LETTERS

Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution

Debbie Lindell[1]†, Jacob D. Jaffe[3]†, Maureen L. Coleman[1], Matthias E. Futschik[5], Ilka M. Axmann[5], Trent Rector[4], Gregory Kettler[1], Matthew B. Sullivan[1], Robert Steen[4], Wolfgang R. Hess[6], George M. Church[3] & Sallie W. Chisholm[1,2]

Interactions between bacterial hosts and their viruses (phages) lead to reciprocal genome evolution through a dynamic co-evolutionary process[1–5]. Phage-mediated transfer of host genes—often located in genome islands—has had a major impact on microbial evolution[1,4,6]. Furthermore, phage genomes have clearly been shaped by the acquisition of genes from their hosts[2,3,5]. Here we investigate whole-genome expression of a host and phage, the marine cyanobacterium *Prochlorococcus* MED4 and the T7-like cyanophage P-SSP7, during lytic infection, to gain insight into these co-evolutionary processes. Although most of the phage genome was linearly transcribed over the course of infection, four phage-encoded bacterial metabolism genes formed part of the same expression cluster, even though they are physically separated on the genome. These genes—encoding photosystem II D1 (*psbA*), high-light inducible protein (*hli*), transaldolase (*talC*) and ribonucleotide reductase (*nrd*)—are transcribed together with phage DNA replication genes and seem to make up a functional unit involved in energy and deoxynucleotide production for phage replication in resource-poor oceans. Also unique to this system was the upregulation of numerous genes in the host during infection. These may be host stress response genes and/or genes induced by the phage. Many of these host genes are located in genome islands and have homologues in cyanophage genomes. We hypothesize that phage have evolved to use upregulated host genes, leading to their stable incorporation into phage genomes and their subsequent transfer back to hosts in genome islands. Thus activation of host genes during infection may be directing the co-evolution of gene content in both host and phage genomes.

Prochlorococcus is the dominant photosynthetic organism in vast regions of the world's oceans[7], where T7-like podoviruses are also abundant[8]. Therefore this phage–host system is likely to be of great relevance for bacterial and phage global evolution, for modelling their population dynamics, and for understanding the role of phage in the oceanic carbon cycle.

Phages infecting marine cyanobacteria encode a number of host-like genes including photosynthesis and stress-response genes[5,9–11]. Phage photosynthesis genes are expressed during infection while transcripts of homologous genes in the host decline[12,13], and are hypothesized to facilitate production of carbon and energy through cell photosynthesis for optimal phage production[5,10,12–14]. This physiological interdependence between host and phage is likely to have led to the observed prevalence of photosynthesis genes in cyanophage[10,15], providing a reservoir for genetic exchange, and influencing the co-evolutionary process of both host and phage[14,15]. Although the analysis of single genes has provided insight into this dynamic, a systems approach is essential for a broader understanding of this co-evolutionary process. Here we investigate genome-wide transcriptome dynamics of *Prochlorococcus* MED4 and the T7-like podovirus P-SSP7 over the course of infection—the first such detailed view of infection for any lytic host–phage system.

We first characterized the gross features of the lytic cycle (Fig. 1). Phage genomic DNA (gDNA) began to increase, and host gDNA to decrease, 4 h after infection, and phage progeny were first released into the extracellular medium 8 h post infection (Fig. 1a). Phage

