



**HAL**  
open science

## Virus-host coexistence in phytoplankton through the genomic lens

Sheree Yau, Marc Krasovec, L. Felipe Felipe Benites, Stéphane Rombauts, Mathieu Groussin, Emmelien Vancaester, Jean-Marc Aury, Evelyne Derelle, Yves Desdevises, Marie-Line Escande, et al.

► **To cite this version:**

Sheree Yau, Marc Krasovec, L. Felipe Felipe Benites, Stéphane Rombauts, Mathieu Groussin, et al.. Virus-host coexistence in phytoplankton through the genomic lens. *Science Advances*, 2020, 6 (14), pp.eaay2587. 10.1126/sciadv.aay2587. hal-02557501

**HAL Id: hal-02557501**

**<https://hal.sorbonne-universite.fr/hal-02557501>**

Submitted on 28 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## OCEANOGRAPHY

## Virus-host coexistence in phytoplankton through the genomic lens

Sheree Yau<sup>1,2\*</sup>, Marc Krasovec<sup>1</sup>, L. Felipe Benites<sup>1</sup>, Stephane Rombauts<sup>3,4</sup>, Mathieu Groussin<sup>5</sup>, Emmelien Vancaester<sup>3,4</sup>, Jean-Marc Aury<sup>6</sup>, Evelyne Derelle<sup>1,7</sup>, Yves Desdevises<sup>1</sup>, Marie-Line Escande<sup>1</sup>, Nigel Grimsley<sup>1</sup>, Julie Guy<sup>6</sup>, Hervé Moreau<sup>1</sup>, Sophie Sanchez-Brosseau<sup>1</sup>, Yves van de Peer<sup>3,4,8</sup>, Klaas Vandepoele<sup>3,4</sup>, Sebastien Gourbiere<sup>9</sup>, Gwenael Piganeau<sup>1\*</sup>

Virus-microbe interactions in the ocean are commonly described by “boom and bust” dynamics, whereby a numerically dominant microorganism is lysed and replaced by a virus-resistant one. Here, we isolated a microalga strain and its infective dsDNA virus whose dynamics are characterized instead by parallel growth of both the microalga and the virus. Experimental evolution of clonal lines revealed that this viral production originates from the lysis of a minority of virus-susceptible cells, which are regenerated from resistant cells. Whole-genome sequencing demonstrated that this resistant-susceptible switch involved a large deletion on one chromosome. Mathematical modeling explained how the switch maintains stable microalga-virus population dynamics consistent with their observed growth pattern. Comparative genomics confirmed an ancient origin of this “accordion” chromosome despite a lack of sequence conservation. Together, our results show how dynamic genomic rearrangements may account for a previously overlooked coexistence mechanism in microalgae-virus interactions.

## INTRODUCTION

Phytoplankton-virus interactions are major determinants of geochemical cycles in the oceans. Viruses are responsible for the redirection of carbon and nutrients away from macroorganisms back toward microorganisms via the lysis of their hosts in a process coined the “viral shunt.” Resolving the mechanisms underlying these interactions and their impact on phytoplankton dynamics represents one of the major challenges for future projections of ocean ecosystems under climate change (1, 2). Nucleocytoplasmic large DNA viruses (NCLDV) are estimated to be present at  $10^4 \cdot \text{ml}^{-1}$  in the photosynthetic zone (3), a large proportion of which are prasinoviruses, a monophyletic group known to infect Mamiellophyceae, a class of cosmopolitan marine picophytoplankton (4), which can be dominant primary producers in coastal areas (5).

Unlike bacteriophages that may integrate into their hosts as prophages and develop complex lysogeny-lysis switch strategies (6), by which the prophage is excised from the host genome and whose transcription induces cell lysis, all known NCLDV-infecting unicellular organisms replicate by cell lysis (7). Nonetheless, acquisition of viral resistance after initial viral lysis has been reported in several phylogenetically diverse microalga-NCLDV populations, which displayed

continuous production of infective viruses (8). Here, we define as “coexistence” of microalgae and viruses the situation where infective viruses are produced by healthy microalgal cultures grown from one single cell and are thus isogenic except for the presence of spontaneous mutations. The process allowing a sustained virus production in these isogenic microalgal cultures has remained unsolved.

Virus and microalga coexistence has been previously reported in the freshwater green alga *Chlorella variabilis* N64A infected by *Paramecium bursaria* chlorovirus (PBCV-1), and this phenotype was coined a “carrier” state (9). Coexistence was also observed in the haptophyte *Emiliania huxleyi* upon infection by EhV-99B1 over 12 days (8) and subsequently in the marine Mamiellophyceae picoalga *Ostreococcus tauri* upon infection by *Ostreococcus tauri* virus 5 (OtV5) over 18 days (10). More recently, a study of the population dynamics of *C. variabilis* and PBCV-1 reported that stable coexistence evolved 30 days after infection, which was maintained until the end of the experiment 45 days later (11).

*Ostreococcus* and their prasinoviruses represent ecologically relevant cultivable microalga-NCLDV model systems for focused study of host-virus interaction mechanisms and dynamics (12). *Ostreococcus mediterraneus* is the most recently described species, and gene phylogeny of the 18S ribosomal RNA (rRNA) indicated that it is the earliest branching member of the *Ostreococcus* genus (13). The analysis of the complete genome sequence of *O. mediterraneus* (strain RCC2590) unexpectedly revealed the sequence of a previously unidentified prasinovirus, hereafter named OmV2 for *Ostreococcus mediterraneus* virus 2. Because the DNA extraction was performed shortly after the environmental sampling of the strain and because subsequent inspections of the culture revealed the presence of infective OmV2 particles for over a decade, we concluded that this microalga-NCLDV system was isolated as a coexisting system, rather than upon the selection of resistant (R) cells after an initial lytic event. We investigate the evolutionary history of this system by whole-genome analysis and the mechanisms underlying viral production by experimental evolution. Last, mathematical modeling of host-virus coexistence as a resistant-susceptible phase switch in the host allows us to explain the observed microalga-virus dynamics.

<sup>1</sup>Sorbonne Université, CNRS, UMR 7232 Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650 Banyuls-sur-Mer, France. <sup>2</sup>Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), CSIC, Barcelona, Spain. <sup>3</sup>Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark-Zwijnaarde 71, 9052 Ghent, Belgium. <sup>4</sup>VIB Center for Plant Systems Biology, Technologiepark-Zwijnaarde 71, 9052 Ghent, Belgium. <sup>5</sup>Department of Biological Engineering, Massachusetts Institute of Technology, 500 Technology Square NE47-378, Cambridge, MA 02139, USA. <sup>6</sup>Genoscope, Institut de biologie François Jacob, Commissariat à l’Énergie Atomique (CEA), Université Paris-Saclay, Evry, France. <sup>7</sup>Univ. Brest, CNRS, IRD, Ifremer, LEMAR, F-29280 Plouzané, France. <sup>8</sup>Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa. <sup>9</sup>Laboratoire Génome et Développement des Plantes, Université de Perpignan Via Domitia, UMR 5096, 52 Avenue Paul Alduy, 66860 Perpignan, France.

