Supplemental Methods

Culture Growth and Experimental Design

Prochlorococcus MED4 was grown in the Pro99 seawater based medium amended with 10 mM HEPES (pH7.5) and 12 mM sodium bicarbonate at 21 °C under continuous white light at 10-25 µmol photon·m^{-1}·s^{-1} as described in Lindell et al.[12]. For experiments in which host gDNA was quantified and expression analyses carried out, the ratio of infective phage to host (the MOI) was 3.0 to maximize levels of infection. Phage at 3x10^{8} infective particles·ml^{-1} were added to 10^{8} cells·ml^{-1} and samples were collected each hour during the course of infection. Prior to phage addition for the expression experiment, cells were concentrated to 10^{8} cells·ml^{-1} by centrifugation. Experiments were carried out with triplicate independent cultures in paired experiments and control treatments were amended with filter-sterilized spent medium. For experiments in which the length of the lytic cycle and the timing of phage gDNA replication were determined, the ratio of infective phage to host was 0.1. Phage at 10^{7} infective particles·ml^{-1} were added to 10^{8} cells·ml^{-1} and allowed to adsorb for 1 h. The phage-cell mix was then diluted 100 fold and samples collected at different times after phage addition.

Quantification of phage particles and phage and host genomic DNA

Extracellular Phage Quantification

Phage particles from the extracellular medium were quantified using a quantitative PCR (qPCR) method. Samples were filtered over 0.2 µm sterile syringe filters (Tuffryn HT), and the filtrate, containing phage particles, was collected. To prevent PCR inhibition by the seawater based growth medium, the filtrate was diluted 20-100 fold in 10 mM Tris pH8, 10 µl of which was used in triplicate qPCR assays with the P-SSP7 specific DNA polymerase primers (see Suppl. Table 7 for primer sequences). Quantification was achieved using a standard curve of phage particles in 10 mM Tris pH8 that had been enumerated by epifluorescence microscopy after SYBR staining (see below).

This qPCR method was compared to standard methodology for determining phage numbers in the extracellular medium (Suppl. Fig. 5). The number of infective phages was determined by the Most Probable Number (MPN) assay[26]. Briefly, phage samples were serially diluted and added to exponentially growing MED4 cells in 96-well plates. The clearing of wells, as compared to control wells, was monitored using a Synergy HT Biotek fluorescence plate reader. The number of cleared wells at the appropriate phage dilutions was used to calculate the most probable number of infective phage in the undiluted sample. Total phage particles were enumerated using epifluoresence microscopy after DNA staining of the phage[27]. Briefly phage samples were filtered onto a 0.02 µm Anodisc (Whatman) filter with a vacuum of 7 in of Hg and allowed to dry. The filter was then stained with SYBR Green I (Molecular Probes), and the sample enumerated by epifluorescence microscopy after addition of anti-fade solution containing 0.1% p-phenylenediamine.

Intracellular phage and Prochlorococcus DNA quantification

Prochlorococcus cells were collected onto 0.2 µm pore-sized polycarbonate filters (Osmonics) by filtration at 8-10 in of Hg. The filters were washed 3 times with sterilized seawater to reduce the presence of extracellular phage, once with 3 ml preservation solution (10 mM Tris, 100 mM EDTA, 0.5 M NaCl; pH8) and then frozen at -80 °C. A heat lysis method was used to extract DNA from *Prochlorococcus* cells[28]. Briefly, the polycarbonate filter with *Prochlorococcus* cells was immersed in 650 µl of 10 mM Tris pH8, and agitated in a mini-bead beater for 2 min at 5000 rpm without beads. Five hundred µl of the sample was removed from the shards of filter and heated at 95 °C for 15 min. Ten µl was used in triplicate qPCR reactions. Phage DNA was amplified with P-SSP7 specific DNA polymerase primers, and *Prochlorococcus* DNA with *rbcL* primers (see Suppl. Table 7 for primer sequences).

Quantitative PCR protocol

Triplicate real-time PCR assays were carried out for each sample using Qiagen's QuantiTect SYBR Green PCR kit, primers at 0.3-1.0 µM and 10 µl samples (in 10 mM Tris pH8) in 25 µl volume reactions run on an DNA Engine Opticon (MJ Research). After 15 min denaturation at 95 °C, 40 amplification cycles were carried out as follows: Denaturation (95 °C for 15 sec); annealing (56 or 58 °C for 30 sec); elongation (72 °C for 30 sec); and fluorescence plate read (for quantification of SYBR green incorporation into double stranded DNA), and were followed by 5 min at 72 °C and melt curve analysis (read every degree from 50-90°C). Quantification of template was determined from standard curves produced with dilution series of P-SSP7 phage particles or *Prochlorococcus* MED4 genomic DNA. Melt curve analysis was used to verify that a single product was amplified.

RNA extraction

Samples were collected by centrifugation (12,400 Xg for 15 min at 20 °C), resuspended in storage buffer (200 mM sucrose, 10 mM sodium acetate pH5.2, 5 mM EDTA), snap frozen in liquid nitrogen and stored at -80 °C. Prior to RNA extraction the storage buffer was removed after spinning the cells for 2 min at 20,000 Xg at 20 °C. RNA was extracted using Ambion's *mir*Vana RNA isolation kit. DNA was removed by DNase I digestion using the Turbo DNA-*free* kit (Ambion). For microarray analysis 8 µg of the nucleic acid extract was digested with 6 U of Turbo DNase during a 60 min incubation at 37 °C followed by DNase I inactivation with inactivation slurry. The RNA was purified and concentrated by sodium acetate/ethanol precipitation. DNA removal was verified by gel electrophoresis. For RT-PCR analysis DNA was removed from 0.1-0.5 µg of the nucleic acid extract using the above procedure but without the precipitation step. DNA removal was verified by running no RT controls followed by qPCR for each sample (see RT-PCR validation of array results).

Array experimentation

Transcriptional analysis was carried out using a custom-made high density antisense Affymetrix array – MD4-9313. Synthesis of complementary DNA (cDNA), labeling, hybridization, staining and scanning was carried out according to Affymetrix protocols for *E.coli* (http://www.affymetrix.com/support/technical/manual/expression_manual.affx) with minor changes. Total RNA (2 µg) was denatured at 70 °C and annealed to random hexamer primers (25 ng/µl) at 25 °C for 10 min. The RNA was reverse transcribed to produce cDNA with Superscript II (25 U/µl – Invitrogen Life Technologies) and 0.5 mM dNTPs in the presence of 1 U/µl RNase Out RNase Inhibitor (Invitrogen). The mix was incubated at 25 °C for 10 min followed by 60 min incubations at 37 °C and 42 °C respectively. Superscript II was inactivated with a 10 min incubation at 70 °C. Sodium hydroxide (0.25 N) was used to remove RNA during a 30 min incubation at 65 °C, followed by neutralization with HCl. The cDNA was purified with MinElute PCR purification columns (Qiagen). Fragments of cDNA, 50-200 nt long, were produced from a 10 min incubation at 37 °C with DNase I (0.6 U per µg cDNA), followed by heat inactivation of the DNase I enzyme (10 min at 98 °C). The cDNA fragments were end-labeled with biotin using the BioArray Terminal Labeling Kit (Enzo) during a 60 min incubation at 37 °C. The reaction was stopped by freezing at –20 °C. The quality of biotin end-labeling was verified by gel-shift assays with NeutrAvidin (Pierce Chemicals) on 1% TBE agarose gels.

The cDNA was hybridized to the MD4-9313 custom Affymetrix array (see below for array description) in aqueous hybridization solution (100 mM MES, 1 M NaCl, 20 mM EDTA, 0.01% Tween-20, 0.1 mg/mL Herring Sperm DNA, 0.5 mg/mL BSA, 7.8 % DMSO and 3 nM prelabeled Affymetrix hybridization B2 oligo control probe mix) during a 16 h incubation at 45 °C in a GeneChip Hybridization Oven 320 rotating at 60 rpm. Washes and stains were carried out on a GeneChip Fluidics Station 450 (Affymetrix) following the ProkGE_WS2v3 Affymetrix protocol. Briefly, following two stringency washes the array was sequentially incubated with 10 µg/mL streptavidin (Pierce Chemical), 5 µg/mL biotinylated anti-streptavidin goat antibody (Vector Laboratories) and 0.1 mg/mL goat IgG (Sigma) and 10 µg/mL streptavidin-phycoerythrin conjugate (Mol. Probes) each for 10 min at 25 °C. After a final wash the arrays were scanned with the GeneChip Scanner (Affymetrix) at a 2.5 µm resolution with excitation set for 570 nm.

Array Design

The MD4-9313 (MD4-9313a520062) array is a custom-made high density antisense Affymetrix array. This array detects labeled cDNA that is antisense to the original RNA. It contains probes for the genomes of two cyanobacterial strains, *Prochlorococcus* MED4 and *Prochlorococcus* MIT9313[29], as well as two dsDNA phages that infect *Prochlorococcus* MED4 – the podovirus P-SSP7 and the myovirus P-SSM4[11]. For the *Prochlorococcus* genomes, the array contains probe sets to detect all predicted open reading frames with probe pairs approximately every 80 bases. For short open reading frames (ORF), the length of the gap was reduced to ensure a minimum of 11 probe pairs per ORF where possible. Probes were designed for intergenic regions longer than 35 bases and were spaced every ca 45 bases on both strands. For short intergenic regions the gap between probe pairs was reduced to ensure a minimum of 4 probe pairs where possible. For some short sequences, where insufficient high performance probes were designed using this approach, the best probes possible were designed with no regard for their spacing along the genome feature. For the phage isolates, probe pairs were designed across the genomes for both strands at an approximate interval of 90 bases. Probes are 25 bases long and each probe pair consists of a perfect match probe (identical to the sequence) and a mismatch probe (containing a single base change at the center of the probe).

Array Data Analyses

Normalization and Statistical Analyses

Data analyses were carried out in the statistical language R using several Bioconductor packages[30]. The array data were normalized and probe set summaries calculated from perfect match probe intensities in Affymetrix CEL files using quantile robust multi-array average (RMA) analysis[24] as implemented in the Bioconductor package *affy*[31]. See below for determination of appropriate normalization method for this experiment. Statistical significance of differentially expressed genes between infected and control cells at each time point was determined using the Bayesian t-test function implemented in the GoldenSpike[32] package (originally derived from the *hdarray* package) with the confidence level set to 9[32, 33]. The results are comparable to those obtained when RMAExpress Version 0.3 beta 1[24], Cyber-T[34] and Q-value[35] were used as stand-alone programs (data not shown). Control arrays at 4 and 8 h after infection gave the same expression profiles (data not shown) and were used for comparison with infected cells at 5, 6, 7 h after infection.

Clustering Analyses

Hierarchical clustering of phage genes was carried out using Pearson correlation and average linkage in the *stats* package in R. Input data was the average logged expression values of 3 biological replicates, standardized so that mean expression values for each gene equal zero and standard deviation equals one. The dendogram, visualized with Java TreeView[36], suggests the presence of several distinct expression clusters (Suppl. Fig. 1a). To determine the number of reliable clusters in the data, a resampling approach was applied[37] using the Bioconductor package *clusterStab*[38] whereby randomly selected sub-sets of genes are repeatedly clustered and the extent of similarity between the resulting clusters are examined. Reliable (or stable) clusters are those which repeatedly occur for the random sub-sets of genes. The similarity between clusters of different repeats was measured by the Jaccard coefficient ranging from zero (no similarity) to one (identical clustering). This resampling strategy was used for a range of number of clusters (k=2 to k=5) and the resulting distribution of Jaccard coefficients compared. If an adequate number of clusters is chosen, the distribution of coefficients will show an enrichment of values equal or close to one. A comparison of the histograms for the Jaccard coefficients strongly indicated the existence of three stable expression clusters for P-SSP7 genes (Suppl. Fig. 1b). Their temporal profiles are shown in Suppl. Fig. 1c. While this normalization methodology is the most appropriate for cluster analysis, we show temporal phage gene expression profiles in Fig. 2a using minimum-maximum normalized data as these more appropriately describe the dynamics of phage genome expression from a biological perspective.

The same strategy was used to determine the number of stable clusters for upregulated MED4 genes. Here the histograms indicate that two stable expression clusters exist (Suppl. Fig. 4).

Significance of co-expression of ‘bacterial-like’ phage genes

Clustering analysis indicated that the ‘bacterial-like’ phage genes *nrd, hli, psbA* and *talC* are temporally coexpressed (Suppl. Fig. 1, Fig. 2). To stringently assess the validity of this co-expression, two approaches were used. First, we assessed the reliability that the four genes are assigned to the same expression cluster (namely cluster 2) by a bootstrap approach using the Bioconductor package *hopach*[39] whereby genes are assigned to a particular cluster when only partial time series are used. Reliable cluster assignments should not depend on single data points and should therefore be found using only partial data. Thus, reliability can be examined by repeated bootstrap sampling and re-clustering of genes with subsequent calculation of cluster memberships. Cluster membership is defined here as the percentage of bootstrap samples that were assigned to the same original cluster. A cluster membership close to one indicates reliable assignment of a gene to the cluster. Applying this bootstrap approach, we detected high membership values for the phage genes *nrd, hli*, *psbA* and *talC* (1.000, 0.998, 0.9984 and 0.9999) for cluster 2 (Suppl. Fig. 2a). This strongly indicates that the 4 ‘bacterial-like’ phage genes are co-transcribed together within cluster 2 despite their spatial separation on the genome.

