LETTERS

Photosystem I gene cassettes are present in marine virus genomes

Itai Sharon[1,2]*, Ariella Alperovitch[1]*, Forest Rohwer[4,5], Matthew Haynes[4], Fabian Glaser[3], Nof Atamna-Ismaeel[1], Ron Y. Pinter[2], Frédéric Partensky[6], Eugene V. Koonin[7], Yuri I. Wolf[7], Nathan Nelson[8] & Oded Béjà[1]

Cyanobacteria of the *Synechococcus* and *Prochlorococcus* genera are important contributors to photosynthetic productivity in the open oceans[1–3]. Recently, core photosystem II (PSII) genes were identified in cyanophages and proposed to function in photosynthesis and in increasing viral fitness by supplementing the host production of these proteins[4–7]. Here we show evidence for the presence of photosystem I (PSI) genes in the genomes of viruses that infect these marine cyanobacteria, using pre-existing metagenomic data from the global ocean sampling expedition[8] as well as from viral biomes[9]. The seven cyanobacterial core PSI genes identified in this study, *psaA, B, C, D, E, K* and a unique *J* and *F* fusion, form a cluster in cyanophage genomes, suggestive of selection for a distinct function in the virus life cycle. The existence of this PSI cluster was confirmed with overlapping and long polymerase chain reaction on environmental DNA from the Northern Line Islands. Potentially, the seven proteins encoded by the viral genes are sufficient to form an intact monomeric PSI complex. Projection of viral predicted peptides on the cyanobacterial PSI crystal structure[10] suggested that the viral–PSI components might provide a unique way of funnelling reducing power from respiratory and other electron transfer chains to the PSI.

Bacteriophages have the ability to manipulate the life histories and evolution of their hosts[11] and evolved many adaptation and defence mechanisms for efficient survival and multiplication. Most of these involve manipulation of the host DNA, as well as the incorporation, into the phage genomes, of bacterial genes that encode proteins with a potential to facilitate bacteriophage reproduction[12]. Recently, it was discovered that marine cyanophages (bacteriophages that infect cyanobacteria) carry photosynthetic genes, and it was suggested that these genes increase phage fitness[4–7]. Cyanobacterial photosynthetic membranes contain two photosystems, of which PSII mediates the transfer of electrons from water, the initial electron donor, to the plastoquinone pool, whereas PSI mediates electron transfer from plastocyanin to ferredoxin, thereby generating reducing power needed for CO_2 fixation in the form of NADPH. Although PSII is known to be sensitive to photodamage, PSI is considered to be a more stable complex. The PSII gene *psbA* coding for the labile D1 protein is readily detected in various cultured and environmental cyanophages infecting *Prochlorococcus* and *Synechococcus*[4,6,13,14]. Furthermore, other photosynthesis genes encoding the PSII D2 protein[4,6], high-light inducible proteins, pigment biosynthesis proteins (Ho1, PebA and PcyA), or the photosynthetic electron transport proteins plastocyanin (PetE) and ferredoxin (PetF) were also identified in several cyanophage genomes[6,15].

To assess the possible presence of other photosynthesis-related genes in viruses, we set up a designated search scheme for publicly available metagenomic data. Initially we searched for the cyanobacterial PSI gene *psaA*. Together with PsaB, the PsaA protein forms the heterodimeric core of PSI that binds the primary electron donor P700, formed by a special chlorophyll pair[10]. Using tBLASTx, different *Synechococcus* and *Prochlorococcus psaA* gene sequences were used as queries against the global ocean sampling (GOS) expedition[8] data set.