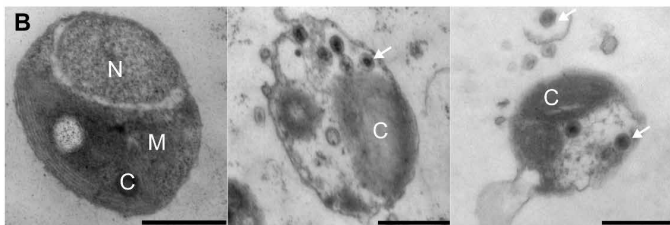
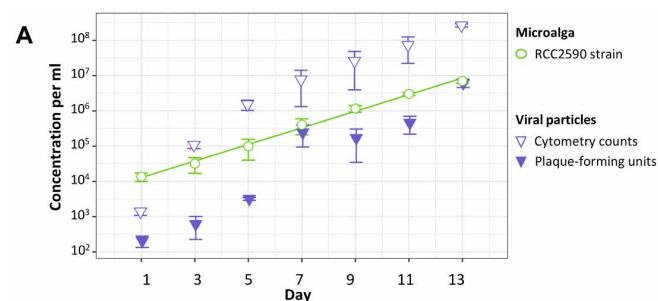
\*Corresponding author. Email: gwenael.piganeau@obs-banyuls.fr (G.P.); syau@icm.csic.es (S.Y.)

## RESULTS

**Production of virions is the consequence of lysis of a subpopulation of microalgae**

Population dynamics of microalgal cells and OmV2 in culture was followed over 2 weeks. The logarithm of microalgal concentration,  $C$ , increases linearly with time,  $t$ , in RCC2590 cultures (Fig. 1A;  $\ln(C) = 0.557 \times t + 7.559$ ,  $R^2 = 0.94$ ). Virus concentration was estimated by flow cytometry and by plaque assay, which measure the total number of DNA-containing particles in the culture medium and the number of infectious viruses, respectively. Despite variation in the estimation of the number of viruses between replicates and between the two methods, the number of viruses increases with the number of microalgae, providing evidence of continuous production of viruses in culture.

We performed transmission electron microscopy (TEM) on these cultures to examine how the virus was being produced. The majority of cells had typical morphology (Fig. 1B) described for *Ostreococcus*, while 0.9% of cells (5 of 573 individuals) contained icosahedral particles of approximately 110 nm in diameter within the cytoplasm (Fig. 1, C and D). Viral particles could also be visualized in the medium (Fig. 1D), often as aggregates around cellular debris. The capsid size and icosahedral morphology, as well as virion development within the cytoplasm, are consistent with the lytic replication cycle observed in other isolated prasinoviruses (14). This led us to conclude that viruses are produced from the lysis of a subset of virus-susceptible cells in the majority of virus-resistant RCC2590 culture. We did not observe nonlytic virus production from budding or extrusion through the cell membrane. Since the cultures are isogenic, having been isolated from one single cell, this raised the question of how susceptible (S) cells originate from a majority of R cells.



**Fig. 1. Population dynamics of the algae and virus.** (A) Population dynamics of microalgae (circles) and prasinovirus (triangles) concentrations over 14 days in the *O. mediterraneus* RCC2590 culture. The line is fitted via ordinary least square regression. (B) Transmission electron micrographs of *O. mediterraneus* RCC2590 cells from an actively growing culture. The majority of cells have typical morphology (left), while 0.9% of cells contained visible viral particles (middle and right) and free viral particles were observed in the medium aggregated around infected cells (right). Scale bars, 500 nm; C, chloroplast; M, mitochondrion; N, nucleus. White arrows indicate examples of virus particles.

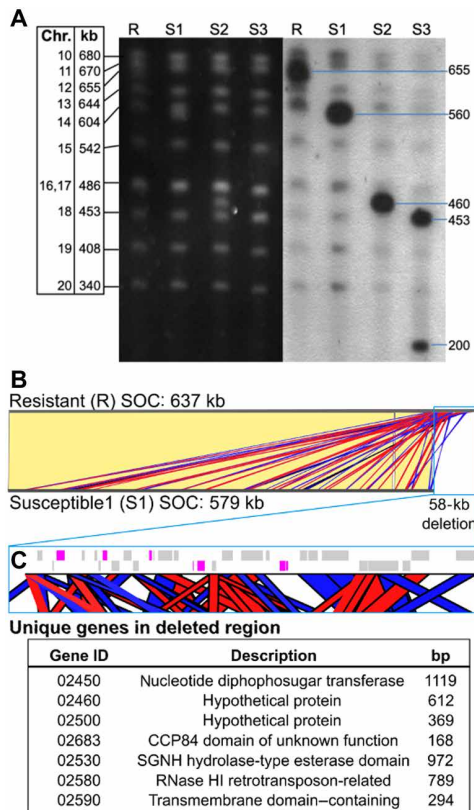
**Susceptible (S) cells derive from resistant (R) cells and vice versa**

To validate the existence of a switch by which S phenotypes could be generated by R cells within isogenic cultures, we performed experimental evolution. First, we established a clonal culture line that was free of viruses by diluting the original OmV2-producing RCC2590 culture to a single cell. This parent line (hereafter referred to as ML for “mother line”) was confirmed to be virus-free by polymerase chain reaction (PCR) screening for OmV2. The ML was not lysed after readdition of purified OmV2, showing that the ML remained R. Furthermore, this demonstrated that uninfected individual cells were in the original RCC2590 culture that could be physically separated from OmV2. Second, the virus-free ML was used to generate hundreds of clonal cultures, again from a single cell, in four independent tests. As the phenotype at the level of a single cell cannot yet be determined, we screened these newly cloned cultures for susceptibility to OmV2 after 7 days of growth; thus, the phenotype of the majority of the cells in culture would reflect the phenotype of the starting cell. In the first test, the ML was repeatedly cloned via single-cell bottlenecks every 2 weeks for 1 year in multiple lines as part of a mutation accumulation experiment in which 3 of 24 independent lines were visibly lysed upon addition of OmV2 (15). To determine whether S phenotypes could evolve over short time periods, three independent experiments were performed, where the ML was subcloned once by end-point dilution to generate a total of 204 clonal cultures and tested for susceptibility to OmV2. Visual inspection revealed no lysis in 159 clones, while 45 clones lysed with the addition of OmV2. This demonstrated that cultures containing a majority of S cells evolved from a single R cell at a 12.5 to 22% frequency.

To test whether cultures containing a majority of S cells similarly could regenerate an R phenotype, the three independent S lines from the first mutation accumulation experimental evolution study were exposed to OmV2. Infected S cultures showed a sharp decrease in cell concentration in the first 15 days, consistent with a lytic event, dropping as low as 167 cells ml<sup>-1</sup>, after which cultures recovered and cell concentrations reached that of uninfected controls (fig. S1). These cultures subsequently showed no visible lysis when rechallenge with OmV2, demonstrating that S cultures can reliably switch to an R state. Exposure to OmV2 was performed independently a second time on 24 S lines and confirmed systematic resistance acquisition.

**What is the genomic signature of the R-to-S switch?**

The three S lines—S1, S2, and S3—obtained from the mutation accumulation experimental evolution study were maintained in batch culture and intermittently tested for susceptibility to OmV2. Viral lysis was observed at each test, confirming that these cultures retained a majority of S cells. Pulsed-field gel electrophoresis (PFGE) karyotype analysis of the ML and three S lines revealed a visible size reduction of 95 to 435 kb in one chromosome, while no difference in karyotype could be observed in the OmV2-R ML, which remained unchanged from the OmV2-producing original RCC2590 (Fig. 2). Hybridization using specific probes designed from the complete genome sequence identified the chromosome with a reduced size to be the “small outlier chromosome” (SOC) of *O. mediterraneus*. The term SOC was coined to describe a chromosome found in all Mamiellales genomes sequenced so far—including *Ostreococcus* (16), *Micromonas* (17), and *Bathycoccus* (18)—because of its modest size (<300 kb), as well as its lower guanine-cytosine (GC) content, gene density, and expression rate as compared to the other chromosomes.