In addition, a regression approach was applied to determine the degree of certainty of co-expression of the four ‘bacterial-like’ genes (Suppl. Fig. 2b, 2c). In this analysis we wished to estimate the time points at which expression of each P-SSP7 gene changed from being non-expressed to expressed – termed here the switch time t*. If genes are co-regulated, we expect them to have the same time t* within acceptable confidence intervals. Here we defined the switch time t* as the time at which expression values reach half maximum values. We used the following procedure to estimate t*: After averaging expression values for the 3 biological replicates, values for each gene across the time series were normalized to a minimum value of zero and maximum value of 1. All P-SSP7 genes displayed sigmoidal expression patterns. To improve the fitting by sigmoidal curves, expression values across the time series were truncated to values ranging from 0.01 to 0.95 .The truncated expression values *y* were fitted to the sigmoidal function; y(t) = exp(a.t+b)/(exp(a.t +b) + 1), where a and b are fitting parameters and t is the time after infection. To allow fitting of the data by linear regression (which simplifies the calculation of confidence intervals), the data were transformed such that
y’ = log(y/(1-y)). Subsequently, linear regression given by y’= a.t + b, was performed and confidence intervals were calculated for each gene. Seeing as y=0.5 (half maximal expression) corresponds to y’=0, we can use the confidence intervals for y’ to assess whether the induction of genes occurred at the same time t*. The confidence intervals for *nrd* (020), *hli* (026), *psbA* (027) and *talC* (054) all overlap indicating that they are turned on simultaneously, together with the remaining genes in cluster 2 (Suppl. Fig. 2b). An example of the fitting procedure is illustrated in Suppl. Fig. 2c for the *nrd* gene.

Determination of Appropriate Normalization Method

Analysis of microarray data after implementation of various normalization methods showed differences in putative expression patterns, in particular for down-regulated genes (Suppl. Fig. 8). Therefore to ascertain which normalization method should be used for this dataset, normalized expression patterns for select genes were compared to those determined empirically with RT-PCR (see below for RT-PCR methodology). Normalization procedures tested were: RMA with quantile normalization at the probe level; RMA with normalization based on positive hybridization control spikes (AFFX-Bio* and AFFX-Cre*); Goldenspike which computes an expression summary based on 8 different normalization methods[32]; and Goldenspike without the second loess normalization at the summary level – as the assumption for this summary level normalization, that the majority of genes is not differentially expressed, may not hold for this experiment.

Comparisons were carried out on representative genes with the following expression patterns: (a) 1 unchanged, internal control gene; (b) 4 down-regulated genes and (c) 7 up-regulated genes. See Suppl. Table 5 for a list of the genes tested. (Note that PMM0550 and PMM1629 are considered both up- and down-regulated based on RMA quantile normalization.) See the "RT-PCR validation of array results" section below for more details of the genes chosen for RT-PCR validation. We compared the performance of the different normalization schemes to detect differential expression as validated by RT-PCR. These analyses show that RMA and the two versions of Goldenspike performed similarly for up-regulated expression, whereas differential expression patterns for down-regulated genes were best represented by RMA with quantile normalization (Suppl. Table 5, and see Suppl. Figs. 8, 9).

Initially, the superior performance of RMA with quantile normalization was somewhat surprising, seeing as it assumes similar overall distribution of probe intensities in different arrays, and we observed downregulation of a large number of genes. However it is important to note that a considerable percentage (25%) of the host MED4 genes were not significantly down-regulated at 8 h after infection. More importantly, however, is that the array used in this study includes a large number of probe sets other than for the host genes. It contains probe sets for intergenic regions, for the P-SSP7 phage genome and for an additional *Prochlorococcus* and phage strain. In fact, most probes on the microarray are not assigned to the organisms examined in our study. Therefore, most expression signals on the array are not expected to change and the underlying assumption of quantile normalization may hold. To assess this issue further, we compared the frequency of probe intensities for the whole array to the subset of intensities for MED4 genes after normalization. The density plots show that the signal distributions for the subset of MED4 probe sets are distinct for different arrays even though the overall distributions are similar for all arrays due to quantile normalization (Suppl. Fig. 10). Thus, quantile normalization can still be applied to our study as it did not erase differences in expression for the host genes.

RT-PCR validation of array results

Total RNA (2-5 ng) was reverse transcribed with Superscript II (Invitrogen) using 2 pmol gene-specific reverse primers in 20 µl reactions following the manufacturer's instructions. The resultant cDNA was diluted with 80 µl 10 mM Tris pH8, and 10 µl was used in each of 3 triplicate quantitative PCR reactions using Quantitect Sybr Green 2x kit (Qiagen) and 0.5-1.0 µM primers (see Suppl. Table 7 for primer sequences) in 25 µl reactions, such that cDNA resulting from 0.2-0.5 ng total RNA was used in each qPCR reaction (see above for quantitative PCR protocol). Results were further normalized to *rnpB* transcript levels which served as an internal control. No RT controls, carried out under identical conditions but without the reverse transcriptase enzyme, indicated that gDNA contamination was less than 1 % of the RT-PCR signal in all samples. Standard curves were carried out with MED4 gDNA or P-SSP7 phage particles. Expression levels from infected cells at different times after infection was compared to that for control cells. Significance of differential expression was determined from two-tailed t-tests.

Phage genes chosen for RT-PCR validation included the first gene of each expression cluster as well as the last gene in the genome (*talC*) as a representative of the 3 genes transcribed out of order on the genome. RT-PCR validation of host genes included downregulated and upregulated genes from both up-regulated expression clusters and were chosen to span low, medium and high array signal intensities as well as to include genes of potential biological interest where possible.

Promoter analysis

Bioinformatic analysis of phage P-SSP7 transcriptional signals

The computational prediction of bacterial promoters was based on a position specific weight matrix established for the -10 box of *Prochlorococcus* MED4 promoters[40]. Bacterial terminators were found by using the TransTerm algorithm[41], which detects rho-independent transcription terminators by searching for stem-loop-structures (inverted repeats) followed by a row of T's in the genome. Putative recognition sites for the phage RNA polymerase were searched *in silico* with a consensus sequence for T7 RNA polymerase allowing substitutions at positions, which are not common among all 47 natural phage promoters[42].

Experimental detection of 5' transcript ends

The 5' ends of mRNA transcripts from P-SSP7 and MED4 were mapped using the 5' Rapid Amplification of cDNA ends (RACE) technique, described previously by Bensing et al.[43] and modified for *Prochlorococcus* by Vogel et al.[40]. Briefly, 0.7-1.5 µg total RNA was used to cleave the 5' triphosphate, found in primary transcripts, with tobacco acid pyrophosphatase (TAP) (Epicentre, Madison, Wisconsin USA). The resulting 5' monophosphate was subsequently ligated, using T4 RNA ligase (Epicentre, Madison, Wisconsin USA), to the 3' hydroxyl group of an RNA oligonucleotide (5' adaptor: GAU AUG CGC GAA UUC CUG UAG AAC GAA CAC UAG AAG AAA). A gene-specific DNA primer (see Suppl. Table 8 for gene specific primer sequences) was used for reverse transcription followed by PCR amplification with a nested gene-specific primer and the 5' adaptor primer (ATA TGC GCG AAT TCC TGT AGA ACG AAC ACT AG). The amplification products were cloned and sequenced, and the first nucleotide downstream of the 5' adaptor RNA was assigned as the 5' end. This method enables the differentiation between transcription initiation sites of primary transcripts and RNA processed sites. For primary transcripts (carrying a 5' triphosphate) the TAP treated samples (TAP+) yield a specific or strongly enhanced amplification product relative to untreated samples (TAP-), whereas amplification products of equal intensity found for both TAP treated and untreated RNA samples are indicative of processed 5' ends that already carried a monophosphate at the 5' end.

Protein Analyses

Protein Extraction and Digestion to Peptides

Cells were collected by centrifugation and stored as described for RNA work, prior to lysis in 3 M urea, 0.05% SDS, and 50 mM Tris-HCl pH 8.0. Protein levels were quantified using the bicinchoninic acid method (Pierce). Samples were digested with sequencing grade trypsin (Promega) (protein:trypsin = 137.5:1) overnight at 37 °C, reduced with 10 mM DTT, alkylated with 50 mM iodoacetamide, and acidified to < pH 3.0 with HCl.

Peptide Fractionation and identification by Ion Trap Mass Spectrometry (MS)

Two phage infection time points (3 h and 7 h post infection) were subjected to comprehensive MS/MS sequencing experiments. For this purpose, each sample was adjusted to 25% acetonitrile and centrifuged to remove particulates. The entire sample was subjected to two-dimensional chromatographic fractionation (strong cation exchange followed by reversed phase) as in Jaffe et al. [44]. The eluate of the nano-flow reversed phase column was coupled directly to a LTQ linear ion trap mass spectrometer (ThermoElectron, Waltham, MA) where the top 10 most abundant MS ions were sampled for MS/MS sequencing in each scan cycle. Dynamic exclusion was employed to increase depth of coverage. In all, 60 Strong Cation Exchange (SCX) fractions were analyzed for each sample. The accumulated spectra were analyzed with SEQUEST[45], searching against a database of all predicted proteins from *Prochlorococcus* MED4 as well as the entire genomic sequence of cyanophage P-SSP7 prepared for proteogenomic mapping as in Jaffe et al.[44]. Criteria for valid spectra assignments and creation of proteogenomic maps for the phage were as in Jaffe et al.[44]. All validated peptides were considered to be potentially ‘present’ in subsequent analyses. It should be noted that detection of peptides is dependent on its ionization properties and on it being of suitable length, therefore the inability to detect a particular peptide using this methodology is not a definitive indication of the lack of its presence.

Phage Particle Purification for protein analysis

Prochlorococcus MED4 was infected with P-SSP7 and harvested once the culture had cleared. The cell lysate was centrifuged at 12,000 Xg for 30 min to remove unlysed cells and cellular debris. The supernatant was incubated for 30 min at 25 °C with 1 µg DNase I (Sigma) to degrade host gDNA from lysed cells. The salt concentration of the solution was brought up to 2 M with NaCl and incubated at 25 °C for an additional 30 min and then spun at 15,000 Xg for 30 min and the pellet discarded. Triton X-100 (0.1 % v/v final concentration) and PEG 8000 (10 % w/v final concentration) was added to the supernatant and stirred gently until fully dissolved and incubated overnight at 4 °C. Phages were collected by centrifugation at 12,400 Xg for 30 min at 4 °C and resuspended in Pro99 medium. Phage particles were purified on a cesium chloride step gradient (rho=1.4/1.6) prepared in 0.2 µm filtered seawater amended with 50 mM $MgCl_2$, 50 mM Tris pH8 and 0.1 % Triton X-100, and spun at 150,000 Xg for 2 hours. Purified phage particles were dialyzed in a step-wise fashion against 1 l of 1M NaCl, 50 mM $MgCl_2$, 50 mM Tris pH8 for 1 hour and twice for an hour each against 1 l of 100 mM NaCl, 50 mM $MgCl_2$, 50 mM Tris pH8.

Determination of proteins in purified phage particles

The equivalent of 10^{10} purified phage particles was subjected to proteomic analysis. The sample was digested for 18 hours at 37 °C with 0.2 μg of trypsin in a buffer consisting of 3M urea, 25 mM Tris pH 8.0, 25 mM $MgCl_2$, and 50 mM NaCl. The sample was reduced and alkylated as above, except that 5 mM DTT and 12.6 mM iodoacetamide were used. The sample was desalted using an Oasis HLB solid phase extraction column (Waters, 10 mg resin) according to the manufacturer's directions, reduced to dryness by vacuum centrifugation, and resuspended in 10 μl of 5% acetonitrile/5% formic acid. The sample was analyzed with 3 x 3μl injection to LCMS as above using a top 10 MS/MS method on an LTQ-FT mass spectrometer. Spectra were extracted and searched using SpectrumMill (Agilent, Palo Alto, CA) against the same hybrid database of phage and host proteins described above. Standard SpectrumMill autovalidation parameters were used to select confidently identified proteins and peptides.

Supplemental References

26. Taylor, J. The estimation of numbers of bacteria by tenfold dilution series. Journal of Applied Bacteriology 25, 54-61 (1962).
27. Noble, R. T. & Fuhrman, J. A. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. Aq. Microb. Ecol. 14, 113-118 (1998).
28. Zinser, E. R. et al. *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. Applied and Environmental Microbiology 72, 723-32 (2006).
29. Rocap, G. et al. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424, 1042-1047 (2003).
30. Gentleman, R. C. et al. Bioconductor: Open software development for computational biology and bioninformatics. Genome Biology 5, R80 (2004).
31. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy - analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307-315 (2004).
32. Choe, S. E., Boutros, M., Michelson, A. M., Church, G. M. & Halfon, M. S. Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. Genome Biology 6, R16 (2005).
33. Baldi, P. & Long, A. D. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. Bioinformatics 17, 509-519 (2001).
34. Long, A. D. et al. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Journal of Biological Chemistry 276, 19937-19944 (2001).
35. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences U S A 100, 9439-9445 (2003).
36. Saldanha, A. J. Java Treeview - extensible visualization of microarray data. Bioinformatics 20, 3246-3248 (2004).
37. Ben-Hur, A., Elisseeff, A. & Guyon, I. A stability based method for discovering structure in clustered data. Pacific Symposium on Biocomputing (2002).
38. Smolkin, M. & Ghosh, D. Cluster stability scores for microarray data in cancer studies. BMC Bioinformatics 4, 36-42 (2003).

39. Pollard, K. S. & van der Laan, M. J. Cluster Analysis of Genomic Data. *In:* Bioinformatics and Computational Biology Solutions Using R and Bioconductor; R. Gentleman, V. Carey, W. Huber, R. Irizarry, S. Dudoit (eds.) Springer, 209-229 (2005).

40. Vogel, J., Axmann, I. M., Herzel, H. & Hess, W. R. Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. Nucleic Acids Research 31, 2890-9 (2003).

41. Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O. & Salzberg, S. L. Prediction of transcription terminators in bacterial genomes. J. Mol. Biol. 301, 27-33 (2000).

42. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. Biochemistry 39, 10419-10430 (2000).

