 0.

References

McKew BA, Lefebvre SC, Achterberg EP, Metodieva G, Raines CA, Metodiev MV *et al* (2013). Plasticity in the proteome of *Emiliania huxleyi* CCMP 1516 to extremes of light is highly targeted. *New Phytol* 200: 61-73.

Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F *et al* (2013). Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature* 499: 209-213.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.

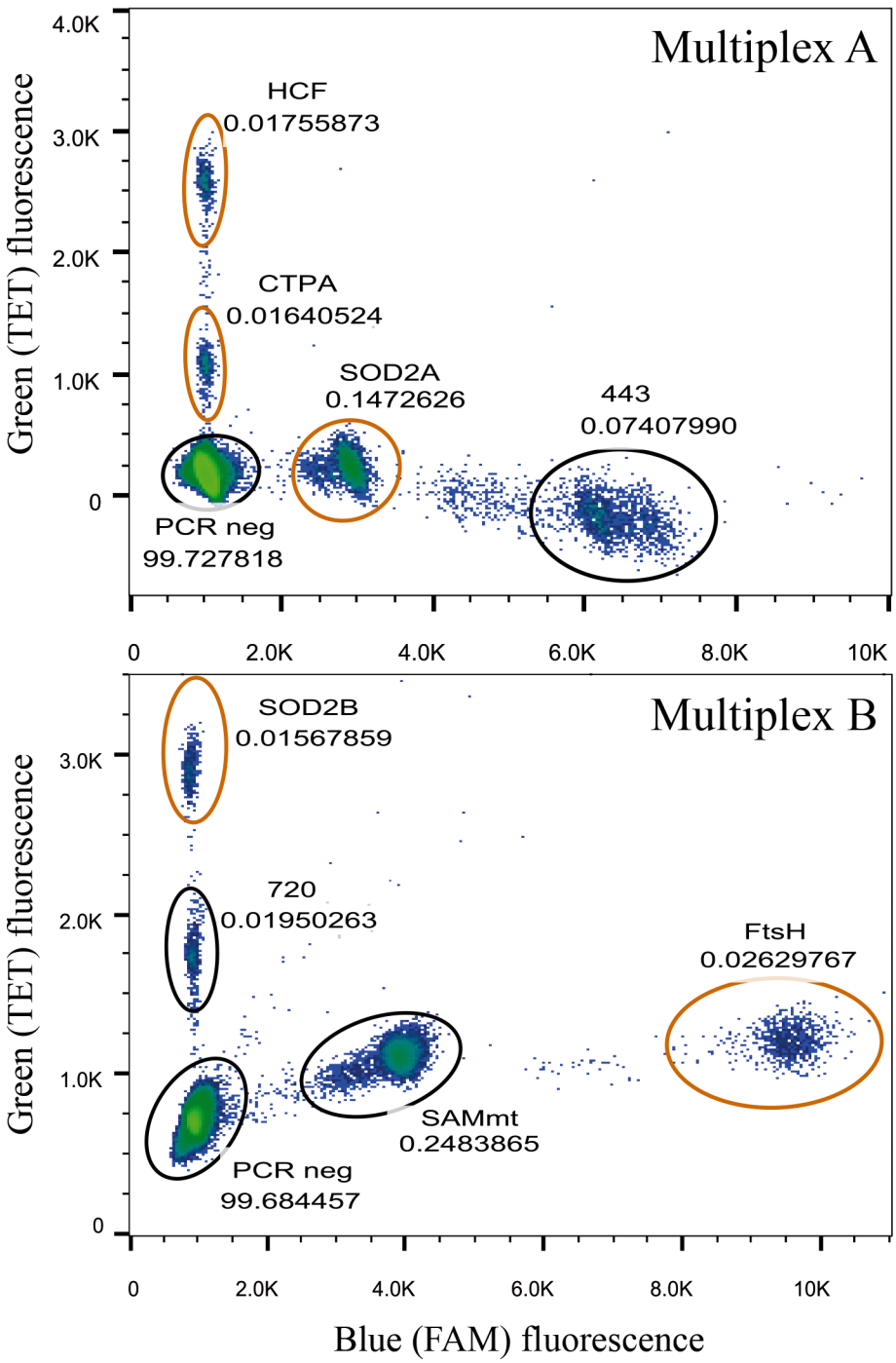

Figure S1. Examples of digital PCR (dPCR) populations from Multiplex A and B viewed as scatter plots. Populations gated with orange ovals represent PCR positive droplets from expression assays in this study. Numbers beneath gate titles show the percentage of droplets in a given population relative to droplet total. Cluster differentiation was achieved by varying probe concentrations. PCR negative populations (lower left) contain background fluorescence levels and hence represent droplets containing non-target templates. PCR positive populations gated in black represent multiplexed expression assays for genes not presented in this study.