**Fig. 2. Comparison of SOC from *O. mediterraneus* RCC2590 virus-resistant (R) and virus-susceptible (S) lines.** (A) Karyotyping of R and S lines by PFGE (left; black background) and in-gel hybridization with SOC-specific probes (right; gray background). (B) Nucleotide sequence alignment of the SOCs of R and S1 lines shows that they are virtually identical (identical regions linked with yellow blocks) except for a 58-kb deletion at the end of the S1 SOC. The deletion consists of short sequences with significant identity to the rest of the SOC, shown by red and blue blocks denoting sense and antisense matches, respectively. (C) Zoomed-in view of the genomic map of the 58-kb deletion (blue outline). Gray blocks are genes with significant identity to other SOC loci. Magenta blocks show unique genes deleted in the S1 SOC with the table below listing their gene identifiers (prefixed by Ostme30g), putative functional description, and gene length. bp, base pairs; RNase, ribonuclease.

To determine whether a change in SOC size was exclusive to an R-to-S switch, we karyotyped a selection of 10 lines from the mutation accumulation experiment that remained R after passage through single-cell bottlenecks. R lines showed variation of  $\pm 5$  kb in the size of the SOC compared to the ML (fig. S2). To determine whether S-to-R changes were also associated to a SOC size switch, three S lines that had regenerated the R phenotype after reintroduction to OmV2 (S to R) were karyotyped, and no visible karyotype change could be observed (fig. S2). Together, this supports there being minor size variation in the SOC in R lines, a large reduction in SOC size linked only to the R-to-S switch, and no systematic visible SOC change linked to an S-to-R switch.

One S line, S1, was selected for long-read genome resequencing to elucidate the changes involved in the R-to-S switch. This line was chosen, as it had the smallest deletion in the SOC and could show the smallest change linked to a switch in phenotype. Resequencing showed that 58 kb was deleted from one of the extremities of this chromosome (Fig. 2). The SOC comprised a mosaic of short repeated sequences, mainly of intrachromosomal origin,

with some contributions from other chromosomes. As a consequence, the deleted fragment was a pastiche of redundant SOC sequences, with only seven genes deleted that were not found elsewhere on the SOC corresponding to 4323 base pairs (bp) of coding sequence (Fig. 2). Three of the unique deleted genes encoded proteins with unknown functions. Those with putative functional matches included a nucleotide diphosphosugar transferase, an esterase domain-containing family, transmembrane domain-containing protein, and a ribonuclease H1 related to retrotransposases.

### Mathematical modeling of S-R switch

To investigate the consequences of such a phase switch on the microalga and virus population dynamics, we set up a host-virus interaction model of this system (Fig. 3A). S hosts are infected by viral particles present in the environment, which ultimately leads to the lysis of the infected cell and the release of  $v_S$  new viral particles that survive in the environment at a rate  $s_V$ . While reproducing asexually at rates  $a_S$  and  $a_R$ , S algae produce a fraction  $e_R$  of R cells, while R algae generate a fraction  $e_S$  of S cells.

In the absence of viruses, this model predicts the exponential growth typically observed in microalgae populations (Fig. 3B). The virus-free population growth rate converges toward an asymptotic value

$$\lambda_{free} = \frac{1}{2} \left\{ (A + B) + \sqrt{(A - B)^2 + 4 a_S a_R e_S e_R} \right\}$$

where  $A = (1 - e_R)a_S$  and  $B = (1 - e_S)a_R$  represent the rates at which S and R cells produce new progeny cells of the same kind as themselves. In these circumstances, the S/R ratio equals  $e_S a_R / (\lambda_{free} - A)$ , which, for presumably small switching rates ( $e_S, e_R$ ) and considering a cost to resistance, i.e.,  $a_S > a_R$ , leads to a larger fraction of S cells in the population, as shown in Fig. 3B.

When the virus is introduced, the population size drops as S cells are eliminated from the population (Fig. 3A) but only until the R cells, which were previously produced by the switch of S cells, take over the population (Fig. 3B). During this transitory stage that only lasts for a few days, the S/R ratio changes and converges toward  $e_S / (1 - e_S)$ , which, for small values of  $e_S$ , indicates that R cells dominate the population. The dynamics then enters a second asymptotic stage where the infection is maintained because R cells produce S cells that sustain the viral population (Fig. 3A). The growth rate of such infected population quickly converges to

$$\lambda_{inf} \approx B$$

and the proportion of virus,  $V$ , to microalga,  $M$  ( $M = S + R$ ), can then be approximated by

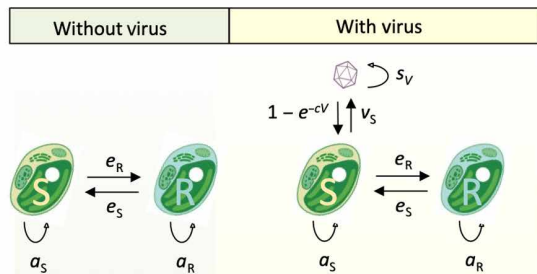
$$\frac{V}{M} \approx \frac{v_S e_S}{a_R (1 - e_S)}$$

The proportion of microalga relative to viruses is then defined by the growth rate of R microalga, the burst size ( $v_S$ ), and the switch rate from R to S ( $e_S$ ). Numerical fit of model behavior to observations of Fig. 1 provides indirect estimates of  $e_S$  between 0.1 and 0.2 and  $v_S$  between 5 and 30, in accordance with the expected range of biological values in this system.

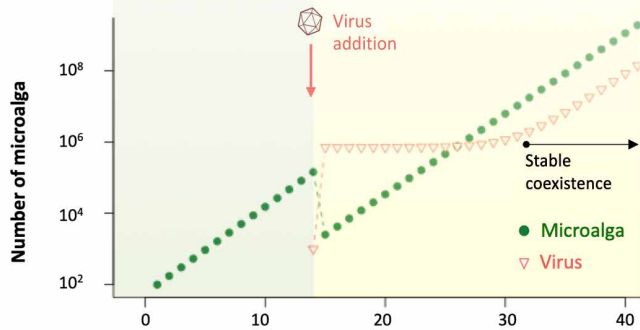
### Genome sequences of *O. mediterraneus* and OmV2

To investigate the genomic bases of this interaction, we performed thorough genome assembly and annotation of the original RCC2590

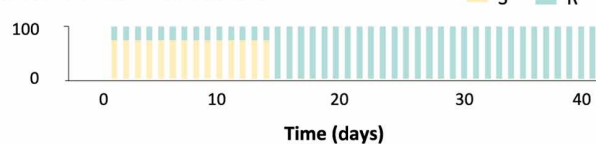
## A Model of Resistant (R) Susceptible (S) phase switch