43. Bensing, B. A., Meyer, B. J. & Dunny, G. M. Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. Proceedings of the National Academy of Sciences U S A 93, 7794-7799 (1996).

44. Jaffe, J. D., Berg, H. C. & Church, G. M. Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics 45, 59-77 (2004).

45. Eng, J. K., McCormack, A. L. & Yates, J. R. r. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. Journal of the American Society for Mass Spectromotry (1994).

46. Steglich, C., Futschik, M., Rector, T., Steen, R. & Chisholm, S. W. Genome-wide analysis of light sensing in *Prochlorococcus*. Journal of Bacteriology 188, 7796-7806 (2006).

47. Dunn, J. J. & Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166, 477-535 (1983).

Supplementary Table 1: Detection of phage proteins inside the host cell during infection (infection) or in the purified phage particle (virion). Number of unique peptides detected: 1 = +; 2-4 = ++; 5-10 = +++; more than 10 = ++++. Transcription cluster designations as per Fig.2 and Suppl. Fig. 1.

ORF ID	Gene Name – Product	Infection	Virion	Transcription cluster
PSSP7_001	unknown	++		Cluster 1
PSSP7_002	unknown	++		
PSSP7_003	unknown	++		
PSSP7_004	unknown	++		
PSSP7_005	unknown			
PSSP7_006	unknown	+++		
PSSP7_007	unknown			
PSSP7_008	unknown			
PSSP7_009	unknown	+		
PSSP7_010	unknown			
PSSP7_011	*gene 0.7* – MarR family of transcriptional regulators			
PSSP7_012	*int* – integrase	+		
PSSP7_013	*gene 1* – RNA polymerase	++		Cluster 2
PSSP7_014	*gene 2.5* – ssDNA binding protein	+++		
PSSP7_015	*gene 3* – endonuclease	+		
PSSP7_016	*gene 4* – primase/helicase	++++		
PSSP7_017	*gene 5* – DNA polymerase	+++		
PSSP7_018	unknown	++		
PSSP7_019	*gene 6* – exonuclease	+++		
PSSP7_019A	unknown	++		
PSSP7_020	*nrd* – ribonucleotide reductase	+++		
PSSP7_020A	unknown	+		
PSSP7_020B	unknown	+		
PSSP7_021	unknown	+		
PSSP7_022	unknown	++		
PSSP7_023	unknown	++	+++	
PSSP7_024	*gene 8* – head-to-tail connector	+++	++++	
PSSP7_025	*gene 9* – capsid assembly protein (scaffolding protein)	+++	+	
PSSP7_026	*hli* – high-light inducible protein	+		
PSSP7_027	*psbA* – D1 photosystem II reaction center protein	++		
PSSP7_028	Unknown			Cluster 3
PSSP7_029	*gene 10* – capsid protein	++++	++++	
PSSP7_030	*gene 11* – tail tubular protein A	+	++++	
PSSP7_031	*gene 12* – tail tubular protein B	++++	++++	
PSSP7_032	unknown (putative *gene 13?)*			
PSSP7_033	unknown (putative *gene 14*? – internal core protein)	++	++++	
PSSP7_034	*gene 15* – internal core protein	+++	++++	
PSSP7_035	*gene 16* – internal core protein	++++	++++	
PSSP7_036	*gene 17* – tail fiber	++++	++++	
PSSP7_037	unknown	++	+	
PSSP7_038	unknown	+	+++	
PSSP7_039	unknown	+	++++	
PSSP7_040	unknown	+		
PSSP7_041	unknown		+	
PSSP7_042	unknown		+	
PSSP7_043	unknown			
PSSP7_044	unknown	+		
PSSP7_045	unknown			
PSSP7_046	unknown		+++	
PSSP7_047	unknown			
PSSP7_048	unknown		++	
PSSP7_049	possible endonuclease			
PSSP7_050	unknown	++	++++	
PSSP7_051	*gene 19* – DNA maturase	+		
PSSP7_052	unknown			Cluster 2
PSSP7_053	unknown			
PSSP7_054	*talC* – transaldolase family protein	+++		

Supplementary Table 2: Promoter analyses: predictions and experimental detection of 5’ ends upstream of the denoted ORF. Promoter analysis: nd – not determined. None = tested but no 5’ end found. Processed 5’ end = product found in both TAP- and TAP+ treatments. Motif = conserved motif found in proximity of processed 5’ end.

ORF ID	Product	Bioinformatic Predictions	Experimental detection of 5’ ends
PSSP7_001	unknown	Bacterial -10 box: 52..57	Processed 5’ end=113 motif=91..116; Processed 5’ end=44686; motif=44664..44689
PSSP7_002	unknown		nd
PSSP7_003	unknown	Terminator: 1675..1691	Processed 5’ end=1672; motif=1650..1675
PSSP7_004	unknown		Same site as per PSSP7_003
PSSP7_005	unknown		nd
PSSP7_006	unknown		nd
PSSP7_007	unknown	Bacterial -10 box (x2): 2591..2596; 2629..2634	nd
PSSP7_008	unknown		Nothing conserved; Processed 5’ ends (5x)=2444; 2446; 2450; 2451; 2459; Nothing conserved; Processed 5’ ends (3x)= 2724; 2725; 2726
PSSP7_009	unknown		nd
PSSP7_010	unknown		nd
PSSP7_011	*gene 0.7* - MarR transcriptional regulator	Bacterial -10 box (x3): 3566..3571; 3577..3582, 3585..3590	none
PSSP7_012	*int* - phage related integrase		nd
PSSP7_013	*gene 1* - RNA polymerase	Terminator: 5005..5028 Bacterial -10 box: 5034..5039	Bacterial tis=5045 Non-Processed (x2) 5’ ends=5032; 5060
PSSP7_014	*gene 2.5* - ssDNA binding protein		none
PSSP7_015	*gene 3* - endonuclease		nd
PSSP7_016	*gene 4* - primase/helicase		nd
PSSP7_017_018	*gene 5* - DNA polymerase		Nothing conserved; ; Processed 5’ ends (3x)=9690; 9692; 9695
PSSP7_019	*gene 6* - exonuclease		none
PSSP7_019a	unknown		nd

PSSP7_020	*nrd* - ribonucleotide reductase domain	Possible -10 box (13160..13165), identical to that found experimentally for rpl21[40].	Nothing conserved;
			Processed 5’ end=12818;
			Non-Processed 5’ end=12928

PSSP7_020a	unknown		nd
PSSP7_021	unknown		nd
PSSP7_022	unknown		nd
PSSP7_023	unknown		nd
PSSP7_024	*gene 8* - head-to-tail connector		none
PSSP7_025	*gene 9* capsid assembly protein		nd
PSSP7_026	*hli* - high-light inducible protein		none
PSSP7_027	*psbA* - D1 photosystem II reaction center protein		none
PSSP7_028	unknown	Bacterial -10 box: 19430..19435	none (no signal found further upstream of Processed 5’ end: 19749)
PSSP7_029	*gene 10* - capsid protein		Processed 5’ end: 19749 motif: 19725..19750
PSSP7_030	*gene 11* - tail tubular protein A	Terminator: 21031..21046	none
PSSP7_031	*gene 12* - tail tubular protein B		nd
PSSP7_032	Unknown (*gene 13??)*	Terminator: 24626..24650	none
PSSP7_033	unknown (*gene*14??)		nd
PSSP7_034	*gene 15* - internal core protein		nd
PSSP7_035	*gene 16* - internal core protein		nd
PSSP7_036	*gene 17* - tail fiber		none
PSSP7_037	unknown		nd
PSSP7_038	unknown		nd
PSSP7_039	unknown		nd
PSSP7_040	unknown		nd
PSSP7_041	unknown		nd
PSSP7_042	unknown		nd
PSSP7_043	unknown		nd
PSSP7_044	unknown		nd

PSSP7_045	unknown		nd
PSSP7_046	unknown		nd
PSSP7_047	unknown		nd
PSSP7_048	unknown		nd
PSSP7_049	possible endonuclease		nd
PSSP7_050	unknown	Terminator: 39402..39417	none
PSSP7_051	*gene 19* - DNA maturase	Terminator: 41165..41176	none
PSSP7_052	unknown	Bacterial -10 box (x2): 42978..42983; 42988..42993	none
PSSP7_053	unknown		nd
PSSP7_054	*talC* - transaldolase		nd
		Terminator: 44063..44075	

Supplementary Table 3: Upregulated *Prochlorococcus* MED4 genes determined from microarray analysis. Fold change (infected/control) with time (h) after infection. Positive and negative values indicate an increase and decline in transcript levels respectively. Significant increases in fold change are shown in blue and the level of significance is shown: * for q<0.05; ** q<0.01; ***q<0.001.

ORF – gene name, possible product and function				**Fold**	**Change**	**(inf/ctrl)**			
TRANSCRIPTION GROUP 1	**0 h**	**1 h**	**2 h**	**3 h**	**4 h**	**5 h**	**6 h**	**7 h**	**8 h**
PMM0549 – *csoS1* carboxysome shell protein 1, carbon fixation	1.70	**1.31****	-1.22	-1.59	-1.76	-2.07	-2.03	-2.16	-2.58
PMM0550 – *rbcL* rubisco large subunit, carbon fixation	1.79	**2.01*****	**1.38****	**1.18***	-1.41	-1.69	-1.62	-2.00	-2.05
PMM0551 – *rbcS* rubisco small subunit, carbon fixation	1.63	**1.85*****	**1.36****	**1.17***	-1.43	-1.67	-1.62	-2.04	-2.01
#PMM0815/PMM1396 – *hli19/09* high-light inducible stress response protein	1.15	**1.23***	-1.27	-1.45	-1.25	-1.38	-1.45	-1.58	-1.39
#PMM0816/PMM1397 – *hli18/08* high-light inducible stress response protein	1.09	**1.29****	-1.12	-1.26	-1.26	-1.50	-1.56	-1.64	-1.56
#PMM0817/PMM1398 – *hli17/07* high-light inducible stress response protein	1.16	**1.35*****	-1.17	-1.19	-1.24	-1.47	-1.46	-1.64	-1.60
#PMM0818/PMM1399 – *hli16/06* high-light inducible stress response protein	1.14	**1.33****	-1.18	-1.24	-1.28	-1.53	-1.49	-1.66	-1.70
PMM0970 – *urtA* urea ABC transporter periplasmic binding protein	1.34	**1.43*****	1.01	-1.24	-1.51	-1.77	-1.59	-1.75	-1.52
PMM1135 – *hli14* high-light inducible stress response protein	1.24	**1.26****	-1.17	-1.26	-1.54	-1.76	-1.73	-1.88	-1.94
PMM1483 – *rpoC2* RNA polymerase subunit, transcription	1.01	**1.26***	-1.02	-1.28	-1.49	-1.59	-1.61	-1.67	-1.67
PMM1536 – *rps11* ribosome small subunit protein 11, translation	1.35	**1.19***	-1.09	-1.35	-1.80	-1.94	-1.95	-2.05	-2.09
PMM1544 – *rpl6* ribosome large subunit protein 6, translation	1.02	**1.23***	-1.05	-1.44	-1.70	-1.74	-2.04	-1.86	-1.97
PMM1545 – *rps8* ribosome small subunit protein 8, translation	1.07	**1.21***	-1.19	-1.39	-1.68	-1.86	-1.93	-1.85	-1.90
PMM1546 – *rpl5* ribosome large subunit protein 5, translation	-1.06	**1.33****	-1.09	-1.33	-1.82	-1.94	-2.22	-1.94	-1.94
PMM1549 – *rps17* ribosome small subunit protein 17, translation	-1.13	**1.22***	-1.17	-1.40	-1.97	-2.09	-2.15	-2.10	-1.98
PMM1629 – *rpoD type II* alternative sigma factor, transcription	1.10	**1.61*****	-1.09	-1.34	-1.38	-1.74	-1.80	-1.84	-1.92
TRANSCRIPTION GROUP 2	**0 h**	**1 h**	**2 h**	**3 h**	**4 h**	**5 h**	**6 h**	**7 h**	**8 h**
PMM0014 – *dus* tRNA dihydrouridine synthase, RNA modification	1.20	-1.28	**1.23***	**1.21***	**1.44***	**1.48****	1.06	**1.49****	1.21
PMM0030 – unknown	-1.10	-1.15	1.09	**1.66****	**1.41***	1.08	-1.41	1.00	1.00
PMM0334 – unknown	1.11	-1.49	-1.07	1.05	1.20	**1.26***	-1.11	1.04	-1.17
#PMM0368 – unknown	1.02	-1.05	**1.21***	**1.70*****	**1.85****	**1.61*****	**1.68*****	**1.70*****	1.32
PMM0426 – *sun* tRNA and rRNA methyltransferase, RNA modification	1.09	-1.66	-1.48	-1.20	**1.48***	**1.41****	1.23	**1.81*****	**1.49***
#PMM0684 – unknown (homologous to PMM0819 and PMM1134)	1.01	-1.06	**1.27***	**1.58*****	1.35	1.16	1.08	1.07	-1.16
#PMM0685 – unknown (homologous to PMM1427)	-1.05	-1.36	**1.77*****	**2.65*****	**2.37****	**2.01*****	**1.44***	**1.49***	**1.39***
#PMM0686 – *clpS-like* protease adaptor, protease inhibition and redirection	-1.07	-1.48	**3.91*****	**8.95*****	**14.00*****	**11.71*****	**8.55*****	**10.82*****	**8.75*****
#PMM0819 – unknown (homologous to PMM0684 and PMM1134)	1.28	-1.01	**1.52*****	**2.16*****	**2.32****	**1.84*****	**1.66*****	**1.57*****	1.21
PMM0830 – *DHPS-like* folate biosynthesis, nucleotide & amino acid synthesis	1.16	-1.84	-1.69	-1.38	**1.38***	**1.36****	1.10	1.18	-1.12