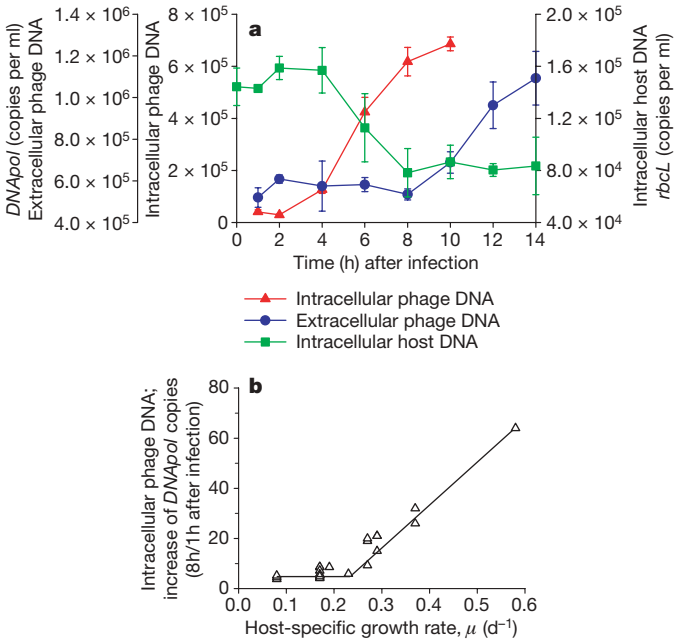

Figure 1 | Infection dynamics of *Prochlorococcus* MED4 by podovirus P-SSP7. a, Timing of phage gDNA replication (intracellular phage DNA) and length of the lytic cycle (extracellular phage DNA) was determined by quantifying the phage DNA polymerase gene (*gene 5*/*DNApol* gene copy number). Host gDNA degradation (intracellular host DNA) was determined by disappearance of the host *rbcL* gene. Average and s.d. of three biological replicates. **b**, Dependence of phage gDNA replication on host growth rate. Phage *DNApol* intracellular copy number was measured 8 h after infection and normalized to that at 1 h after infection as a measure for phage gDNA replication. $n = 24$.

[1]Department of Civil and Environmental Engineering, [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Genetics, and [4]BioPolymers Facility, Department of Genetics Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Institute for Theoretical Biology, Humboldt University, Berlin 10115, Germany. [6]Institute of Biology, University of Freiburg, Freiburg 79104, Germany. †Present addresses: Department of Biology, Technion—Israel Institute of Technology, Haifa 32000, Israel (D.L.); The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02141, USA (J.D.J.).

replication is a function of host growth rate (Fig. 1b), which, together with a dependence on photosynthesis[12], suggests an intimate link between phage fitness and host physiology. We wanted to know: what are the dynamics of the phage and host transcriptome during infection, and where do phage-encoded 'bacterial-like' genes fit into this transcriptional program?

The genome content and architecture of the cyanophage P-SSP7 are similar to that of the *Escherichia coli*-infecting T7 podovirus[11]. As in T7 (ref. 16), the P-SSP7 genome was transcribed linearly from the left to the right of the genome map over the course of infection (with important exceptions, see below) (Fig. 2), and three expression clusters were discerned (Fig. 2a, and Supplementary Fig. 1). The first cluster contains a putative *marR* transcriptional regulator gene, a T7 homologue (g0.7) suggesting a role in redirecting transcription from the host towards the phage. The second cluster contains genes involved in DNA metabolism and replication (Fig. 2a, and Supplementary Fig. 1) as well as RNA polymerase (RNAP), which may be involved in RNA transcription and/or DNA replication[16]. The third cluster consists of genes involved in phage particle formation and DNA maturation. Proteins encoded by this latter cluster were detected in the mature phage particle (Fig. 2b, and Supplementary Table 1), further supporting that many are phage structural genes. Thus the three expression clusters in this cyanophage are analogous to T7-coliphage class I, II and III genes in both gene content[11] and the timing of genome expression (Fig. 2). That these fundamental operational properties are conserved across cyanophage and enteric phage, the hosts of which are drastically different with respect to energy source (autotroph versus heterotroph), habitat (nutrient-poor oceanic waters versus the nutrient-rich human gut), and growth rate (generation time of a day versus less than an hour), is remarkable.