## B Microalga and virus dynamics



## C Proportion of S and R microalga (%)



**Fig. 3. Model predictions of microalga-virus population dynamics.** (A) Model parameters with and without virus. (B) Simulation of the evolution of the number of microalga cells during 40 days; viruses are introduced after 13 days, and the system converged to its virus/microalga equilibrium behavior after 18 days ( $a_S = 1.75$ ,  $a_R = 1.7$ , and  $e_S = e_R = 0.01$ ). (C) Evolution of the proportion of susceptible (S) and resistant (R) microalga; initial S and R values have been set to reflect equilibrium proportion without virus.

culture. Probes designed to unique regions of three chromosomes are as follows: the chromosome carrying the candidate sexual mating-type locus (17, 19); the predicted SOC; and chromosome 17, carrying the 18S rRNA gene, hybridized to chromosomal bands in the gel corresponding to their assembled sizes (fig. S3). In addition, 16 scaffolds were bordered by two telomeric sequences, providing further confirmation that this genome assembly comprised a majority of complete chromosome sequences. Complete circular assemblies were also obtained for the chloroplast and the mitochondrion, which were similar to the organellar genomes of other *Ostreococcus* species in size, topology, the presence of segmental duplications, and gene content (20). Together, *O. mediterraneus* displayed a similar number of chromosomes and genes, as well as the hallmarks of genome streamlining seen in other Mamiellophyceae genomes, including relatively small overall genome size and high coding density (Table 1). Two notable features of this genome were the presence of the largest SOC so far recorded, which was more than three times the average size of other sequenced Mamiellophyceae, and a relatively low GC content (56%) compared to other *Ostreococcus* species (Table 1). On the basis of an alignment of protein-coding regions between *O. tauri* and *O. mediterraneus*, all chromosomes, except for the SOC and the putative mating locus, shared a high degree of colinearity (fig. S4).

The assembly of OmV2 resulted in a linear 192-kb-long sequence with a mean GC content of 40.7% and encoded 221 predicted genes, similar to other sequenced prasinoviruses (14, 21–23). Visualization of the DNA extraction of *O. mediterraneus* cultures by PFGE and subsequent hybridization with OmV2-specific probes revealed a band at ~200 kb (fig. S3), consistent with the assembled genome size and a linear structure. This hybridization demonstrated that the viral genome was not integrated into one of the *O. mediterraneus* chromosomes. The abundance of OmV2 in the DNA extracted for the PFGE was so low as not to be visually detectable as a band by ethidium bromide staining (indicating <1 ng of OmV2 in the PFGE). The coverage of the OmV2 genome was ~0.5× that of the microalga nuclear genome, suggesting that, at the time of sequencing and following culture centrifugation and DNA extraction protocol, OmV2 genome copies were half as abundant as the microalga.

Genomic insights into an ancestral *Ostreococcus*-prasinovirus interaction

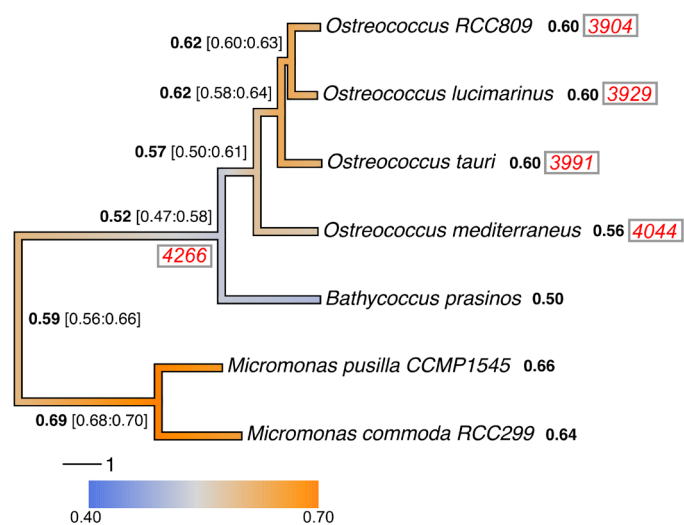
We took advantage of the available Mamiellales and prasinovirus genomes to infer the ancestral *Ostreococcus* genome architecture and test the long-term coexistence scenario. Phylogenomic analysis of 188 single-copy gene families shared among all Chlorophyta (fig. S4) and from 2161 single-copy Mamiellales orthologs placed *O. mediterraneus* at the earliest diverging branch within the *Ostreococcus* genus (Fig. 4), consistent with 18S ribosomal DNA phylogeny (13). Protein families gained, lost, or expanded were retrieved in *O. mediterraneus* by comparative protein family analysis on the fully sequenced Mamiellales genomes. This showed that all *Ostreococcus* species have lost 261 of the 5096 protein families present in all Mamiellales (Fig. 4), which represent probable ancestral Mamiellales functions and thus reflect the general trend of genome reduction in *Ostreococcus* (Table 1). Nonetheless, *O. mediterraneus* has retained relatively more of these ancestral protein families, of which 72 are uniquely present in *O. mediterraneus* and lost in the other *Ostreococcus* species.

The lower genomic GC content of *O. mediterraneus* as compared to other *Ostreococcus* species raised the question of whether it descended from a high-GC content ancestor and experienced a shift toward a lower GC composition or whether it descended from a low-GC content ancestor and that an increase in GC occurred in the rest of the *Ostreococcus* clade. To discriminate between these two alternative hypotheses, we inferred the ancestral genomic GC content for all ancestors of the species tree using ancestral sequence reconstruction techniques. Because current GC contents vary between 0.5 and 0.65, we reconstructed ancestral sequences with a nonhomogeneous codon substitution model to allow each lineage to vary in global nucleotide compositions. This analysis showed that the last common ancestor of the *Ostreococcus* clade had a GC content of ~0.57, similar to that of *O. mediterraneus* and lower than that of *O. tauri*, *Ostreococcus lucimarinus*, and *Ostreococcus* sp. RCC809 (~0.60) (Fig. 4). This global trend is consistent across all codon positions although to a much lesser degree for the highly constrained second codon position (fig. S4). This genomic GC content increase occurred twice within Mamiellophyceae, once within *Ostreococcus*, and once in the other in the branch leading to the *Micromonas* lineage (Fig. 4).

Whole-genome alignment of OmV2 against the model *Ostreococcus* virus OtV5 (14) revealed large syntenic regions and overall conservation of gene order interspersed with virus-specific gene gains and losses (Fig. 5A and table S1). Phylogenetic analysis of the widely

**Table 1. Comparison of the genomic features of sequenced Mamiellophyceae.** chr. no., chromosome number; –, not determined.