PMM0936 – *umuD* SOS response to DNA damage	1.01	-1.55	1.31*	1.39***	1.75**	1.53***	1.15	1.43**	1.10
PMM1114 – unknown	1.26	-1.81	-1.05	-1.18	1.22	1.29*	1.05	1.04	1.02
PMM1115 – *crtH*, phytoene dehydrogenase, secondary metabolite biosynthesis	1.01	-1.42	-1.15	-1.07	1.26	1.17*	-1.08	1.00	-1.14
PMM1187 – AAA ATPase family, protein turnover, stress response	1.01	-1.39	1.09	1.21*	1.56*	1.45**	1.00	1.27*	1.15
#PMM1201 – dTDP-D-glucose 4,6-dehydratase, cell envelope biogenesis	1.27	-1.63	-1.43	-1.16	1.22	1.27**	1.02	1.17	-1.08
#PMM1248 – unknown	1.08	-2.27	-1.76	-1.43	1.37*	1.60***	1.12	1.33*	1.25
PMM1284 – *phoH-like* phosphate stress ATPase	1.13	-1.34	1.06	1.11	1.80**	1.56***	1.12	1.26*	1.07
#PMM1403 – HNH nuclease domain, site-specific endonuclease	1.11	-1.88	-1.56	-1.60	1.24	1.45**	1.11	1.56**	1.51*
#PMM1426 – unknown	1.47	-2.50	-2.05	-1.82	1.16	1.49**	1.18	1.31*	1.36
#PMM1427 – unknown (homologous to PMM0685)	1.00	-1.20	1.23*	1.71***	1.93**	1.74**	1.52**	1.54**	1.35
PMM1428 – unknown	1.02	-1.27	-1.03	1.03	1.60**	1.50**	1.16	1.32*	1.04
PMM1501 – *rne* RNase E, mRNA degradation	1.01	-1.16	2.44***	3.51***	3.43***	2.91***	1.83***	2.09***	1.72*
PMM1502 – *rnhB* RNase HII, DNA replication and repair	1.13	-2.16	1.47*	2.25***	3.55***	2.91***	1.64**	2.52***	2.09**
PMM1517 – unknown	-1.06	-1.58	-1.32	1.03	1.39*	1.31*	1.10	1.06	1.15
PMM1529 – *prfA* peptide release factor, translation	1.31	-1.11	1.10	1.52***	1.67**	1.29**	-1.04	1.17	-1.10

#Genes found in genome islands as per Coleman et al.[6].

Supplementary Table 4: High-light inducible genes (*hli*) in *Prochlorococcus* MED4 and their expression patterns during exposure to environmental stressors.

MED4 Gene IDs	High Light Stress	Nitrogen Stress	Phage Infection
PMM0093 - *hli01*			
PMM0064 - *hli02*			
PMM1482 - *hli03*			
*PMM1118 - *hli04*	+		
* #PMM1404 - *hli05*	+		
* #PMM0818/PMM1399 - *hli06/hli16*	+		+
* #PMM0817/PMM1398 - *hli07/hli17*	+		+
* #PMM0816/PMM1397 - *hli08/hli18*	+		+
* #PMM0815/PMM1396 - *hli09/hli19*	+		+
* #PMM1390 - *hli10*		+	
* #PMM1385 - *hli11*	+		
* #PMM1384 - *hli12*	+		
PMM1317 - *hli13*			
*PMM1135 - *hli14*	+		+
* #PMM1128 - *hli15*	+	+	
PMM0471 - *hli20*			
* #PMM0690 - *hli21*	+	+	
* #PMM0689 - *hli22*	+	+	

*Clusters with phage *hli* genes as per Lindell & Sullivan et al.[5]; #Found in genome islands as per Coleman et al.[6]. High-light stress[46], Nitrogen stress[23], Phage infection (this study).

Supplementary Table 5: Comparison of array normalization methods to RT-PCR. Significance is assigned to differentially expressed genes determined by RT-PCR (p-values) normalized to *rnpB* and microarray analysis (q-values) normalized using different methods (Quantile RMA; Hyb Ctrl;.Golden Spike (GS) regular[32]; GS without 2nd Loess)

Expression	ORF	Time	RT-PCR	RMA	Hyb Ctrl	GS w/	GS w/o
Pattern	*gene*					2nd loess	2nd
						(regular)	Loess
Unchanged	PMM_rnpB	0	0.528	0.912	0.759	0.991	0.978
	rnpB	1	0.530	0.818	0.674	0.419	0.795
		3	0.938	0.805	0.158	0.614	0.550
		4	0.878	0.489	0.712	0.884	0.986
		8	0.970	0.756	0.281	0.627	0.903
DownRegulated	PMM0496	0	0.171	0.529	0.231	0.995	0.998
	rpoD	4	0.000	0.003	0.470	0.539	0.123
		8	0.001	0.003	0.043	0.966	0.100
	PMM0627	0	0.422	0.337	0.150	0.655	0.769
	pcb	4	0.004	0.001	0.449	0.572	0.208
		8	0.000	0.000	0.007	0.152	0.048
	PMM1309	0	0.034	0.405	0.072	0.813	0.803
	ftsZ	4	0.000	0.008	0.602	0.703	0.113
		8	0.000	0.000	0.003	0.453	0.011
	PMM1629	0	0.195	0.231	0.123	0.889	0.895
	rpoD type II	1	0.388	0.000 up	0.636	0.000 up	0.044
	(up- and down-	3	0.000 dn	0.001 dn	0.005	0.457	0.057
	regulated)	8	0.000 dn	0.001 dn	0.054	0.991	0.042
UpRegulated	PMM0550	0	0.041 up	0.974	0.000 up	0.022 up	0.016 up
	rbcL	1	0.066	0.000 up	0.283	0.000 up	0.000 up
	(up- and down-	3	0.317	0.010 up	0.021	0.000 up	0.170
	regulated)	8	0.269	0.000 dn	0.051 dn	0.809	0.050
	PMM0684	0	0.095	0.999	0.200	0.992	0.989
	unknown	3	0.049	0.000	0.000	0.000	0.005
		4	0.001	0.088	0.223	0.000	0.171
		8	0.075	0.269	0.154	0.000	0.702
	PMM0686	0	0.467	0.337	0.231	0.801	0.624
	clpS-like	4	0.006	0.000	0.000	0.000	0.000
		8	0.000	0.000	0.000	0.000	0.000
	PMM0819	0	0.599	0.719	0.101	0.373	0.476
	unknown	4	0.000	0.002	0.035	0.000	0.003
		8	0.013	0.164	0.014	0.000	0.230
	PMM0936	0	0.439	0.193	0.416	0.889	0.921
	umuD	1	0.276	0.000 dn	0.041	0.251	0.038 dn
		3	0.001	0.000	0.005	0.006	0.059
		4	0.005	0.005	0.101	0.003	0.075
		8	0.001	0.497	0.350	0.459	0.843
	PMM1284	0	0.002	0.486	0.105	0.707	0.765
	phoH-like	4	0.002	0.002	0.037	0.003	0.111
		8	0.738	0.569	0.336	0.256	0.754
	PMM1501	0	0.451	0.297	0.403	0.993	0.988
	rne	3	0.025	0.000	0.000	0.000	0.000
		8	0.026	0.019	0.036	0.027	0.160

p determined from 2-tailed t-test for RT-PCR results and q determined using Cyber-T and Q-value for microarray results.

Supplementary Table 6: Mass spectrometric detection of previously unannotated proteins from the P-SSP7 phage genome. Detected tryptic peptides are in bold and underlined. The entire ORF inferred from these peptide detections are shown. For PSSP7_033, detection of the peptide suggests an N-terminal extension of the previously annotated protein (shown in italics).

GENE ID	PROTEIN SEQUENCE
PSSP7_19A	MTTRKK**NQSFGPPPPITK**LTTEQDFKLRQLEILLSKPETRKEDIAIVMIALQEQAFVLSNCIKNLIEKWPKPPTTTDPR**TTNEVPLMFGILLETK**DSDFTSET
PSSP7_20A	MKYLGEVVRTVTVPAFYTLILITPILLTSCKSKDKNSHGLNDVWTSPENSGLIQKLEQR**KQLYKELLGETSGSTK**
PSSP7_20B	**METESIQTSVLR**FTCPHAERASYSTLICQPVVSATYEKVCVKVCQICASSIVGQGLKNLESILHQISTGKLDSDS
PSSP7_033	MCLGAAAKAANENARRRYKYENERRER**NW*MQTMSIYNAQK****VKYDEDVQNAGLAQAQVKTDQQEAMDLARGEAQIKYAELFRKLLNDSTYGKLVASGQTGQSTRRRATMDYAKYGRDVSDIARRLTLNDRELARKSSEQISKYKQFKDEAFKVAFQPIPDVAPPQPVMRNVGAEAFMGALSIASNVATMGGQSGFGWWGG*

Supplementary Table 7: Primers used for RT-PCR verification of microarray results and normalization methods for representative phage and host genes.

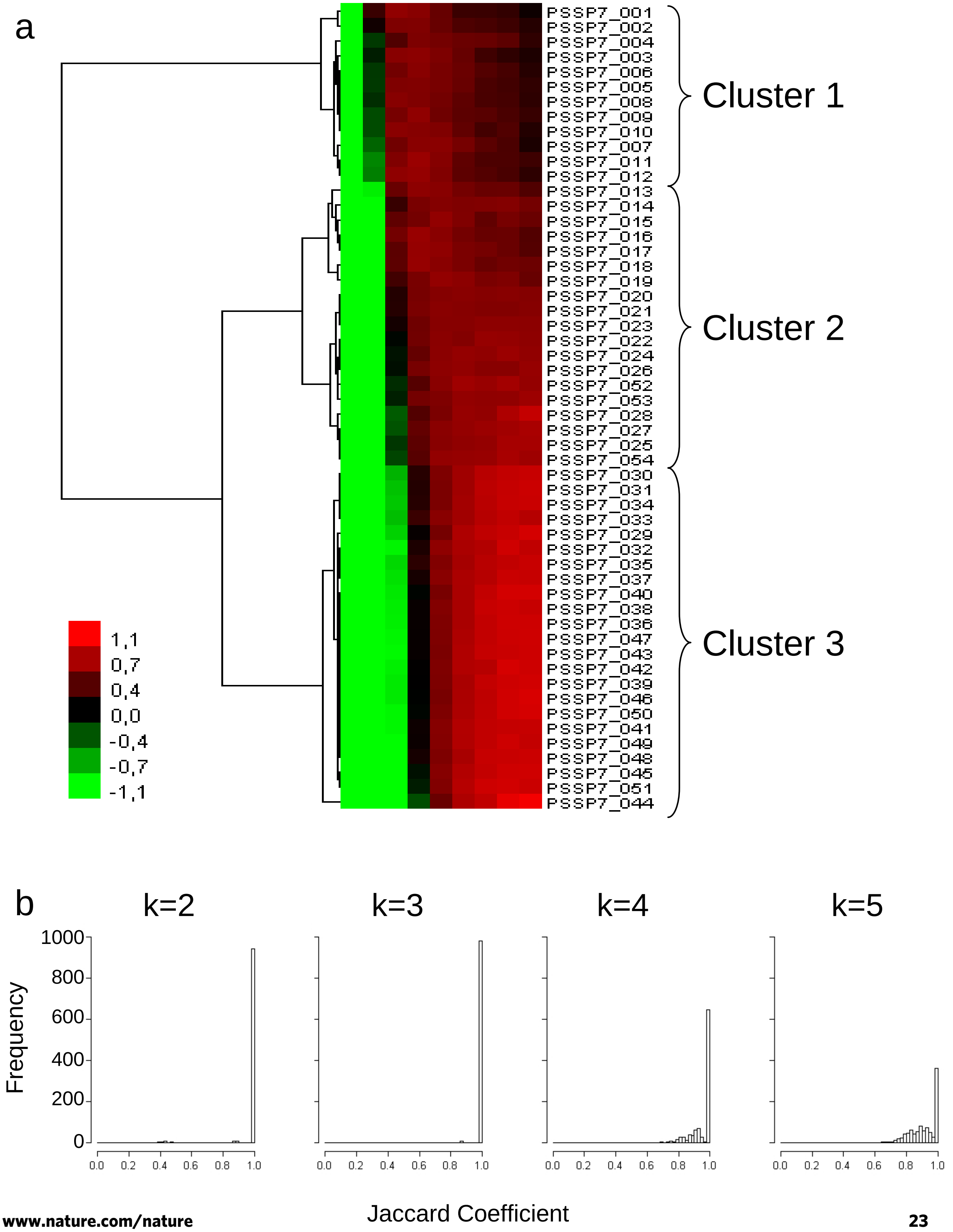
Target ORF ***gene* – product**	**Primer Direction**	**Primer Sequence 5' – 3'**
Phage P-SSP7 genes		
PSSP7_001 Unknown	F R	CCAAGCCAAAGGCTACACAT GCATCCCTTGATTCATTGCT
PSSP7_003 Unknown	F R	ATGGTTCACTTCCTAACCAAGC CCCCCTTACCCATAGGTGTT
PSSP7_013 *gene 1* – RNA polymerase	F R	CGACTATGGAGGAGCGGTTA GTCTGCTGCTTCCCAATCTC
PSSP7_017 *gene 5* – DNA polymerase	F R	AAACACTTCCGCCCTTACCT CTGCAACGAAAGGGAATTGT
PSSP7_020 *nrd* – ribonucleotide reductase	F R	TTGTGCAAGCTCCATAGTCG GCCTTACCAAACTCGGCATA
PSSP7_027 *psbA* – D1 photosystem II protein	F R	CTCTGCTATGCACGGAAGTT GCAGATTCCCATGGAGGTAA
PSSP7_029 *gene 10* – capsid protein	F R	GGCTTCCAGCATGAAACAAT TGGTCTTCTCTGCAACTGGA
PSSP7_054 *talC* – transaldolase family protein	F R	TGGTCGAAAATACGGAGAGG TACGTAGCACCAGCATGAGC
Host *Prochlorococcus* MED4 genes		
PMM_rnpB *rnpB* – RNA of RNase P	F R	TTGAGGAAAGTCCGGGCTC GCGGTATGTTTCTGTGGCACT
PMM0496 *rpoD* – principle RNA polymerase sigma factor	F R	AATCAGAGCTGCCGAAAATA TGATCTGCTATCGCTCGTGT
PMM0550 *rbcL* – rubisco large subunit	F R	CCTGAATATGTCCCCCTCGA CCGCTGCTGCAACTTCTTCT
PMM0627 *pcb* – chlorophyll a/b binding protein	F R	TCATGTCGCTCATGCAGGG GACCCATTGGGACACTGGG
PMM0684 Unknown	F R	CGCAAGGCAGCTTTTTAATC TCCATGTTTCAAACGCAGAG
PMM0686 *clpS-like* – protease adaptor	F R	CAGTTGTAGATCCAAAGACAACG CAAGACAATTTGCTACGTGTTCA
PMM0819 Unknown	F R	CCCAAGTGGTTGGCTTCTTA ATCCCAGGCTTTTTCCAAAT
PMM0936 *umuD* – SOS response to DNA damage	F R	GTGATTCGGTCTCAGCAGGT TTCTCCATCTATCATCGCAATA
PMM1284 *phoH-like* – phosphate stress induced ATPase	F R	GTTTGTGCCGCCAGATTATT TGCTAATGGTGCGACTTCAA
PMM1309 *ftsZ* – cell division protein	F R	AATGACTGAAGCTGGCACTGC ACTATTCATTGCGGCTTGAGC
PMM1501 *rne* – RNase E	F R	AACCGCCTAGCACAGGATTA TGCTTTTTCGAGAGCGATTT
PMM1629 *rpoD type II* – alternative sigma factor	F R	GAGTTGCCCGAAGATGATGT ACATTGGCTCATCTCCATCC