We detected 574 *psaA*-containing GOS scaffolds. These were further screened to identify those that were likely to originate from viruses using tBLASTx against refseq_viral, a database that contains all known viral genomes. This procedure reduced the number of suspected scaffolds to five. The PsaA homologues encoded by these sequences showed only 65–75% identity to *Prochlorococcus* or marine *Synechococcus* PsaA proteins. On a maximum-likelihood tree, four of these proteins clustered together on a well-supported branch related to *Prochlorococcus* PsaA, whereas the fifth sequence (JCVI_SCAF_1096628008692) was retrieved near the base of the *Synechococcus* branch (Fig. 1). Because the GOS general scaffold assembly represents reads that come from different GOS sample sites or from different clones and hence are chimaerical by definition, we restricted all further analysis to sequences assembled from single clone reads only. Analysis of the GOS clones containing the modified *psaA* genes confirmed their viral origin (probably cyanophages of the *Myoviridae* family), as indicated by the presence, in the vicinity of *psaA*, of typical viral genes, such as *nrdA* and *B* (that encode the α2 and β2 subunits of viral ribonucleoside diphosphate reductase, respectively) or the T4-like neck gp13 protein gene (Fig. 2). In addition to *psaA*, these clones contained clusters of PSI genes, including *psaB*, *psaC*, and a unique fused version of the *psaF* and *psaJ* genes (*psaJF*). An analysis of the GOS data sets with other PSI peptides as baits showed the presence of several other PSI clusters also containing *psaE*, *psaK* and *psaD* genes (see distribution in the different GOS sites in Supplementary Table 1). Like with PsaA, phylogenies made with these extra PSI protein sequences showed that they were all clustered at a distance from the homologous proteins of *Prochlorococcus* and *Synechococcus*, with the exception of PsaC and PsaD from GOS clone 1061008099984 (hereafter described as clone 9984; a clone used to build the previously mentioned scaffold JCVI_SCAF_1096628008692), which were retrieved closer to corresponding cyanobacterial sequences than to other viral sequences (Supplementary Fig. 1). Examining the *Prochlorococcus* and *Synechococcus* genome arrangements (Fig. 2, middle panel) or gene-pairs frequency modelling showed that the organization observed in most viral clones, *psaJF-C-A-B-K-E-D*, differs from that observed in these cultured cyanobacterial genomes and in most other (probably cyanobacteria-derived) GOS sequences (Fig. 3).

[1]Faculty of Biology, [2]Faculty of Computer Science, [3]Bioinformatics Knowledge Unit, Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel. [4]Department of Biology, [5]Center for Microbial Sciences, San Diego State University, San Diego, California 92182, USA. [6]CNRS and UPMC-Université Paris 6 (UMR 7144), Station Biologique, 29682 Roscoff, France. [7]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. [8]Department of Biochemistry, George S. Wise Faculty of Life Sciences, Daniella Rich Institute for Structural Biology, Tel Aviv University, Tel Aviv 69978, Israel. *These authors contributed equally to this work.