|                             | <i>O. mediterraneus</i> | <i>O. tauri</i> | <i>O. lucimarinus</i> | <i>Ostreococcus</i> sp. | <i>B. prasinus</i> | <i>M. pusilla</i> | <i>M. commoda</i> |
|-----------------------------|-------------------------|-----------------|-----------------------|-------------------------|--------------------|-------------------|-------------------|
|                             | RCC2590                 | RCC4221         | CCE9901               | RCC809                  | RCC1105            | CCMP1545          | RCC299            |
| Size (Mb)                   | 13.86                   | 12.91           | 13.20                 | 13.30                   | 14.95              | 21.9              | 20.9              |
| No. chromosome              | 20                      | 20              | 21                    | 20                      | 19                 | 19                | 20                |
| GC content (%)              | 56                      | 60              | 60                    | 60                      | 48                 | 65                | 64                |
| Genes                       | 8,110                   | 7,749           | 7603                  | 7,492                   | 7,847              | 10,587            | 10,286            |
| Genes/kb                    | 0.58                    | 0.60            | 0.58                  | 0.56                    | 0.52               | 0.48              | 0.49              |
| Monoexonic genes (%)        | 77                      | 81              | 79                    | 81                      | 86                 | 47                | 63                |
| BOC size (kb)<br>[chr. no.] | 874 [4]                 | 1,217 [2]       | 895 [2]               | 822 [2]                 | 663 [14]           | 2,172 [2]         | 2,053 [1]         |
| Candidate mating type       | plus                    | minus           | minus                 | plus                    | –                  | –                 | –                 |
| SOC size (kb)<br>[chr. no.] | 637 [12]                | 312 [19]        | 150 [18]              | 102 [18]                | 146 [19]           | 242 [19]          | 195 [17]          |
| Mitochondrion size (bp)     | 49,312                  | 44,237          | –                     | 43,168                  | 43,614             | –                 | 47,425            |
| Chloroplast size (bp)       | 74,215                  | 71,666          | –                     | 67,238                  | 72,700             | –                 | 72,585            |

**Fig. 4. Phylogeny of fully sequenced Mamiellales genomes, gene family gain and loss, and reconstructed ancestral global GC contents and protein families.**

The phylogenetic tree was calculated from random subsampling of 2161 Mamiellales single-copy orthologous gene alignments. The branch colors correspond to the estimated changes in GC over time. The average genome-wide GC contents of extant genomes are shown after the species in bold, and the ancestral estimates are shown at the nodes with ranges in square brackets (see Materials and Methods for the GC evolution estimates). The total number of protein families common to *Bathycoccus* and *Micromonas* are shown in red italics between parentheses at the node leading to those genera, and the numbers of those orthologs present in extant *Ostreococcus* are shown after the species names.

used NCLDV taxonomic marker, the DNA polymerase B (*PolB*) gene, placed OmV2 with viruses infecting *Ostreococcus* spp. within the *Prasinovirus* (Fig. 5B). The *PolB* phylogeny regrouped viruses infecting the microalgae from the same genus, while some viral

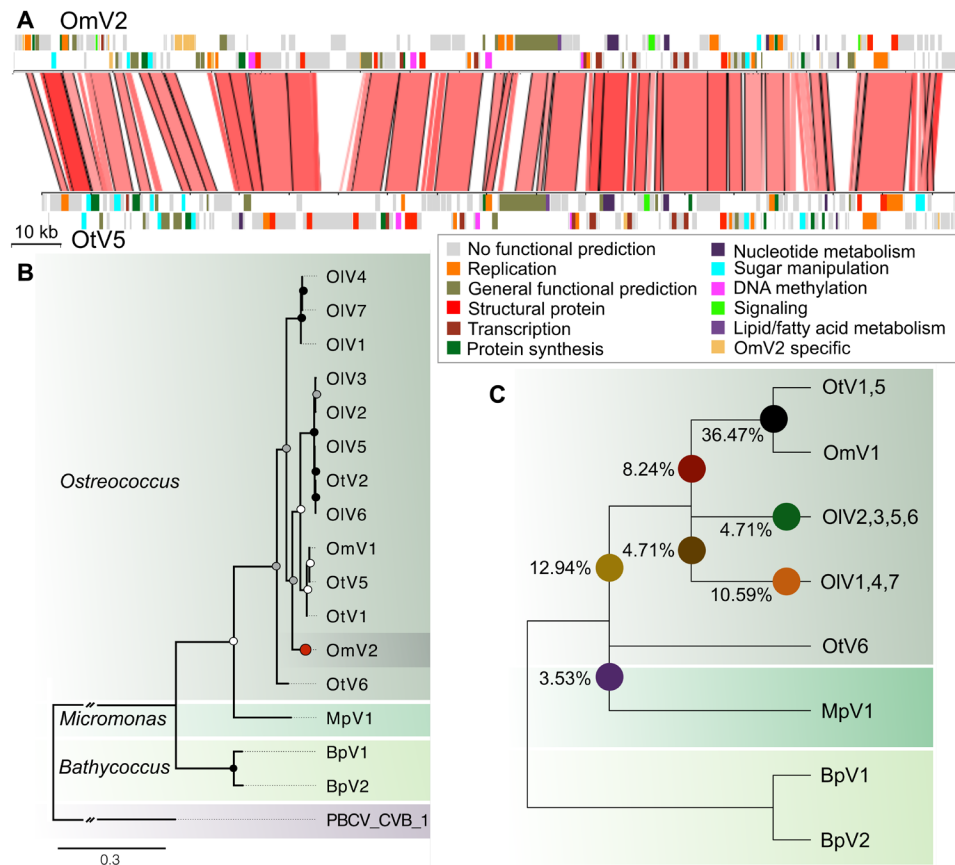
clades contain members that were isolated from different species. To explore the evolutionary relationship of OmV2 with other prasinoviruses, we identified 85 core orthologous proteins shared between the 15 complete prasinovirus genomes. Phylogenies computed for each of these genes showed that most orthologs supported the monophyly of *Ostreococcus* viruses, with only 3.5% clustering with *Micromonas pusilla virus 1* (MpV1) (Fig. 5C). To highlight the incongruities of the gene tree topologies within the *Ostreococcus* virus phylogeny, we estimated the percentage of genes supporting the position of OmV2 on different branches. The majority (36%) of OmV2 genes clustered within the OmV1, OtV1, and OtV5 clade (Fig. 5B), as compared to 8% of OmV2 genes supporting a more basal position.

## DISCUSSION

### A decade of coexistence of microalgae and prasinovirus reveals a reversible susceptible-resistant phase switch

The study of a microalgae-virus system maintained for more than a decade provided evidence of two phenotypes: an S and an R phenotype, as well as the existence of an R-to-S phase switch within an isogenic culture undergoing clonal reproduction. Previous analyses of the related *O. tauri*–OtV system have shown that *O. tauri* strains resistant to a broader range of virus strains tend to have larger SOC sizes (19) and increased differential expression of SOC-encoded genes (24). As the spontaneous mutation rate associated to random errors of the replication machinery (25) is too low to account for the rapidity and prevalence of the acquisition of viral resistance (text S1), we suggest that the phase switch is associated to an immunity-specific mutation process involving a programmed modification of the SOC, as observed in *O. tauri* switching from S to R (24) but, in *O. mediterraneus*, was linked to an R-to-S switch.

The cell immunity toolkit is notorious for combining genetic and nongenetic mechanisms to produce phenotypic plasticity from a single genome (26). While we can only speculate about the genes involved,



**Fig. 5. Phylogenomic analysis of prasinovirus OmV2.** (A) Whole-genome alignment of OmV2 compared to the *O. tauri* virus, OtV5. Genomic regions with 60 to 100% nucleotide identity are joined by red blocks with higher identity indicated by darker red. Genes are colored according to functional groups. (B) Maximum likelihood phylogenetic tree of amino acid sequences of the DNA polymerase B gene from prasinoviruses infecting the genera *Ostreococcus*, *Micromonas*, and *Bathycoccus*. The tree was rooted with PBCV-1 (Chlorovirus) with the connecting branch truncated for display. The red circle highlights the position of OmV2. Circles at the nodes represent bootstrap support of 50 to 70% (white), 70 to 90% (gray), and 90 to 100% (black). The scale bar shows substitutions per site. (C) Cladogram representing the relationships between the fully sequenced prasinovirus clades from 85 core genes. Colored circles mark the branching positions of the percentage of the OmV2 orthologs.

as there are no orthologous genes in the SOC of the Mamiellales sequenced to date, the same kinds of gene functions are overrepresented in the SOCs of all of these species, suggesting that variation in these functional genes imparts viral immunity in Mamiellales. Our results support the proposed function of the SOC as a specialized chromosome involved in antiviral immunity, evidence for which has been provided in two species so far: *O. tauri* and *O. mediterraneus*.