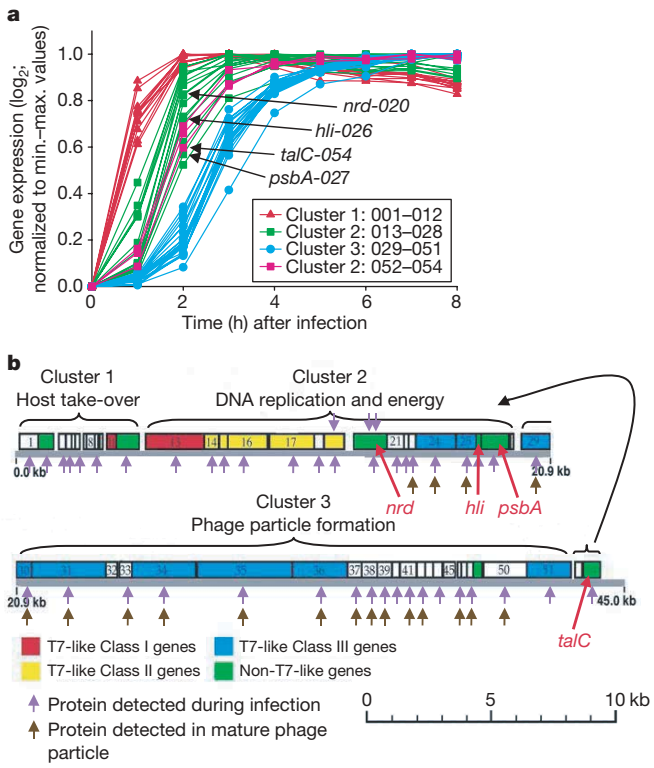

Figure 2 | Temporal expression dynamics of P-SSP7 phage genes during infection of *Prochlorococcus* MED4. a, Transcript levels with time after infection reveal three transcription clusters (see Supplementary Fig. 1). Profiles determined from microarrays, were normalized to minimum–maximum values for each gene. Average of three biological replicates; Supplementary Fig. 6 shows RT–PCR verification of results. **b**, Genome map[11] highlighting the position of *talC* at the end of the genome, even though it is transcribed in cluster 2. Protein detection during infection (purple arrows) and in mature phage particle (brown arrows) showing that 74% of phage genes produced proteins, including three overlapping genes that escaped previous annotation (Supplementary Table 6). Supplementary Table 1 encompasses gene identification and peptide detection.

Despite the similar overall infection strategies of P-SSP7 and T7, transcription cluster 2 in the cyanophage displays novel features in both gene content and regulation and bears signatures of host–phage co-evolution unique to the marine ecosystem. This cluster contains four 'bacterial-like' genes: the ribonucleotide reductase gene *nrd* (ORF 020), the high-light-inducible stress response gene *hli* (ORF 026), the photosystem II gene *psbA* (ORF 027), and the transaldolase gene *talC* (ORF 054). Although *nrd*, *hli* and *psbA* are in the middle of the genome, *talC* is at the end[11] (Fig. 2b). The co-transcription of these four genes, despite their physical separation (Fig. 2, and Supplementary Fig. 2), suggests that they are functionally linked[3].

Clues as to the function of the 'bacterial-like' genes are given by their position in the transcriptional and translational program of the entire host–phage system. First, the proteins encoded by these genes are present during infection but absent from the mature phage particle (Fig. 2b, and Supplementary Table 1), indicating that they function intracellularly. Second, these genes are transcribed together with DNA replication genes, and include ribonucleotide reductase, which converts host ribonucleotides, recycled from degraded RNA (see below), to deoxynucleotides. The photosynthesis genes found in this cluster are thought to be involved in the production of energy[5,10,12–14] and transaldolase may function in the host's pentose phosphate pathway to produce reducing power (NADPH) and/or ribose substrates for nucleotide synthesis[11]. Together, these findings suggest that these genes form a functional unit to produce energy and deoxynucleotide carbon substrates necessary for cyanophage DNA replication in the resource-poor oceans.

The bacterial-like metabolism genes found in P-SSP7 are also commonly found in myoviruses that infect marine cyanobacteria, despite drastic differences in their core genome content[9,11]. In some myoviruses, however, the genes are situated together on the genome[10,11]. Therefore we may be seeing a snapshot of evolution in progress, from spatial separation with cotranscription in P-SSP7, to physically linked genes in other cyanophage genomes.