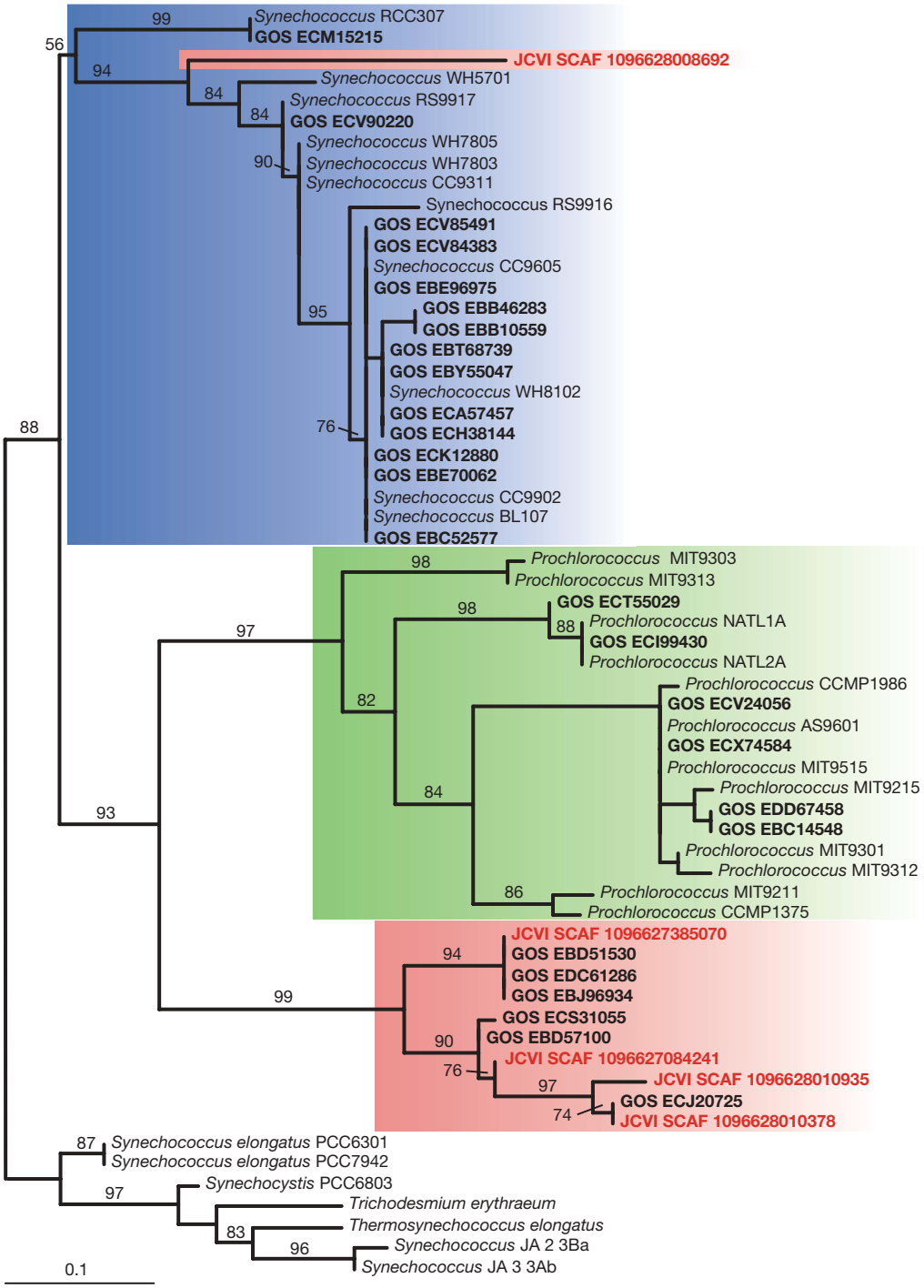

Figure 1 | A maximum-likelihood phylogenetic tree of *psaA*-deduced amino acid sequences obtained from the GOS expedition. PsaA sequences from the 27 fully sequenced and annotated *Synechococcus* (blue background) and *Prochlorococcus* (green background) genomes are shown. Sequences from the GOS expedition are shown in bold, and sequences from the original five scaffolds obtained in this study are indicated in red. For clarity, the tree shows only a subset of the 583 partial PsaA sequences found in the GOS data set. The tree is on the basis of an alignment of 94 shared amino acids. See Supplementary Methods for description of tree construction.

The PSI genes found on clone 9984 (represented by GOS reads 1095964115098 and 1095975140994 in Fig. 2) had a different order (*psaD-C-A)* than on the other clones, consistently with their distinct positions in phylogenetic trees (Supplementary Fig. 1).

To validate the viral origin of these genes and their unique cluster organization, data obtained from the GOS project were cross-referenced with recently released 454 pyrosequencing metagenomic sequences obtained from a variety of marine and non-marine viral and microbial biomes data sets[9]. This was a critical step in increasing the credibility of the results because the two approaches each introduce different biases[16]. The various viral-suspected PSI GOS clones identified were used to recruit reads from these different data sets. Marine virome fragments were readily recruited to all of the viral GOS clones regions, whereas virome or microbiome fragments coming from other environments were scarcely recruited (Table 1), with a much lower identity (Fig. 4a), further supporting a marine viral origin for the PSI clones. The overall coverage measure of viromes and microbiomes to all different GOS clones containing PSI genes (Fig. 4b) clearly points to two distinguished populations, one from bacteria (cyanobacteria) and one from viruses (phages). Except for clone 9984, all our identified viral clones are falling in the viral population. Furthermore, marine virome fragments were also recruited to regions between the photosynthesis genes, linking neighbour genes in the observed viral cassette (Fig. 4a and Supplementary Table 2)—an observation that supports the gene cluster organization observed on the GOS clones.

To validate the juxtaposition of the genes in the identified viral–PSI gene clusters, DNA from the Northern Line Islands marine virome[17] was used to perform 'continuous' overlapping and long PCR with primers assigned to the different genes (Supplementary Table 5). The results of the 'continuous' overlapping PCR (Fig. 2, bottom