Analysis of the whole gene content of *O. mediterraneus* confirmed its phylogenetic position at the base of the *Ostreococcus* lineage. Maximum likelihood inference suggests that the *Ostreococcus* ancestor had a GC-poorer genome and contained a larger number of genes than the present species, revealing a global trend toward genome streamlining (27). The comparison of the genome sequence of its resident prasinovirus, OmV2, with available prasinovirus sequence reflected the story of a highly dynamic virus genome with recombination between related viral clades (28) and independent horizontal gene transfers from different microalga hosts into the virus genome (29). Together, phylogenetic analysis of OmV2 genes dismisses coevolution of this host-virus pair since the speciation of *O. mediterraneus*. Instead, it is consistent with substantial gene exchange between the ancestors of present-day species-specific prasinoviruses (22).

### How does this phase switch translate into microalga-virus population dynamics?

Mathematical modeling of the phase switch is consistent with observed continuous virus production of OmV2 by *O. mediterraneus*, which is a consequence of the production of new viruses by S cells that are generated from R cells. This model does not generate oscillatory dynamics after an initial lysis event, contrary to many previous host-lytic virus coexistence models (11). The recently proposed “leaky resistance” model from a bacteria-bacteriophage system does also predict continuous virus and host growth and could also be invoked to describe the R-to-S switch (30). The population collapse observed upon infection in culture is only predicted if S microalgae make up most of the population. This is predicted either if there is a bias between the S-to-R and the R-to-S switch ( $e_S > e_R$ ) or if S microalgae have a higher growth rate than R microalgae, producing a trade-off between resistance and growth. While there is evidence of this trade-off in many bacterial-bacteriophage systems (31), there is yet little evidence for a similar trade-off in microalgae (32). Our model is consistent with these findings in predicting, over longer time periods, an equilibrium S/R ratio depending on four parameters that are independent of the phenotype of the founder cell. Long-term competition experiments that started with an equal

number of S and R *O. tauri* cells resulted in the evolution of cultures with a majority of S cells (10), consistent with either a higher growth rate of S cells and thus a cost of resistance or  $e_S > e_R$ . In the present microalga-virus system, the cost of resistance may be metabolic, as a consequence of an increase in the size of the SOC, which requires additional energy and resources for cell replication.

One fundamental property of a life history strategy is its evolutionary stability, which can be estimated by investigating whether a population adopting one strategy can be invaded by individuals playing another (33). A microalga with a pure S strategy (it cannot convert from S to R) cannot invade a culture with the resistant-susceptible switch strategy (hereafter referred to as RS strategy) in the presence of viruses, as the viruses will lyse it. It will, however, invade an RS culture in the absence of viruses since microalgae with a pure S strategy divide more rapidly than the R microalgae. Now, a microalga with a pure R strategy (not generating S cells) will not invade an RS population in the absence of viruses because of its lower growth rate than the S microalgae. However, it will invade the RS population in the presence of viruses, as the R cells that do not switch to S will invade the population. The RS strategy that we describe is therefore not evolutionarily stable, neither in the presence nor in the absence of viruses, so the most obvious explanation for its persistence comes from the evolutionary theory of bet-hedging (34) that demonstrates that a strategy may be suboptimal under one environmental condition, but advantageous in the long run, when environmental conditions alternate randomly. The RS strategy outcompetes both a pure S and a pure R strategy if virus presence and absence alternates (34). A droplet of seawater can certainly experience changes in virus prevalence over short time scales, by mechanisms of marine virus decay (35) or mixing of water masses or rain (36). Last but not least, the RS strategy may confer a substantial advantage by lysing competitors because strains propagating a virus have a higher probability of invasion in the face of competitors that may be susceptible to the virus produced (37).

### What is the prevalence of the RS strategy in the environment?

Within the simplified one-microalga, one-virus experimental set up, our model makes two predictions. First, it predicts the systematic recovery of cell density after virus-induced lysis. This has been reported in many different species from divergent branches of the eukaryotic tree of life: Haptophyceae, Stramenopiles, and Chlorophytes (8, 24, 38). Second, it predicts the steady production of viruses in apparently healthy cultures. While this has been explicitly reported in *E. huxleyi*, *Phaeocystis pouchetii* (8), and *O. tauri* (10), it has also been observed in the dynamics of *C. variabilis* (11). However, it is difficult to extend this single-species dynamic to the community scale, where several microalga-virus pairs compete for the same resources (39). A general prediction that we venture is that introducing RS microalgae in community models may increase the overall stability of the host-virus dynamics. This model is consistent with previously reported non-“boom and bust” dynamics in the natural environment (40), including persistent infection (41). Transcriptomes of several hundred species from the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) initiative revealed the expression of core NCLDV genes in 13 cultures (42). While this may merely indicate infection, one of the haptophyte strains sequenced had been described as persistently infected (42). Screening six *O. mediterraneus* transcriptomes from the MMETSP (table S2)

for OmV2 sequences, we found evidence for a virus in strain RCC2596. This strain was isolated on the same day and site as the genome-sequenced strain RCC2590, and although closely related, they are nonidentical, having approximately five differences per 10,000 bp and different karyotypes (13). RCC2596 was still available in culture in our laboratory, and we could thus experimentally confirm the presence of viral DNA in the culture by PCR. Given that these strains were independently isolated, this implies that stable coexistence was widespread in this population in the environment but is simply not detected during the isolation of algae as the bulk culture grows reliably. We thus speculate that this phase switching is a common strategy within NCLDV-microalga interactions, in addition to mutation and selection mechanisms that lead to pure R microalga strains and the arms race type of interaction. This phase switch is the result of a yet to be described molecular mechanism of cell immunity in microalgae, mirroring the complexity of the NCLDV genomes that they have been coevolving with since the early stages of eukaryotic life (43).

## MATERIALS AND METHODS

### Cultures and growth conditions

All *O. mediterraneus* strains were maintained in liquid L1 medium (National Center for Marine Algae and Microbiota, Bigelow Laboratory for Ocean Sciences, USA) without addition of  $\text{Na}_2\text{SiO}_3$ , prepared using autoclaved seawater from offshore Banyuls Bay (MOLA station: 42°27'11"N, 3°32'36"E) diluted 10% with Milli-Q water, and filtered before use through 0.22- $\mu\text{m}$  filters, under a 12-hour light/12-hour dark regime. Incubation temperature was 15°C for measuring growth curves in population dynamics experiments and 20°C for all other experiments. The strains used in this study and the genomic and transcriptomic data obtained for them are listed in table S2.