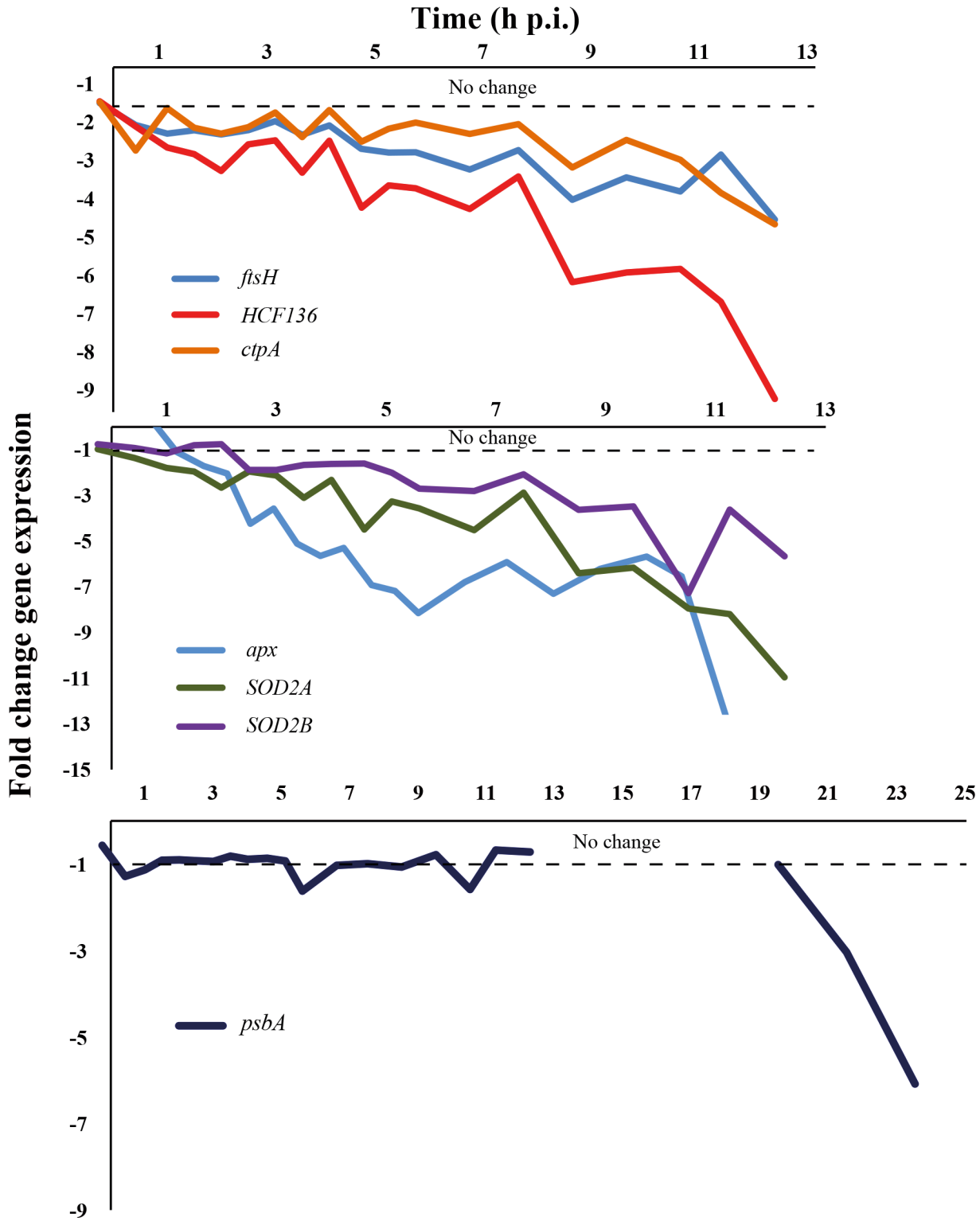

Figure S2. Fold changes in gene expression of infected cultures relative to control cultures over the first 12 hours post-inoculation with EhV-86. A fold change value of -1 represents no change in expression, < -1 represents down-regulation and > -1 represents upregulation, respectively. Corresponding absolute gene expression data are shown in Figures 3 and 4 in the main article.

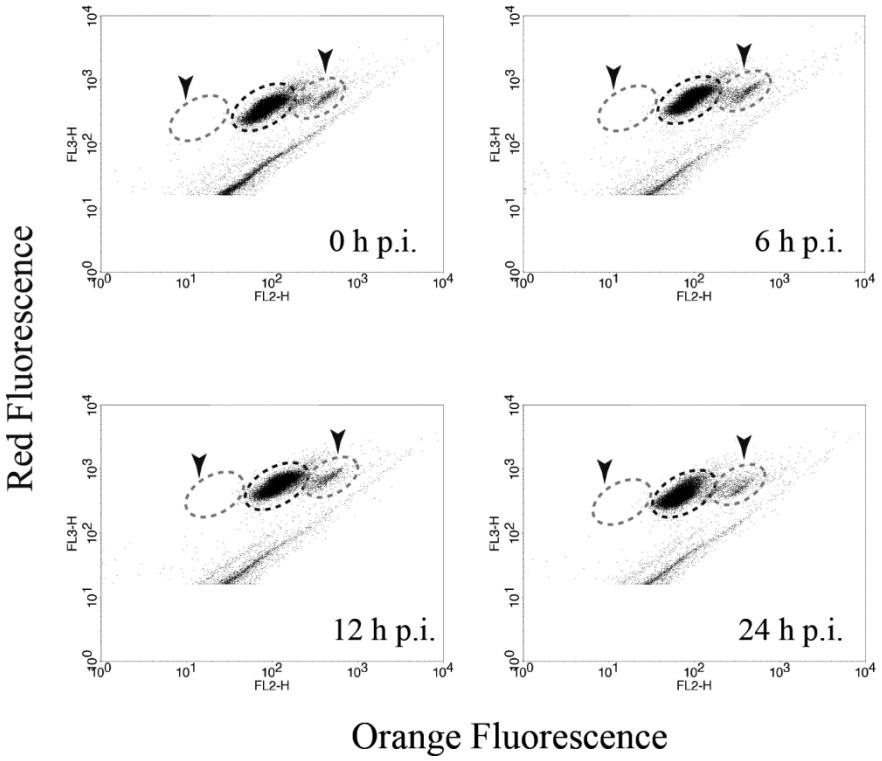

Figure S3. Control *E. huxleyi* FCM plots at selected time points with orange lipid fluorescence (FL2) plotted against red, chlorophyll fluorescence (FL3). Regions marked by grey dashed lines and arrows show the locations of visibly infected populations. As the infection progresses, changes in lipid distribution and composition cause cells to transition to very high FL2 and then to low FL2 (right and left grey gates with arrows, respectively; see Fig. 1C). Control cultures contain small subpopulations of cells with very high FL2, but unlike infected cultures, this subpopulation remains contstant over time. No low FL2 population was observed in the control cultures.