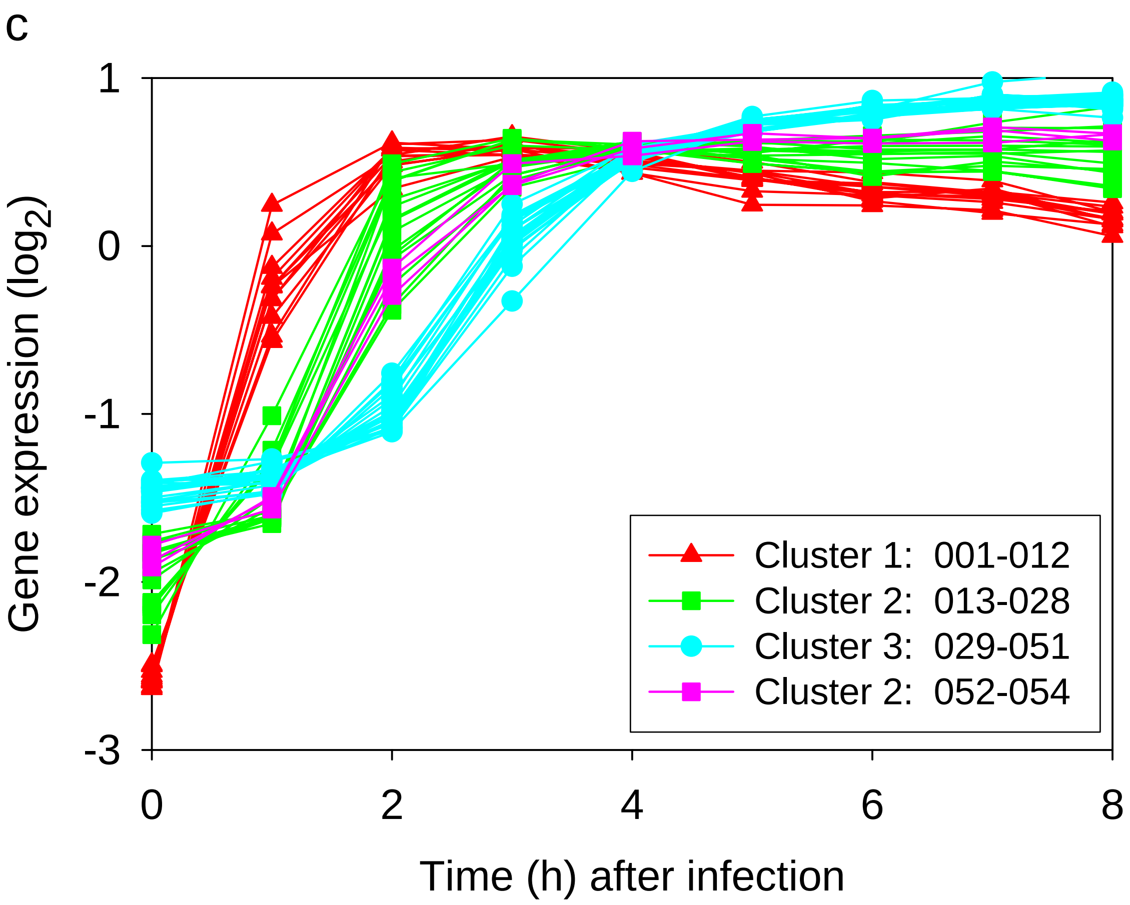
Supplementary Table. 8. Primers used in 5' RACE analysis for phage P-SSP7 and host *Prochlorococcus* MED4 genes. Primers used for reverse transcription are designated by “rt”, whereas those used for nested or second nested PCR are designated by “nest” and “nest2” respectively in the oligonucleotide name. “up” – the primer was designed upstream of the gene.

Gene	Oligo. Name	Oligonucleotide Sequence (5‘–3‘)	Length
Phage P-SSP7 genes			
PSSP7_001	P001rtREV	TCTCCATAATTGACCCGCTT	20
	P001nestREV	CTTAATAAAGTCAGTCAGTCCATCCCAGTCAGT	33
	P001nest2rev	TGATGGGAGAGGAATTGAACCTCTCGAT	28
PSSP7_003	P003nestREV	CCTTCTTACCGAATTTTCCTATGGTGTACTTAG	33
	P003rtREV	TCTCCCCCTTACCCATAG	18
PSSP7_004	P004nestREV	GGAATCGTTTACAGGGAATAACTCAGGCTCG	31
	P004rtREV	TTGTGCTTGGGTTTCTCTCA	20
PSSP7_008	P008nestREV	CGAAGGCTTCATAAGGCATCGAAGGAAAG	29
	P008rtREV	GGTATCTGATAGCCATAAATCTT	23
PSSP7_011	P011nestREV	CAGTCAGTATTACGGCTACCACTTGCGC	28
	P011rtREV	TCAATCTATGAAACTCACTGAG	22
PSSP7_013	P013rtREV	TCTTTACGTGGGGAGAATAG	20
	P013nestREV	GGTTATCTTTGCAGTTATAGCTCCTTGCGATTCTG	35
PSSP7_014	P014nestREV	CTATAAACTTACCTTCTTCTACCTCCTCCCATG	33
	P014rtREV	CTGGAGGACGCTTATCTTC	19
PSSP7_017	P017nestREV	CAGCATGGGTGAGCCAATGCAGAGC	25
	P017rtREV	ACAGGTAAATCGTAGCCAATAAT	23
PSSP7_019	P019nestREV	CTTAAGTTCCTGTATGACACGTTTGTATCCAC	32
	P019rtREV	CATCTGCTTCTAGAGTATCTC	21
PSSP7_020	P020rtREV	GTCCTGATGCAACGAGAG	18
	P020nestREV	CCCTTATTTGTTCTTGTTCCCGCCGGTC	28
PSSP7_020 up	P020nest2REV	CCGAGGTGAAATCCGAGTCCTTGGTC	26
	P020rtREV	ACCCTGCTCTGCATGTGT	18
PSSP7_024	P024nestREV	GGAGGTAGAACTGCGAGCATGAGCTTC	27
	P024rtREV	CCTAACTTGTCATCACGTAC	20
PSSP7_026	P027nest2REV	CAGCCATTAAATCTTTCTGCTTCTGGTGACATTAG	35
PSSP7_027	P027nestREV	CCTATTGCATTGGAGCTTGGAACTACTGC	29
	P027rtREV	CGGCTTCCCAGATCGG	16
PSSP7_029 up	P029nest2REV	GGACTCGCACCTCCCTGATCGG	22
PSSP7_029	P029rtREV	GCCTTGTCAGCATTACCC	18
	P029nestREV	GGAAGGAACTTGTCATACGACCTGTGTAG	29
PSSP7_030	P030nestREV	GCCATCCTTCACCCTGTACATCTTTGTTTG	30
	P030rtREV	TCTGGGGTAACAAGTACATG	20
PSSP7_032	P032nestREV	CATCTTCCTGTACGCCGGCTACTCC	25
	P032rtREV	CGGGTGTACATAGCATCC	18
	P032nest2rev	CTCCATAAGCGGCACATAGTGGGATTACC	29
PSSP7_036	P036nestREV	GTCGAGTTCAACTTTGACATCGACATTCGC	30
	P036rtREV	GGTGTAGTCATTATTTGTTTGAC	23
PSSP7_050	P050nestREV	GTCCATTATTTCAGGGTTAGCTTTTTGTGCAGG	33
	P050rtREV	ACCTTTCCATCTCATTTTAGAGA	23
PSSP7_051	P051nestREV	GGCTTGAATCTGGAGTCTCTTGGGTCC	27
	P051rtREV	CCAAGATTTACCAACACCTC	20
PSSP7_052	P052rtREV	CCTTTGCGTTAAAGATGCTG	20
	P052nestREV	CTTCCATCTCATCCAGGTAGTCCTCAATAGC	31
Host MED4 genes			
PMM0368	PMM0368nestREV	CTATTGCCCAACCATTAGCTCCCTGAGG	28
	PMM0368rtREV	GCTGACACCACTGCCAAA	18
PMM0684	PMM0684nestREV	CTGCCTTGCGGGTTAAGTATCCAGCC	26
	PMM0684rtREV	TTGTAAATAAGAGATGGGTATAAAC	25
PMM0819	PMM0819nestREV	TGACTTGATGACTTGATTGCTGAGGGCTC	29
	PMM0819rtREV	CATTTATCCCAGGCTTTTTCC	21
PMM1500	PMM1500nestREV	CCATCCATATCCTTGCTCTGGGCCG	25
	PMM1500rtREV	GATGAGATTTGACACCCTC	19