### PFGE and in-gel hybridization

PFGE and in-gel hybridization enabled estimation of chromosome sizes in different strains and testing for the presence of the virus genome. PFGE and in-gel hybridization were performed as previously described with modifications as follows (44). Cells ( $8.7 \times 10^7$ ) in exponential growth phase were harvested by centrifugation (8000g for 20 min), resuspended in 150  $\mu\text{l}$  of TE buffer [10 mM tris-HCl and 125 mM EDTA (pH 8)], and then embedded in plugs by mixing with an equal volume of molten low-melting point agarose (1% in TE buffer precooled to 45°C). After setting, the cells were lysed in the agarose plugs by immersing the plugs in proteinase K buffer [10 mM tris HCl (pH 8), 0.5 M EDTA (pH 8), 1% lauroyl sarcosinate, and proteinase K (1 mg  $\text{ml}^{-1}$ ) (final concentration)] at 37°C for 24 hours under constant agitation. The plugs were then washed three times in EDTA solution [0.5 M (pH 8)] at 37°C for 2 hours under constant agitation and then stored in EDTA solution at 4°C. PFGE was performed in 0.8% agarose gels in 0.5 $\times$  tris-borate EDTA buffer [44.5 mM tris, 44.5 mM boric acid, and 1 mM EDTA (pH 8)] with 2 mm of plug from each sample loaded into the wells. Electrophoresis was run in the CHEF-DR III (Bio-Rad) system at 14°C and 3 V  $\text{cm}^{-1}$  with 120° pulse angle for 45 hours with a switch time of 90 s, followed by 27 hours at a switch time of 140 s. After PFGE, DNA was visualized by immersing the gel in a solution of ethidium bromide. The gel was moved to a flat-bottomed dish, and enough 0.4 M NaOH was added to cover the gel for chemical denaturation of the DNA by incubation for 30 min at room temperature with



agitation. Then, the gel was washed three times for 10 min in a 6× SSC solution [20× SSC, 3 M NaCl, and 0.3 M Na<sub>3</sub>citrate (pH 7)]. Last, the gel was dehydrated under vacuum (GD2000, Hoefer Inc., Richmond, CA, USA).

PCR (GoTaq, Promega) was used to amplify specific regions of the SOC, Big Outlier Chromosome (BOC), OmV2, and the 18S rRNA gene (see table S3 for primer sequences) for use as DNA probes with the following conditions: initial denaturation for 2 min at 94°C; 30 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 55°C, and extension for 45 s at 72°C; and final extension for 5 min at 72°C. The PCR product was cleaned with Wizard (Promega) SV Gel and PCR Clean-Up System. The amplicons were randomly labeled with [ $\alpha$ -<sup>32</sup>P]cytidine triphosphate (PerkinElmer catalog no. BLU008H250UC) according to the manufacturer's instructions (Prime-a-Gene kit, Promega, catalog no. U1100) for use as DNA probes. Dried gels were equilibrated in sufficient hybridization buffer [6× SSC, 5× Denhardt's solution, 0.1% (v) SDS, and tRNA (10  $\mu$ g ml<sup>-1</sup>)] to cover the gels for 3 hours at 65°C with constant agitation, after which radiolabeled probes were added and hybridized overnight at 65°C with constant agitation. After hybridization, the gel was subjected to a series of washes with constant agitation: three washes in 6× SSC buffer at room temperature for 30 min, one wash in 6× SSC buffer at 65°C for 1.5 min, followed by a final wash in 3× SSC buffer at room temperature for 5 min. The gel was sealed in a transparent plastic sleeve and then exposed to radiographic film.

### Transmission electron microscopy

RCC2590 culture in exponential growth phase was centrifuged for 20 min at 6000g before inclusion in agar 1% to avoid successive centrifugations. Plugs were embedded in epoxy resin (Epon 812) after fixation with 2.5% glutaraldehyde. Images of individual cells and cells deemed to be visibly infected if clear virus-like particles were clearly present within the cell were captured.

### Genome sequencing and assembly

Genome library preparation, sequencing, assembly, and optical mapping were performed by Genoscope ([www.genoscope.cns.fr](http://www.genoscope.cns.fr)). Two libraries were prepared and sequenced using the Illumina HiSeq platform; a paired-end (PE) read library of 272-bp insert size and a mate-pair (MP) library of five 230-bp insert size. DNA sequence reads were assembled using ALLPATHS-LG assembler (45). At this stage, 99 scaffolds were obtained that aligned with eukaryotic sequences. Optical mapping was performed to further place the *O. mediterraneus* scaffolds in their correct order on the chromosome. This resulted in 19 optically mapped scaffolds corresponding to nuclear chromosomes and several low-GC scaffolds corresponding to fragments of the SOC, organelles, and OmV2. The organelles were assembled using ABySS from the Illumina HiSeq PE reads using the MP library for scaffolding (abyss-pe, lib = "PE," mp = "MP," k = 85, n = 10). Organelle genome completeness was evaluated by alignment to the complete organelle sequences of *O. tauri*. The two large contigs of the OmV2 genome obtained from the Illumina sequencing were scaffolded together manually using the assembly and visualization tools available from the Geneious software ([www.geneious.com](http://www.geneious.com)).

To resolve the highly repetitive SOC sequence, genomic DNA from RCC2590 and the evolved line MA3 were resequenced with the Single Molecule, Real-Time (SMRT) technology PacBio RS II by the GeT-PlaGe sequencing platform in Toulouse, France as previously described (19). The raw data were assembled using HGAP3 (version

2.3.0). The PacBio assemblies were checked manually with assembly and visualization tools available from the Geneious software to identify the scaffolds corresponding to the SOC. Both RCC2590 and MA3 SOCs were assembled from telomere to telomere, the former comprising five scaffolds and the latter a single scaffold.

### Genome annotation

Briefly, gene models were predicted using EuGene (release 4.1; [eugene.toulouse.inra.fr](http://eugene.toulouse.inra.fr)). Transcript information was provided to the gene caller as follows: one PE Illumina RNA sequencing (RNA-seq) dataset was assembled into transcript contigs using a De Bruijn Graph (DBG) assembly approach with high Kmer sizes to boost specificity during the assembly. Six *O. mediterraneus* strains (table S2) had their transcriptome sequenced as part of the MMETSP. The resulting transcriptomes are referenced in table S2 with their respective MMETSP identifiers.

The obtained set of contigs was cleaned from any contigs smaller than 300 bp, completed with expressed sequence tag sequences collected from National Center for Biotechnology Information (>300 bp) and mapped on the genomic scaffolds by combining BLASTN (for rapid detection on regions of interest) and GenomeThreader (for fine mapping, taking splice sites into account) with parameters set to map at least 98% of the transcript sequence on the regions as reported by BLASTN. Besides, RNA-seq was also mapped directly onto the genome using HISAT2, and junctions were extracted from the BAM alignments with RegTools. The obtained junction instances were counted and grouped on the basis of the intron coordinates that was spanned by a read. This count (or coverage) was used to filter for junctions confirmed by a minimum of three reads. Functional descriptions were inferred from filtered best BLAST hits and combined with protein domains obtained with InterPro. This automatic annotation was curated using the Online Resource for Community Annotation of Eukaryotes (ORCAE) portal ([bioinformatics.psb.ugent.be/orcae/](http://bioinformatics.psb.ugent.be/orcae/)). All predicted genes were manually inspected, and from the 8483 protein-coding genes, 1142 have been modified (including gene fusions) and 59 were discarded, resulting in a final set of 8110 protein-coding genes. The functional description was manually edited for 662 genes.