It is not at all clear how the transcription of this cyanophage genome is regulated, and, in particular, how the last three genes are co-regulated with cluster 2 genes. Although we bioinformatically detected host-like RNAP promoters upstream of each phage expression cluster, and ORF 052, no clear phage-like RNAP promoters were detected[17] (Supplementary Table 2). A transcription initiation site consistent with bacterial-like promoters was experimentally mapped upstream of cluster 2 genes, whereas 5′ ends consistent with RNA processing sites and with weak similarity to T7-like promoters, were found upstream of cluster 1 and cluster 3 genes (Supplementary Fig. 3). However, it remains unclear whether these sequences serve as phage promoters and/or RNA processing sites for transcripts generated by either host or phage RNAP.

Given the reliance of phage replication on host physiology (Fig. 1b), the behaviour of the host transcriptome during infection is of interest. Whereas the transcript levels for approximately 75% of the 1,716 host genes declined during infection (Fig. 3), 41 genes were significantly upregulated. This is distinctly different from other lytic host–phage systems where few, if any, host genes become activated[16,18]. The upregulated genes fall into two groups (Fig. 3, and Supplementary Fig. 4 and Supplementary Table 3). The first was transiently upregulated immediately after infection and consists of high-light-inducible stress response (*hli*), carbon metabolism (*rbcLS*), transcription (*rpoC2*, *rpoD*) and ribosome (*rpl5*, *rpl6*, *rps8*, *rps11*, *rps17*) genes. Transcripts of the second group appeared 2 h after infection and included genes involved in RNA degradation and modification (*rne*, *rnhB*, *dus* and *sun*), protein turnover (*clpS*, and an AAA ATPase family gene), stress responses (*umuD* and *phoH*), and those of unknown function. Two of the latter were

transcribed from bacterial-like promoters (Supplementary Fig. 3), suggesting involvement of host RNAP. Upregulated host gene expression may constitute a direct stress response to phage infection, or may have been facilitated by phage factors[16] injected into the cell or expressed from phage expression cluster 1.

Regardless of the mechanism of upregulation, we hypothesize that phages may have evolved to make use of the products of some upregulated host genes as part of the 'arms-race' between host and phage[19]. Certainly, phages are known to exploit host stress-response proteins during infection in other systems[20–22]. The T4 and T7 phages infecting *E. coli*, for example, have evolved to modify host RNase E (involved in RNA degradation) leading to the degradation of host RNA[22]. It is perhaps not coincidental that *rne* (encoding RNase E) is one of the upregulated genes in *Prochlorococcus* during infection. This may have initially served as a host defence mechanism for degrading phage RNA, but could also be exploited by phage to degrade host RNA for use as substrates for phage deoxynucleotide synthesis.

Perhaps the most compelling evidence that upregulated host genes are part of the co-evolutionary process in this system is that 34% of them (more than would occur by chance $P < 0.001$) are found in hypervariable host genome islands (Supplementary Table 3), which are thought to be mobilized by phages[6]. Furthermore, homologues of a number of these host genes are found in phage genomes, including *hli*, *phoH*, and HNH endonuclease and sigma factor genes, as well as RNase H and heat-shock genes[9,11].