PMM1501	PMM1501nestREV	CTATAAAGGCAGCATCAATACCTGGTAGGAC	31
	PMM1501rtREV	GGACCTAGATCTGATACATG	20

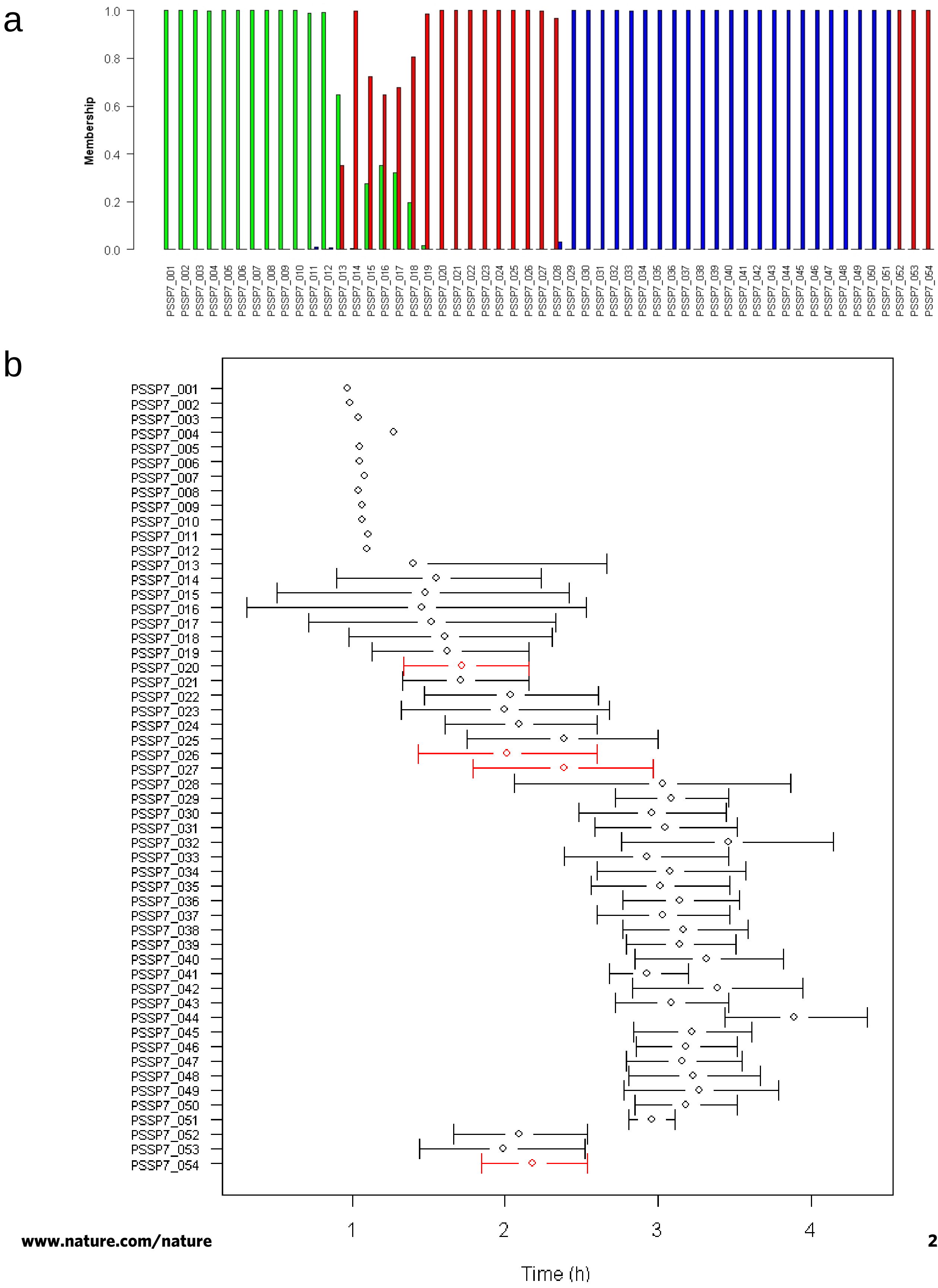
Supplementary Figure 1

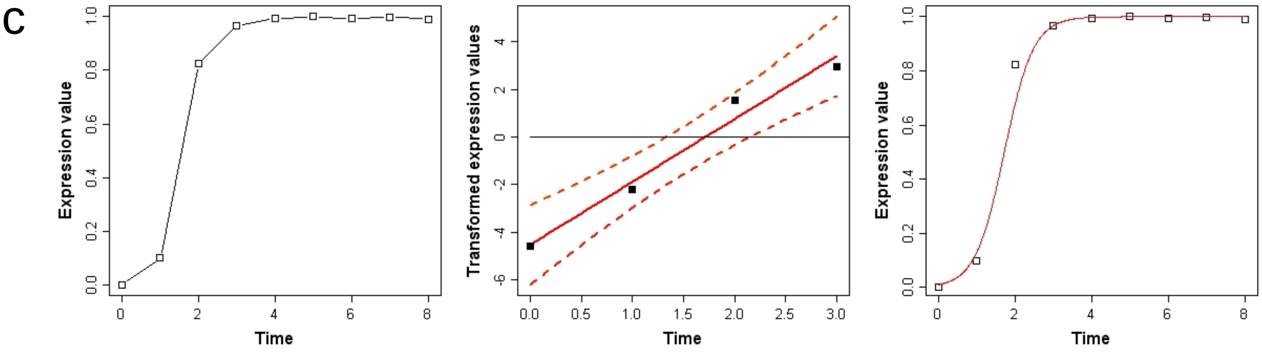




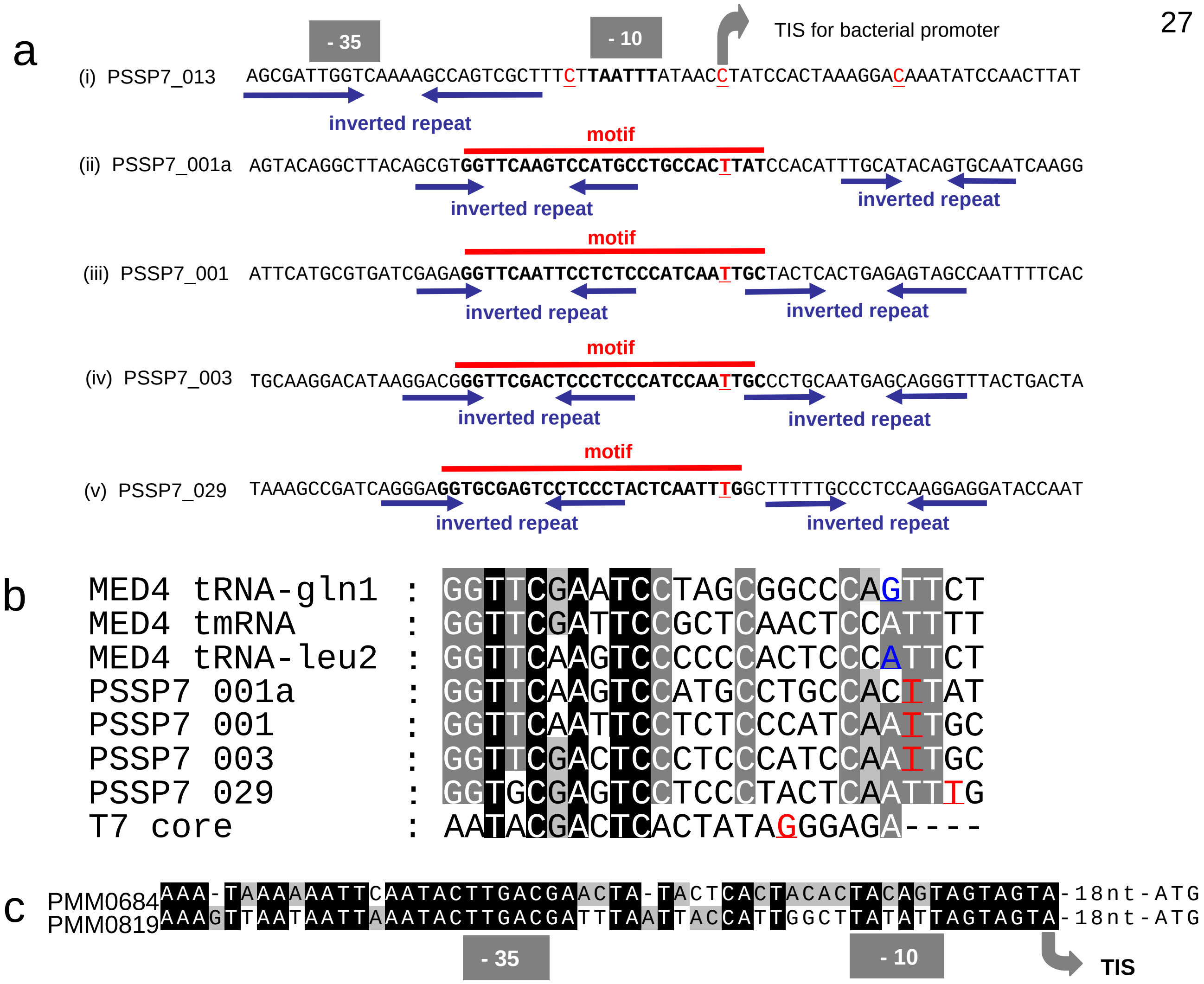
Supplementary Figure 1. Cluster analysis of phage gene expression profiles. (a) Hierarchical clustering of phage gene expression profiles, after standardization of logged data (mean expression equal to zero and standard deviation equal to one) was performed with average linkage and Pearson correlation. (b) Distribution of Jaccard coefficients derived from 1000 random independent resamplings of phage genes. The proportion of genes used for resampling was 0.7. The number of clusters (k) tested ranged from 2 to 5. Average linkage and Pearson correlation was used for the hierarchical re-clustering. For k=3 clusters, 982 out of 1000 Jaccard coefficients equaled 1 indicating that phage genes form three stable clusters. (c) Temporal profiles of the 3 clusters detected by hierarchical clustering. See Figure 2a for a representation of these temporal profiles after minimum-maximum normalization. Note that the last 3 genes in the genome cluster together with genes from cluster 2 and are transcribed prior to genes in cluster 3. See Suppl. Fig. 2 for statistical analysis of the significance of the clustering of all 4 ‘bacterial-like’ genes in cluster 2. See Suppl. Table 1 for gene name and function for each ORF.

Supplementary Figure 2

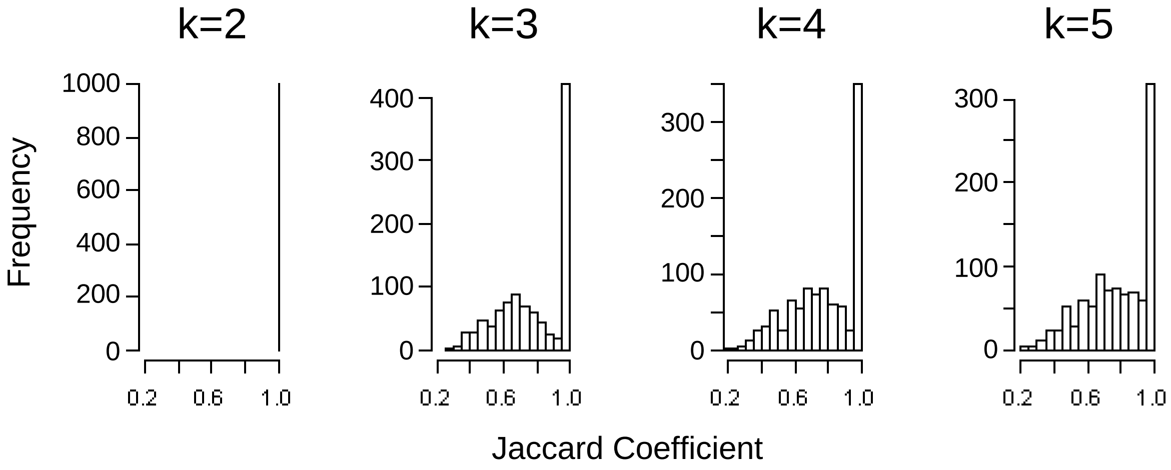




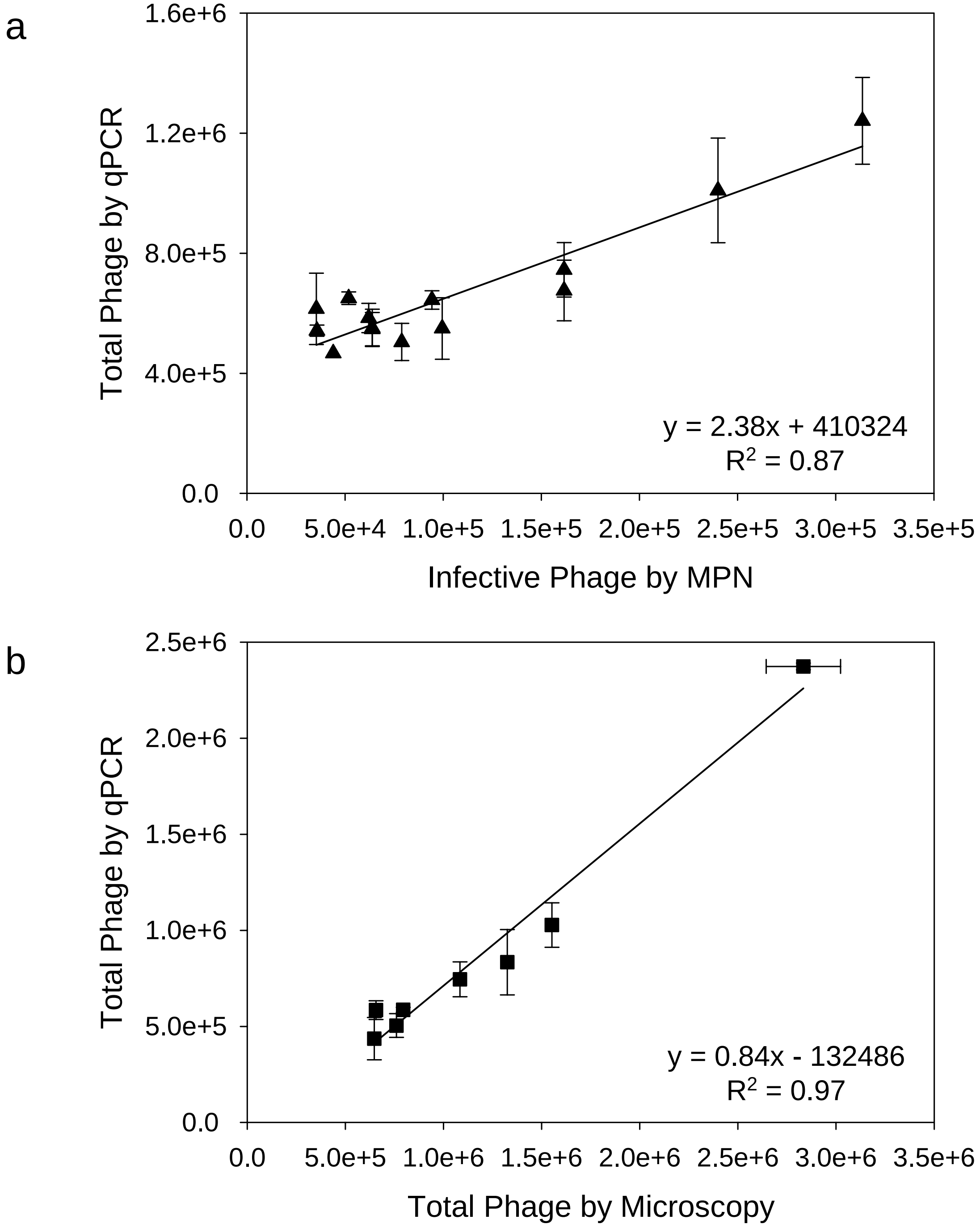
Supplementary Figure 2. Significance of the temporal coexpression of the last 3 genes of the genome with genes in cluster 2 and therefore of the 4 ‘bacterial-like’ genes *nrd* (020), *hli* (026), *psbA* (027) and *talC* (054). (a) Cluster membership for genes in cluster 1 (green), cluster 2 (red) and cluster 3 (blue) were based on 10000 independent bootstrap samplings with replacement of expression values for the same gene. (b) 90% confidence intervals are shown for the switch time (t*) at which transcription of the phage genes went from being non-expressed to expressed – defined here as the time point at which 50% of maximal expression was reached. Confidence intervals for the 4 ‘bacterial-like’ genes are shown in red. Note that no intervals could be derived for ORFs 1-12 due to immediate initiation of expression. (c) An example of the fitting procedure carried out (for *nrd*) to determine the confidence intervals shown in (b). The left panel shows the expression data (y) after normalization so that minimum expression equals zero and maximal expression equals 1. The middle panel shows the linear regression (solid red line) and the confidence intervals (dashed red lines) determined after the transformation y’=log(y/(1-y)). Note time t* is derived as the time point for which the regression line crosses y’=0 (set at half maximum expression). The regressed sigmoidal curve is shown in the right panel. The reliable assignment of the last 3 genes on the genome with genes in expression cluster 2 as well as the overlap of their confidence intervals provides strong evidence for the temporal coexpression of the 4 ‘bacterial-like’ genes in cluster 2 despite their spatial separation on the genome.