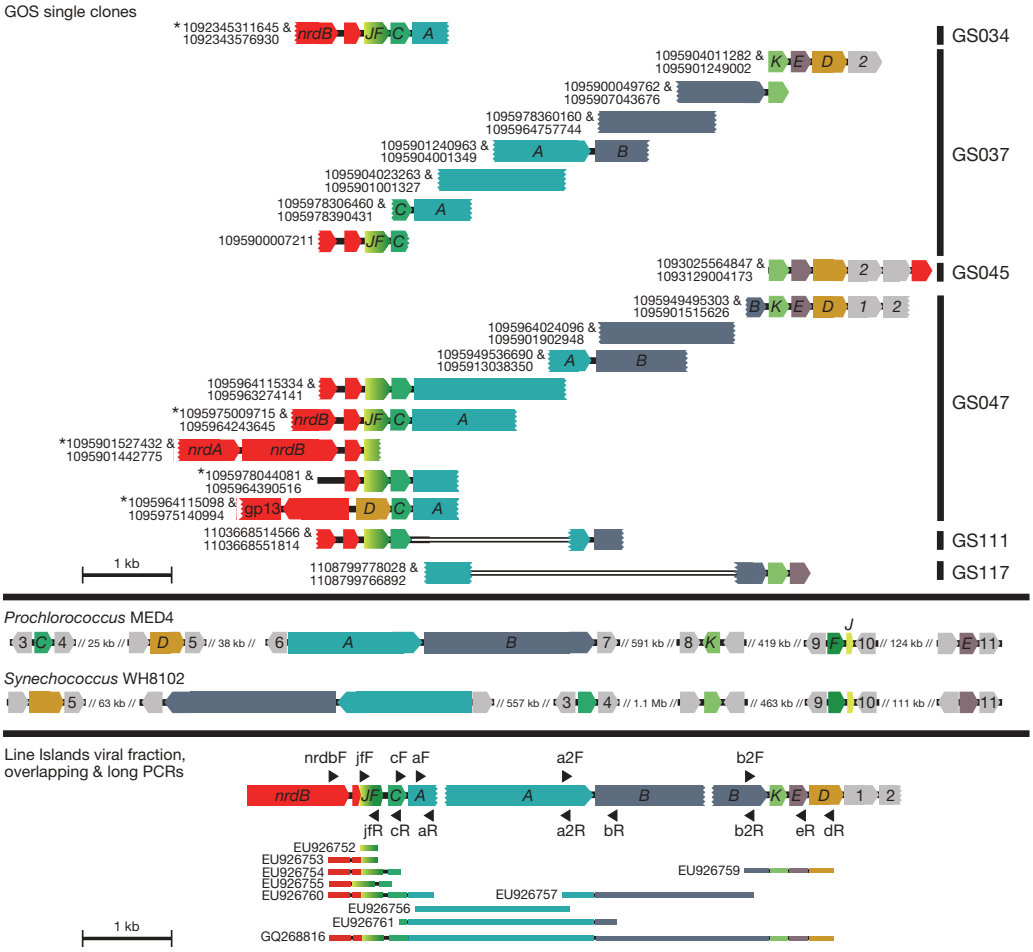

Figure 2 | Schematic physical maps of selected viral-suspected GOS clones (top), *Prochlorococcus* and *Synechococcus* genomes (middle) and environmental PCR products containing PSI genes (bottom). Red arrows represent ORFs with predicted viral origin, and grey arrows represent unknown ORFs. Capital letters represent the corresponding PSI core genes. Gaps shown in Indian Ocean GOS clones (stations GS111 and GS117) are the result of regions that were not covered by the end-reads owing to the size of these clones (5 kb). Primer positions on GOS clones are indicated by triangles, and thick coloured lines denote PCR products.

panel) and the amplification of a ~6.2-kilobase (kb) long PCR amplicon (Supplementary Fig. 2 and Fig. 2, bottom panel) spanning the entire PSI cassette and including the viral *nrdB* gene, show that the different genes in the cluster *nrdB-hyp-psaJF-C-A-B-K-E-D* (and also a new arrangement *nrdB-psaJF-C*; GenBank accession EU926755) are physically linked and exist as one photosynthetic cluster.

Although the data presented here are derived from environmental genomic data sets (non-continuous data), and therefore the lack of genes is not a proof of absence, it is notable that the PSI genes *psaI*, *psaL* and *psaM* were not found in the viral *psa* gene cassettes. The *psaM* gene is naturally absent from plants[18] and its inactivation in cyanobacteria shows that it is mainly required for the formation of stable PSI trimers[19]. Similarly, targeted inactivation of cyanobacterial *psaL* produces functional PSI complexes unable to form trimers, whereas PsaI is mostly required for stabilizing PsaL[20]. Therefore, these three proteins are mainly involved in the trimer formation of cyanobacterial PSI, and their potential absence from the viral clone might indicate the formation of a monomeric PSI complex as in