Thus there seems to be a connection between genes upregulated during infection, their position in the host genome, and the presence of homologues in phage. Although there are a number of possible explanations for this connection, the most parsimonious evolutionary scenario is as follows: Host stress response genes are upregulated in response to phage infection. Phages that have evolved to use these gene products gain a fitness advantage. Random incorporation[3] of these genes into their own genomes would enable phages to more tightly regulate their expression, conferring a fitness advantage, and leading to preferential retention. This retention would increase the probability of transfer back to the host in genome islands, by lysogeny or unsuccessful infection, and those genes beneficial to the host would remain in the host genome. Analysis of the *hli* gene family provides an interesting illustration in support of this scenario. *hli* genes are upregulated in the host in response to phage infection (Fig. 3, and Supplementary Table 3), are common in *Prochlorococcus*-infecting phage genomes[5,11], and multiple phage-like copies[5] are found in *Prochlorococcus* genome islands[6] (Supplementary Table 4). Furthermore, their differential expression in *Prochlorococcus* in response to various environmental stressors (Supplementary Table 4) and the presence of a binding site for the nitrogen transcriptional regulator NtcA upstream of the nitrogen-regulated *hli10* gene[23], suggests that copies acquired from phage[5] have undergone specialization of function in the host. It remains to be seen whether host fitness has been enhanced by the acquisition of these *hli* genes from cyanophages.

This system-wide analysis of the infection of a cyanobacterium by a phage has led to new insights and hypotheses regarding co-evolutionary interactions between host and phage. These interactions clearly shape the gene content of both host and phage, and probably play a role in shaping the distribution and abundance of cyanobacterial ecotypes in the oceans.

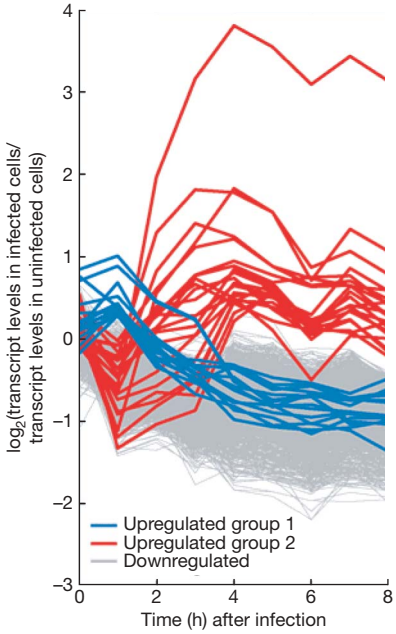

Figure 3 | Transcriptional profiles of *Prochlorococcus* MED4 genes with time after infection by P-SSP7. Transcript levels, determined from microarrays, are presented as $\log_2$-fold change in infected cells relative to uninfected cells over the 8 h latent period of infection. Only genes whose expression levels were significant at a false-discovery rate of $q < 0.05$ are shown. Blue and red indicate significantly upregulated genes in transcription groups 1 and 2, respectively (see Supplementary Table 3). Grey indicates genes significantly downregulated at 8 h after infection. Average of three biological replicates. Supplementary Fig. 7 shows RT–PCR verification of results.

METHODS SUMMARY

Prochlorococcus MED4 was grown at 21 °C under 10–25 μmol photon m^{-1} s^{-1} continuous white light in Pro99 seawater medium with HEPES and sodium bicarbonate. The length of the lytic cycle was determined by quantifying phage DNA in the extracellular medium using a real-time quantitative PCR (qPCR) assay (see Supplementary Fig. 5 for a comparison with standard methods). The timing of phage DNA replication and host DNA degradation were determined intracellularly using qPCR assays for the phage *DNApol* and host *rbcL* genes, respectively. For expression analysis triplicate cultures (10^8 cells ml^{-1}) were infected with the P-SSP7 podovirus (3×10^8 infective phage particles ml^{-1}) and the paired control cultures were amended with filter-sterilized spent medium. Samples were collected by centrifugation, resuspended in storage buffer and snap frozen in liquid nitrogen. RNA was extracted using Ambion's *mir*Vana RNA isolation kit and DNA was removed by DNase I digestion using Ambion's Turbo DNA-*free* kit. Transcriptional analyses were carried out using a custom-made high-density antisense Affymetrix array—MD4-9313. Two micrograms of total RNA were subjected to Affymetrix protocols for *E. coli*. Array analyses were carried out using R and Bioconductor, and array data were normalized and probe set summaries calculated using the robust multi-array average (RMA) procedure[24]. Array results were validated by RT–PCR (Supplementary Figs 6, 7) and the appropriate normalization method was determined by comparing normalized transcription profiles to RT–PCR results (Supplementary Table 5 and Supplementary Figs 8, 9, 10). Promoters were computationally predicted and experimentally assessed using the 5′ RACE technique. For the detection of phage proteins, *Prochlorococcus* cells were harvested 3 and 7 h after infection with phage, and 10^{10} caesium-chloride-purified phage particles were subjected to mass spectrometry proteomic analysis as in ref. 25. See Supplementary Methods for details of all experimental procedures.