a
- 35
- 10
TIS for bacterial promoter
(i) PSSP7_013 AGCGATTGGTCAAAAGCCAGTCGCTTTCTTAATTTATAACCTATCCACTAAAGGACAAATATCCAACTTAT
inverted repeat
motif
(ii) PSSP7_001a AGTACAGGCTTACAGCGTGGTTCAAGTCCATGCCTGCCACTTATCCACATTTGCATACAGTGCAATCAAGG
inverted repeat
inverted repeat
motif
(iii) PSSP7_001 ATTCATGCGTGATCGAGAGGTTCAATTCCTCTCCCATCAATTGCTACTCACTGAGAGTAGCCAATTTTCAC
inverted repeat
inverted repeat
motif
(iv) PSSP7_003 TGCAAGGACATAAGGACGGGTTCGACTCCCTCCCATCCAATTGCCCTGCAATGAGCAGGGTTTACTGACTA
inverted repeat
inverted repeat
motif
(v) PSSP7_029 TAAAGCCGATCAGGGAGGTGCGAGTCCTCCCTACTCAATTTGGCTTTTTGCCCTCCAAGGAGGATACCAAT
inverted repeat
inverted repeat
b
MED4 tRNA-gln1 : GGTTCGAATCCTAGCGGCCCAGTTCT
MED4 tmRNA : GGTTCGATTCCGCTCAACTCCATTTT
MED4 tRNA-leu2 : GGTTCAAGTCCCCCCACTCCCATTCT
PSSP7 001a : GGTTCAAGTCCATGCCTGCCACTTAT
PSSP7 001 : GGTTCAATTCCTCTCCCATCAATTGC
PSSP7 003 : GGTTCGACTCCCTCCCATCCAATTGC
PSSP7 029 : GGTGCGAGTCCTCCCTACTCAATTTG
T7 core : AATACGACTCACTATAGGGAGA----
c
PMM0684 AAA-TAAAAAATTCAATACTTGACGAACTA-TACTCACTACACTACAGTAGTAGTA-18nt-ATG
PMM0819 AAAGTTAATAATTAAATACTTGACGATTTAATTACCATTGGCTTATATTAGTAGTA-18nt-ATG
- 35
- 10
TIS

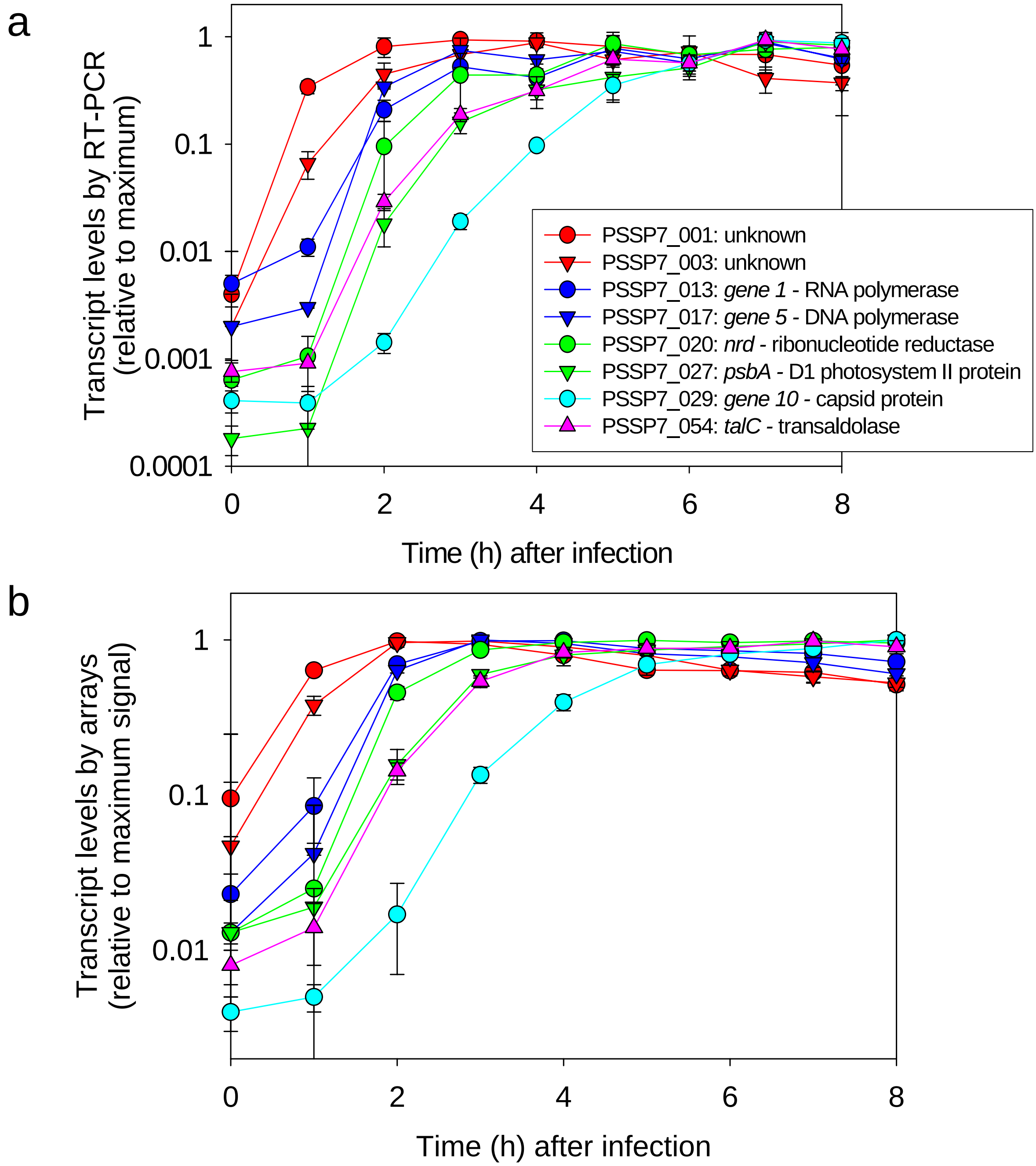
Supplementary Figure 3. RACE mapping of 5' transcript ends for phage and host genes and associated possible regulatory elements. (a) Experimentally mapped 5' ends for phage genes. A typical bacterial promoter was found upstream of the 5'ends of the cluster 2 gene PSSP7_013 coding the RNA polymerase (i). The 5' ends upstream of cluster 1 and 3 genes – ORF PSSP7_001 (2 transcripts), PSSP7_003 and PSSP7_029 – were found in both TAP+ and TAP- treatments (data not shown) suggesting that they are mature transcripts that have undergone post-transcriptional processing (ii to v). The nt at the 5' ends are shown in red and underlined. A 26 nt conserved motif (bold type face and marked with a red line above the sequence) was found in the region of these processed 5' ends, as were inverted repeats (blue arrows below the sequence). (b) Comparison of the conserved motif found upstream of processed transcripts in (a) shows that they have weak similarity to the T7 core promoter. A bioinformatic search for sequence similarity between this motif and the MED4 host genome revealed similarity to putative 3' cleavage sites of 2 tRNA genes and tmRNA, suggesting that the inverted repeats associated with this motif may serve as an RNase III recognition site in a similar fashion to that known for T7 with linked promoter and processing sites[47]. The red underlined nucleotide indicates the 5' end determined by RACE and for the T7 core promoter it indicates the known transcript start. The annotated 3' end of the tRNAs is blue and underlined. (c) Transcription initiation sites for two up-regulated MED4 homologous genes of unknown function (PMM0684 and PMM0819). The consensus bacterial regulatory elements (the -10 and -35 box) are shown upstream of the transcription initiation site.



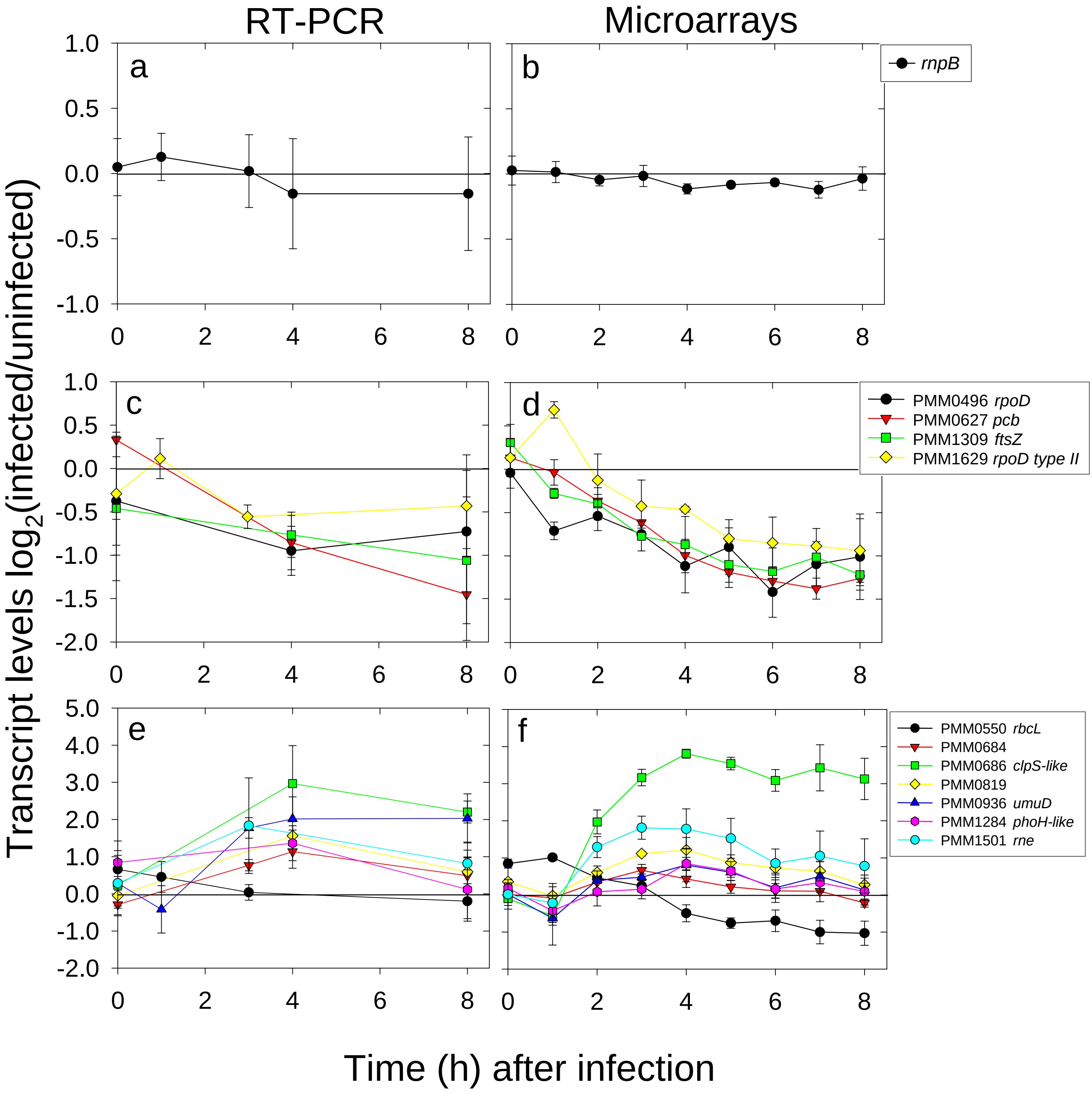
Supplementary Figure 4. Analysis of the number of stable clusters of upregulated host (MED4) genes. The distribution of Jaccard coefficents was derived from 1000 independent random samplings of upregulated host genes. The number of clusters (k) tested ranged from 2 to 5. The proportion of genes used for resampling was 0.7. For hierarchical re-clustering average linkage and Pearson correlation was used. For k=2 clusters the coefficients are concentrated at 1 indicating that upregulated host genes form two stable clusters.