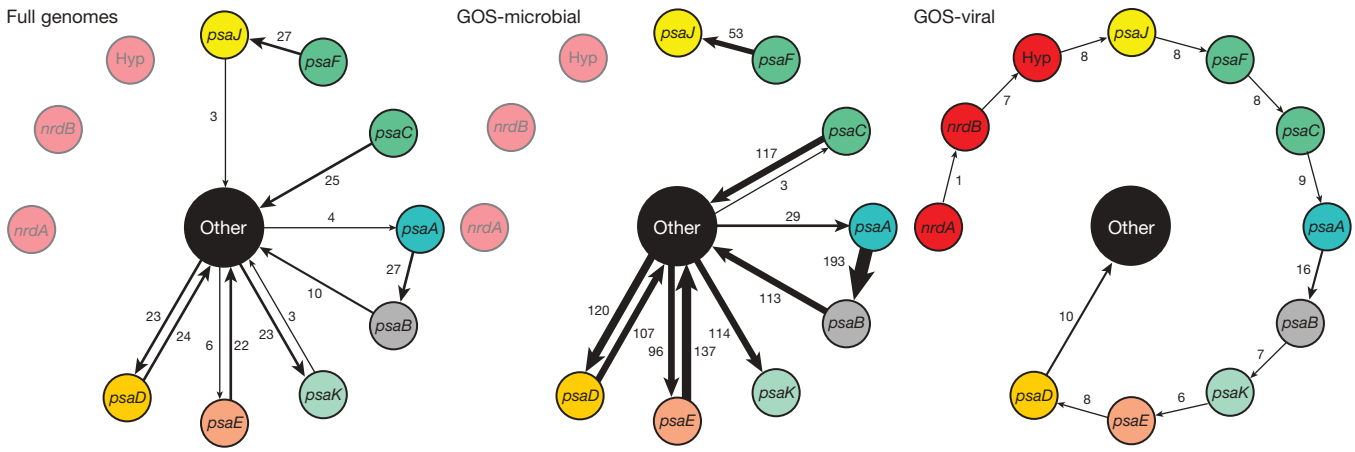

Figure 3 | Distribution of neighbouring genes involving at least one PSI gene. Each arrow connects neighbouring genes, and its thickness represents the number of pairs found in *Synechococcus* and *Prochlorococcus* genomes (left gene-circle), microbial sequences from the GOS metagenome (middle gene-circle), and viral sequences from the GOS metagenome (right gene-circle). Note the uninterrupted clustering of PSI genes in phage genomes that contrasts the scattered arrangement of these genes in cyanobacterial genomes (in both cultures and GOS). Gene connections observed only once are not shown.

Table 1 | Number of different biome reads recruited to GOS-suspected viral–PSI clones

Type	Microbial metagenomes			Viral metagenomes		
	85%	90%	95%	85%	90%	95%
Coral	0	0	0	0	0	0
Fish	0	0	0	0	0	0
Freshwater	2	0	0	1	1	1
Hypersaline	1	0	0	0	0	0
Marine	1	0	0	207	144	91
Microbialites	0	0	0	0	0	0
Terrestrial	0	0	0	0	0	0

plants[21] and not a trimeric complex as in cyanobacteria[22]. All genetic information required to form this putative minimal, monomeric PSI is clustered onto a very small cyanophage genome fragment (~5.9 kb). To our knowledge, gene clusters encoding all the components of a photosystem from an oxygenic phototroph have not been previously reported, and neither have there been reports on cyanobacterial PSI genes outside a cyanobacterial chromosome.

The potential structural consequences of assembling the phage proteins into the PSI complex were modelled in relation to the 2.5 Å structure of PSI from the cyanobacterium *Thermosynechococcus elongatus*[10]. We modelled the PsaJF fusion protein (in which the carboxy terminus of PsaJ is fused to the amino terminus of PsaF) at the position of subunits J and F of PSI. Figure 5 shows that the viral PsaJF fusion protein fits perfectly at the position of subunits J and F in the PSI structure. The only prominent change was the absence of the N terminus of subunit F, which is responsible for the specific binding of the natural electron donor (plastocyanin) of PSI[23]. In chloroplasts of green algae and plants, this part of subunit F is elongated, resulting in higher affinity of plastocyanin to the chloroplast PSI[18,23,24]. Although both plastocyanin and cytochrome c_6 are capable of donating electrons to PSI[24] in *Chlamydomonas reinhardtii*, this site in higher plants is specific for plastocyanin[25]. However, the electron donation to PSI is not at all promiscuous, and several soluble cytochromes, including the respiratory cytochrome *c*, fail to donate electrons to PSI[26]. We propose that the replacement of PsaJ and PsaF with the viral PsaJF fusion protein enables electron donation through extra electron carriers, including cytochromes that usually function as electron donors to cytochrome oxidase.