Received 5 June; accepted 26 July 2007.

1. Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. Prophage genomics. *Microbiol. Mol. Biol. Rev.* **67,** 238–276 (2003).
2. Filee, J., Forterre, P. & Laurent, J. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res. Microbiol.* **154,** 237–243 (2003).
3. Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. The origins and ongoing evolution of viruses. *Trends Microbiol.* **8,** 504–508 (2000).
4. Hsiao, W. W. L. *et al.* Evidence of a large novel gene pool associated with prokaryotic genome islands. *PLoS Genet.* **1,** e62 (2006).
5. Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101,** 11013–11018 (2004).
6. Coleman, M. L. *et al.* Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311,** 1768–1770 (2006).
7. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63,** 106–127 (1999).
8. Breitbart, M., Miyake, J. H. & Rohwer, F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **236,** 249–256 (2004).

9. Mann, N. H. *et al.* The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus*. *J. Bacteriol.* **187,** 3188–3200 (2005).
10. Millard, A., Clokie, M. R., Shub, D. A. & Mann, N. H. Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc. Natl Acad. Sci. USA* **101,** 11007–11012 (2004).
11. Sullivan, M. B., Coleman, M., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* **3,** e144 (2005).
12. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438,** 86–89 (2005).
13. Clokie, M. R. J., Shan, J., Bailey, S., Jia, Y. & Krisch, H. M. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ. Microbiol.* **8,** 827–835 (2006).
14. Zeidner, G. *et al.* Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ. Microbiol.* **7,** 1505–1513 (2005).
15. Sullivan, M. B. *et al.* Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4,** e234 (2006).
16. Molineux, I. in The Bacteriophages (ed. Calendar, R.) 277–301 (Oxford University Press, New York, 2005).
17. Chen, Z. & Schneider, T. D. Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. *Nucleic Acids Res.* **33,** 6172–6187 (2005).
18. Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67,** 86–156 (2003).
19. Lenski, R. E. & Levin, B. R. Constraints on the coevolution of bacteria and virulent phage: A model, some experiments, and predictions for natural communities. *Am. Nat.* **124,** 585–602 (1985).
20. Tabor, S., Huber, H. E. & Richardson, C. C. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262,** 16212–16223 (1987).
21. Tilly, K., Murialdo, H. & Georgopoulos, C. P. Identification of a second *Escherichia coli groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl Acad. Sci. USA* **78,** 1629–1633 (1981).
22. Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329,** 134–141 (2004).
23. Tolonen, A. C. *et al.* Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol. Systems Biol.* **2,** 53 (2006).
24. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31,** e15 (2003).
25. Jaffe, J. D. *et al.* The complete genome and proteome of *Mycoplasma mobile*. *Genome Res.* **14,** 1447–1461 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Steglich, S. Bhattacharya, H. Keller, L. Thompson, P. Weigele, S. Choe, D. Endy, S. Kosuri, M. Shmoish and J. Aach for discussions, and J. Waldbauer and M. Osburne for comments on the manuscript, and the MIT Center for Environmental Health Sciences. This work was funded by the DOE Genomes to Life System Biology Center Grant (G.M.C. and S.W.C.), the Gordon and Betty Moore Foundation's Marine Microbiology Program (S.W.C.), and the National Science Foundation (S.W.C).

Author Information The microarray data have been deposited in the GEO database under the accession number GSE8382. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.W.C. (chisholm@mit.edu).