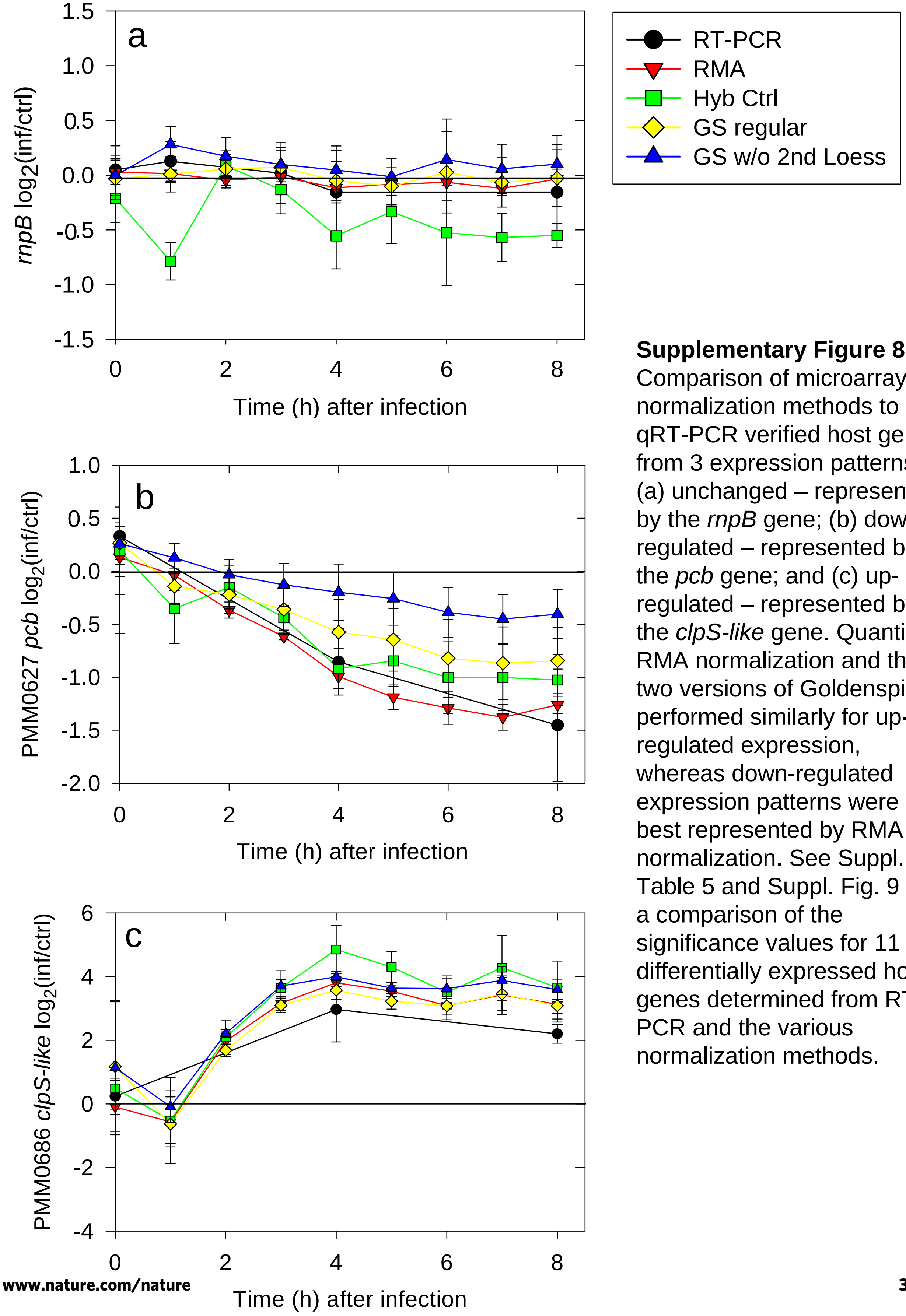
Supplementary Figure 5. Comparison of extracellular P-SSP7 quantification using a quantitative PCR (qPCR) assay for the phage DNA polymerase gene with (a) infective titer determined by the most probable number (MPN) assay and (b) total phage particles after staining with the DNA SYBR Green I stain and enumerated by epifluorescence microscopy. The linear regression for (a) is y = 2.38x + 410324, R^2 = 0.87; and (b) is y = 0.84x – 132486, R2 = 0.97. Note that qPCR quantification provides close to a 1:1 ratio with the SYBR stained particles, but was 2.5 fold higher than infective phage.



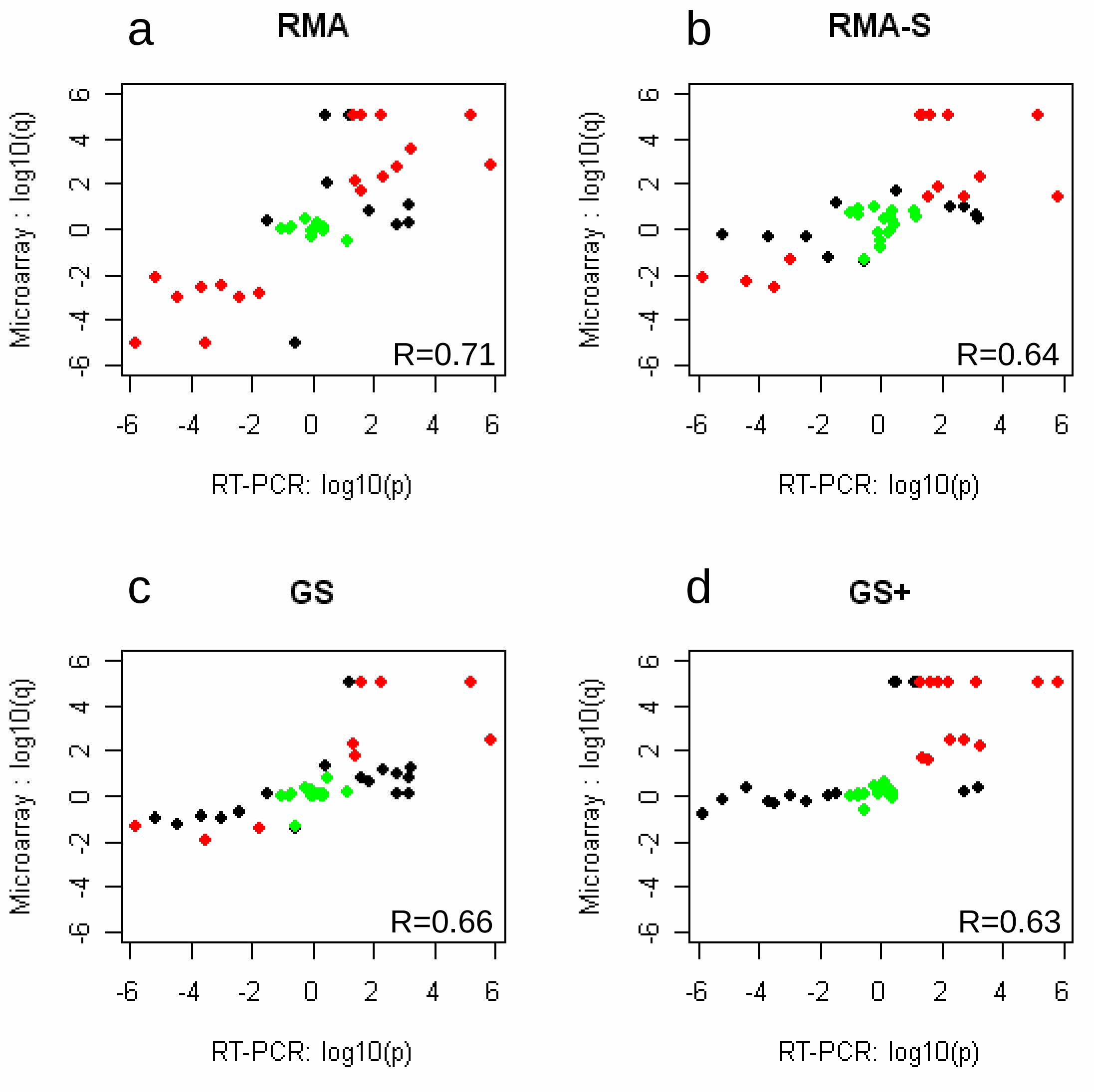
Supplementary Figure 6. qRT-PCR verification of phage gene expression patterns determined from microarray results. (a) Expression profiles for representative genes from each transcription cluster were analyzed by RT-PCR using gene specific primers. The results were normalized to *rnpB* (an internal control gene) to correct for potential differences in input RNA, and are presented relative to maximum levels for each gene. The results are shown on a logarithmic scale to better discern differences in expression patterns at the early time points which are low relative to maximal transcript levels. (b) Microarray results for the same representative genes shown to facilitate direct comparison to the RT-PCR results. Note that, as is commonly found, changes in expression determined by RT-PCR were orders of magnitude greater than by microarray analysis.



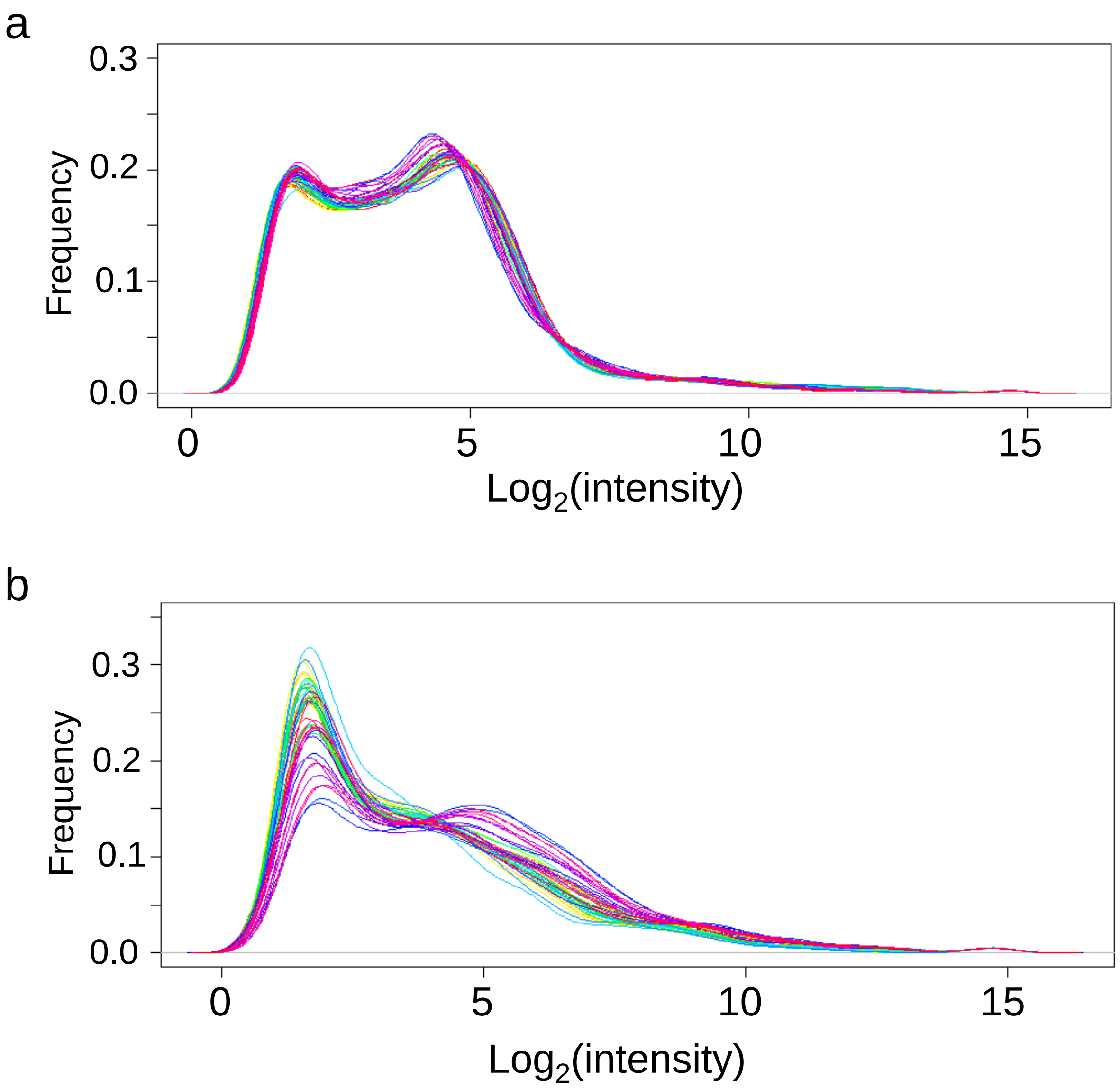
Supplementary Figure 7. qRT-PCR verification of host gene expression patterns determined from microarray analysis. Left panel (a, c, e) shows RT-PCR results and right panel (b, d, f) shows microarray results for representative genes displaying: (a, b) unchanged expression profile; (c, d) down-regulated genes; and (e, f) up-regulated genes. PMM1629 is the only gene whose up-regulated expression was not verified by RT-PCR (compare c and d at T=1 h). Suppl. Table 5 provides a direct comparison of the significance of differentially expressed from qRT-PCR and microarrays after quantile RMA normalization.



Supplementary Figure 8. Comparison of microarray normalization methods to qRT-PCR verified host genes from 3 expression patterns: (a) unchanged – represented by the *rnpB* gene; (b) down-regulated – represented by the *pcb* gene; and (c) up-regulated – represented by the *clpS-like* gene. Quantile RMA normalization and the two versions of Goldenspike performed similarly for up-regulated expression, whereas down-regulated expression patterns were best represented by RMA normalization. See Suppl. Table 5 and Suppl. Fig. 9 for a comparison of the significance values for 11 differentially expressed host genes determined from RT-PCR and the various normalization methods.



Supplementary Figure 9. Comparison of the performance of different microarray normalization methods for detection of significant differences in gene expression as compared to RT-PCR analysis. (a) Quantile RMA normalization at the probe level; (b) RMA normalization based on spiked in hybridization controls; (c) Goldenspike (GS) without summary level normalization; and (d) Goldenspike with summary level normalization. R = Pearson correlation between RT-PCR and microarray analysis. Red and green symbols denote significant and insignificant differences respectively in gene expression called for both RT-PCR and microarray analysis, whereas black symbols denote discrepancies in significance calls between the RT-PCR and microarray analyses. The significance of differential expression for microarray analyses (q-values) was calculated using the Bayes t-test and significance of differential expression for RT-PCR was determined from a standard two-tailed t-test (p-values). Q-values <0.00001 were set to 0.00001 to ensure non-zero values. These findings show that quantile RMA normalization gave the highest correlation and the largest number of correctly identified differentially expressed genes, especially for downregulated genes.



Supplementary Figure 10. Density distribution of signal intensities for probe sets from each microarray after quantile RMA normalization: (a) All probe sets are displayed. The overall distribution for all arrays is similar due to the quantile normalization. (b) MED4 probe sets only are displayed. Note that different arrays displayed various distributions for the MED4 probe sets despite the similar distribution for all probe sets. Therefore quantile normalization can be used for this experiment without erasing differences in MED4 gene expression. Each line represents a different array.