The mechanistic consequence of a less selective electron donation to PSI might be the possibility of sharing reducing power generated by the respiratory chain with the photosynthetic electron transport chain. A similar phenomenon, called chloro-respiration, detected in both cyanobacteria and chloroplasts, was attributed to the plastid terminal plastoquinone oxidase (PTOX)[27]. The electron mediator in

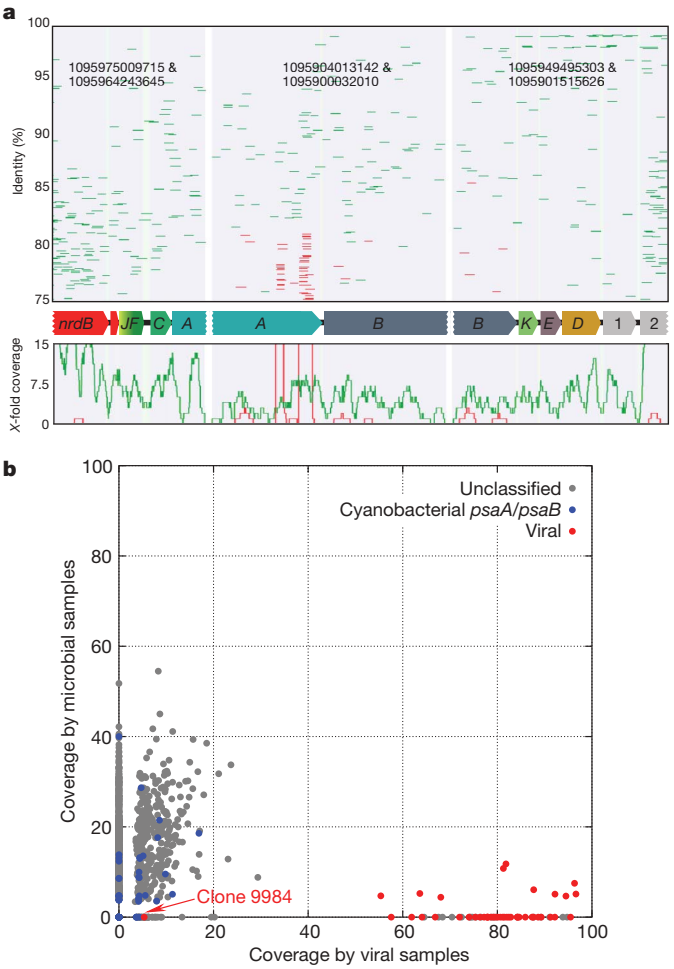

Figure 4 | Recruitment of GOS clones carrying PSI genes with Northern Line Islands biomes. a, Recruitment of three GOS physical clones carrying suspected viral-PSI genes by Northern Line Islands virome (green) and microbiome (red) reads. The top panel shows recruitment at 75–100% identity, and the bottom panel shows the fold-coverage by these reads. Accession numbers of the GOS reads used are presented above each clone (JCVI_READ_#). **b**, Recruitment coverage of GOS single clones carrying PSI genes with Northern Line Islands biomes (viromes (*x*-axis) and microbiomes (*y*-axis)). Coverage is defined as the percentage of GOS clone length covered by at least one recruited read.

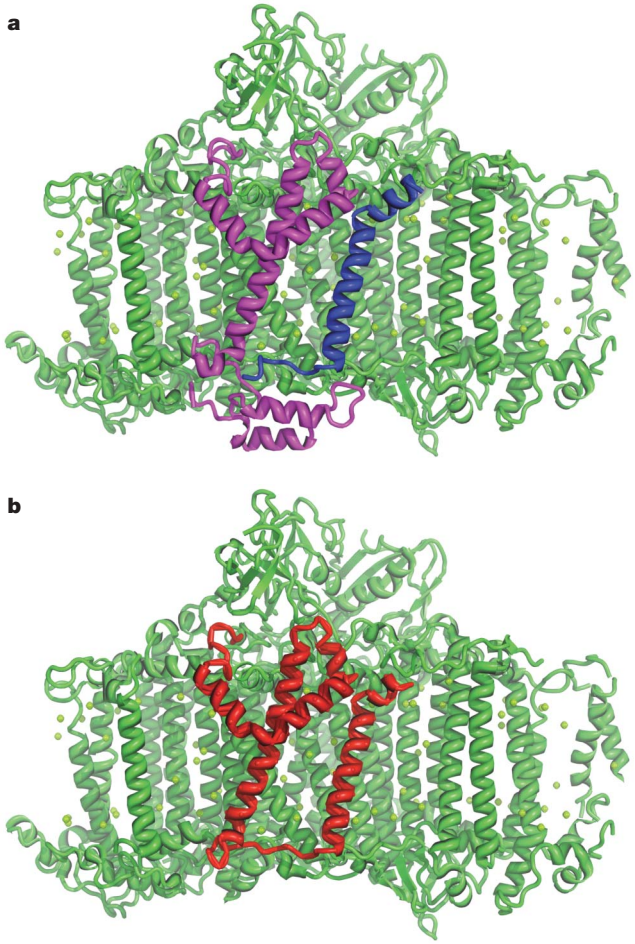

Figure 5 | Structural consequences of assembling the viral fusion protein PsaJF into PSI. a, The structure of *T. elongatus* PSI (subunits) was illustrated by PyMOL (http://pymol.sourceforge.net/) using a PSI monomer (adopted from Protein Data Bank (PDB) accession 1jb0). PsaF is in magenta, PsaJ is in blue, and all of the other subunits are in green. **b**, A model for the structure of the viral PsaJF fusion protein (red) substituting the original PsaF and PsaJ subunits.

this process is plastoquinone, which shuttles between the respiratory-like chain and the chloroplast b_6f complex[27]. After phage infection and the incorporation of the phage gene products into PSI, the function of electron mediation could be carried out by a soluble cytochrome. Moreover, the phage might boost the amount of PSI to lead the infected cyanobacterial cells towards a cyclic photosynthesis for the generation of ATP in expense for the production of reducing power for CO_2 fixation. The PSI levels are notably low in both oceanic *Synechococcus*[28] and in *Prochlorococcus*[29], possibly as a result of adaptation to low iron levels, and it was recently proposed that a compensatory mechanism might exist, involving alternative electron flow to O_2 (ref. 28).

The phage PSI gene fusion *psaJF* described here is, to our knowledge, the first example of a phage gene innovation that involves structural membrane proteins. Modification towards a new function of existing cyanobacterial proteins by their phages was recently demonstrated for the divergent phage PebA homologue[30] (renamed PebS (phycoerythrobilin synthase)). The phage PebS single-handedly catalyses a reaction for which uninfected host cells require two consecutive enzymes, PebA and PebB. Considering these findings and our calculations that suggest a high likelihood of gene cluster formation in phage genomes (see Supplementary Information), the oceanic virome could be an almost unlimited source of naturally bioengineered gene cassettes.

METHODS SUMMARY

Collecting GOS-PSI clones. The following steps were taken to identify viral and non-viral PSI clones in GOS: (1) tBLASTx searches, with e-value threshold of 10^{-20}, of *psaA*, *psaB*, *psaC*, *psaD*, *psaE*, *psaF*, *psaJ* and *psaK* probes against the data set of GOS scaffolds. This step yielded 1,167 scaffolds. (2) Identify all reads composing the scaffolds found in the previous stage and their division into clones. Overall, 3,758 reads from 2,147 clones were found (536 single-read and 1,611 pair-end clones). (3) 'In-clone assembly', the reads of each pair-end clone were aligned (bl2seq) and assembled; 50 Ns were added between non-overlapping reads. (4) Annotation, an iterative procedure was used for gene discovery and annotation: at each iteration all clones were BLASTxed against nr (e-value threshold = 10), first hit for each clone was saved and the clone's segment in the alignment was replaced with Ns. For each clone, the process halted when no new hits were found. (5) All clones with no PSI hit were removed. Overall we were left with 1,585 GOS clones carrying at least one PSI gene.

Collecting preliminary set of viral-PSI sequences in GOS. To find candidate viral-PSI sequences in GOS we have used the following two-step method: first, identify all GOS sequences (scaffolds or clones) containing *psaA* genes (see earlier); and second, identify viral genes on *psaA*-containing sequences. In the second step, all *psaA*-containing sequences were blasted (tBLASTx) against the refseq-viral database (again, with an e-value threshold of 10^{-20}). The initial scan revealed five scaffolds containing both *psaA* and viral genes. These scaffolds were annotated (BLASTx against the nr database) and found to contain both viral genes such as *nrdA* and *nrdB*, as well as PSI genes such as *psaA*, *psaC*, *psaD* and a fusion of *psaJ* and *psaF*.

For details on recruitments against 454 databases, gene organization analysis, abundance measures, estimation of the number of recombination events, and PCR conditions, see Supplementary Methods.

Received 14 June; accepted 9 July 2009.
Published online 26 August 2009.

1. Li, W. K. W., Zohary, T., Yacobi, Y. Z. & Wood, A. M. Ultraphytoplankton in the eastern Mediterranean Sea: towards deriving phytoplankton biomass from flow cytometric measurements of abundance, fluorescence and light scatter. *Mar. Ecol. Prog. Ser.* **102,** 79–87 (1993).
2. Liu, H., Nolla, H. A. & Campbell, L. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat. Microb. Ecol.* **12,** 39–47 (1997).
3. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63,** 106–127 (1999).
4. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Bacterial photosynthesis genes in a virus. *Nature* **424,** 741 (2003).
5. Millard, A., Clokie, M. R. J., 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).
6. Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101,** 11013–11018 (2004).
7. 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).
8. Rusch, D. B. *et al.* The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through the eastern tropical Pacific. *PLoS Biol.* **5,** e77 (2007).
9. Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452,** 629–632 (2008).
10. Jordan, P. *et al.* Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411,** 909–917 (2001).
11. Rohwer, F. & Thurber, R. V. Viruses manipulate the marine environment. *Nature* **459,** 207–212 (2009).
12. Brown, N. F., Wickham, M. E., Coombes, B. K. & Finlay, B. B. Crossing the line: selection and evolution of virulence traits. *PLoS Pathog.* **2,** e42 (2006).
13. 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).
14. Sharon, I. *et al.* Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J.* **1,** 492–501 (2007).
15. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* **3,** e144 (2005).
16. Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10,** R32 (2009).
17. Dinsdale, E. A. *et al.* Microbial ecology of four coral atolls in the northern line islands. *PLoS One* **3,** e1584 (2008).
18. Ben-Shem, A., Frolow, F. & Nelson, N. The crystal structure of plant photosystem I. *Nature* **426,** 630–635 (2003).
19. Naithani, S., Hou, J. M. & Chitnis, P. R. Targeted inactivation of the *psaK1*, *psaK2* and *psaM* genes encoding subunits of Photosystem I in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **63,** 225–236 (2000).
20. Xu, Q. *et al.* Mutational analysis of photosystem I polypeptides in the cyanobacterium *Synechocystis* sp. PCC 6803. Targeted inactivation of psaI reveals the function of psaI in the structural organization of psaL. *J. Biol. Chem.* **270,** 16243–16250 (1995).
21. Kouril, R., van Oosterwijk, N., Yakushevska, A. E. & Boekema, E. J. Photosystem I: a search for green plant trimers. *Photochem. Photobiol. Sci.* **4,** 1091–1094 (2005).
22. Boekema, E. J. *et al.* Evidence for a trimeric organization of the photosystem I complex from the thermophilic cyanobacterium *Synechococcus* sp. *FEBS Lett.* **217,** 283–286 (1987).
23. Hippler, M., Drepper, F., Farah, J. & Rochaix, J. D. Fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii* requires PsaF. *Biochemistry* **36,** 6343–6349 (1997).
24. Nelson, N. & Yocum, C. Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* **57,** 521–565 (2006).
25. Merchant, S. & Sawaya, M. R. The light reactions: a guide to recent acquisitions for the picture gallery. *Plant Cell* **17,** 648–663 (2005).
26. Kerfeld, C. A. & Krogmann, D. W. Photosynthetic cytochromes c in cyanobacteria, algae and plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49,** 397–425 (1998).
27. Rumeau, D., Peltier, G. & Cournac, L. Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* **30,** 1041–1051 (2007).
28. Bailey, S. *et al.* Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochim. Biophys. Acta* **1777,** 269–276 (2008).
29. Partensky, F., La Roche, J., Wyman, K. & Falkowski, P. G. The divinyl-chlorophyll *a*/*b*-protein complexes of two strains of the oxyphototrophic marine prokaryote *Prochlorococcus*—characterization and response to changes in growth irradiance. *Photosynth. Res.* **51,** 209–222 (1997).
30. Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W. & Frankenberg-Dinkel, N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr. Biol.* **18,** 442–448 (2008).

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

Acknowledgements We would like to thank M. Rosenberg for technical support, D. Rusch, J. Zehr and S. Bench for sharing genomic data, D. Lindell and R. Sorek for encouragement and discussions, and U. Pick for the comments on cyclic photosynthesis. This work was supported in part by grants 1203/06 (O.B.) and 356/06 (N.N.) from the Israel Science Foundation, by the Henry Taub Award for Academic Excellence, and by the Technion V.P.R. Fund-Henri Gutwirth Promotion of Research Fund (O.B.).

Author Contributions A.A. devised the initial idea for the project. I.S. and O.B. conceived the experiments. I.S. wrote the code and analysed the raw data, and together with F.G., R.Y.P., E.V.K., Y.I.W., N.N. and O.B. performed the bioinformatics. F.R. collected DNA and phage concentrates from the Northern Line Islands. A.A., N.A.-I. and M.H. conducted the molecular biology experiments, I.S., F.P., E.V.K., N.N. and O.B. co-wrote the paper.

Author Information The PSI sequences reported here have been deposited with GenBank under accession numbers EU926752–EU926761 (overlapping PCRs) and GQ268816 (long PCR). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.B. (beja@tx.technion.ac.il).