Gene prediction in OmV2 was performed using Glimmer implemented within Geneious. Functional description of predicted OmV2 genes was done using BLASTX against the non-redundant (NR) database (from December 2017) with default parameters and against the Pfam version 31.0 database (<http://pfam.xfam.org/>). tRNAs were predicted using the tRNAscan-SE 2.0 web server.

### Comparative genome analysis

The genome and predicted genes from *O. mediterraneus* RCC2590 were loaded into a custom instance of pico-PLAZA ([bioinformatics.psb.ugent.be/plaza/versions/pico-plaza/](http://bioinformatics.psb.ugent.be/plaza/versions/pico-plaza/)) containing 38 eukaryotic genomes (including Metazoa, Fungi, Chlorophyta, Embryophyta, Rhodophyta, Haptobionta, and Stramenopiles; table S4). Following an "all-versus-all" BLASTP (version 2.2.27+; e-value threshold, 10<sup>-5</sup>; max hits, 500) protein sequence similarity search, both TribeMCL (version 10-201) and OrthoMCL (version 2.0; mcl inflation factor, 3.0) were used to delineate gene families and subfamilies, respectively.

Phylogenomic analysis of the *O. mediterraneus* nuclear genome was performed by concatenating the amino acid sequence from 188 single-copy orthologs by TribeMCL within the Chlorophyta genomes (*O. mediterraneus*, *O. tauri*, *O. lucimarinus*, *O. sp. RCC809*, *Micromonas pusilla strain CCMP1545*, *Micromonas sp. RCC299*,

*Bathycoccus prasinos*, *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Coccomyxa subellipsoidea* C-169, *C. variabilis* NC64A, *Oryza sativa*, *Physcomitrella patens*, and *Volvox carteri*) aligned using MUSCLE (version 3.8.31). This resulted in alignment of 148,028 positions of which 88,784 were informative. Similarly, the phylogenetic topology across Mamiellales was determined by concatenating the alignment of 2161 single-copy genes, generated using MUSCLE, resulting in an alignment of 1,267,142 positions of which 444,911 were informative. The phylogenetic trees were generated using RAxML (version 8.2.8) (model, PROTGAMMAWAG) with 100 bootstrap replicates.

The phylogenetic profile of TribeMCL gene families (excluding orphans) and the inferred Mamiellales tree topology were provided to reconstruct the most parsimonious gain and loss scenario for every gene family using the Dollop program from the PHYLIP package (version 3.69).

### Prasinovirus phylogenomics

Eighty-five putative orthologous prasinovirus genes were identified for phylogenetic analyses by TBLASTX of predicted OmV2 genes against a database of all genes from 15 fully sequenced prasinovirus genomes (e-value,  $<10^{-4}$ ; amino acid identity,  $>75\%$ ; and alignment length,  $>50$ ) accepting as “core” those orthologs present in all genomes. Each ortholog family was aligned using MAFFT, and phylogenetic trees were reconstructed using Bayesian inference with MrBayes using a mixed model for amino acid with four chains of  $10^6$  generations, trees sampled every 100 generations, and burnin value set to 20% of the sampled trees. We checked that the SD of the split frequencies fell below 0.01 to ensure convergence in tree search.

### Analysis of the evolution of GC content

To reconstruct genomic GC contents along the tree of *Ostreococcus*, we used an ancestral sequence reconstruction approach to reconstruct gene sequences at the codon level and compute their GC contents at the tree nodes. We included three outgroups (*B. prasinos*, *M. pusilla* CCMP1545, and *Micromonas commoda* RCC299) in the reconstruction of ancestral GC contents of the four *Ostreococcus* lineages and selected all 2161 single-copy genes universally present in these seven genomes, as determined from the comparative genomic analysis. We aligned all 2161 universal single-copy protein families at the codon level with MACSE using default parameters and trimmed alignments for badly aligned positions with BMGE using a BLOSUM62 substitution matrix and other parameters set to their default values.

To both reduce computational burden and produce confidence intervals in estimates of ancestral GC contents, we used a jackknife approach to subsample the 2161 alignments and to build 100 concatenates containing each 100 randomly sampled genes. Each concatenate has an average size of 104,290 nucleotide positions ( $\pm 8644$ ), representing on average 34,763 codons, and so is expected to contain a sufficient amount of phylogenetic information to reconstruct ancestral GC contents. We further verified that these concatenates are faithful representations of each genome in terms of GC contents. We computed correlation coefficients between the average GC contents computed across all 2161 genes and the GC contents of each of the 100 concatenates. We observed that all correlation coefficients were higher than 0.972, showing that these 100 jackknife subsamples have the same properties than the original population of 2161 genes, which allows us to infer ancestral genomic GC contents.

We first reconstructed the phylogenetic tree of each of the 100 concatenates to verify that each concatenate contains the same phylogenetic information on the relationships between the seven considered species. We built trees with FastTree (version 2.1.8) using a generalized time-reversible (GTR) model and the per-site rate CATegories (CAT) approximation with 20 categories to model the variation of evolutionary rates across sites. Each phylogenetic reconstruction yielded a tree consistent with the reference species tree of the seven species. We used this topology to reconstruct ancestral sequences on all concatenates. We used the Yang and Nielsen (YN) substitution model for codons with different base frequencies at the three codon positions (F3x4) to model codon evolution. As the seven species have genomic GC contents that vary from  $\sim 0.5$  to  $\sim 0.65$ , the GC composition has not evolved homogeneously in time, as was previously observed in many prokaryotic and eukaryotic lineages. For this reason, we considered a nonhomogeneous and nonstationary approach in which stationary frequencies are allowed to vary from one lineage to another on the species tree. Each branch of the species tree is assigned its own set of parameters of the YN model, and the stationary codon frequencies at the root of the species are also free to vary. We used maximum likelihood and the BppML program to estimate the parameters of this nonhomogeneous model and the branch lengths of the tree. Using the maximum likelihood values for all these parameters, we reconstructed ancestral sequences with the BppAncestor program, using the marginal ancestral sequence reconstruction (ASR) approach. At a given internal node of the species tree, the codon state having the maximum posterior probability was inferred as the putative ancestral state. We then computed the ancestral global GC content as well as the ancestral GC content of all codon positions from these reconstructed maximum likelihood ancestral sequences. We obtained a distribution of ancestral GCs for each ancestor from which we computed the mean and the 95% confidence interval.

### Population dynamics of virus and microalgae

Microalgae and viruses were counted over 13 days by flow cytometry and plaque assay from triplicate 20-ml cultures in tissue culture flasks at  $15^\circ\text{C}$  in a 12-hour light/12-hour dark regime. Cultures of RCC2590 and S1 were inoculated to have a starting density of  $10^4$  cells  $\text{ml}^{-1}$ . In addition, we passed the RCC2590 culture through a  $0.2\text{-}\mu\text{m}$  filter to remove cells retaining free virions in the medium and kept this filtrate in the same conditions as the cultures to estimate viral particle stability over the duration of the experiment. For flow cytometry counts of cells and viruses,  $250\ \mu\text{l}$  was sampled every 2 days, fixed for 15 min in the dark with a final concentration of electron microscopy-grade glutaraldehyde of 0.25% and Pluronic F-68 of 0.01%, flash-frozen in liquid nitrogen, and stored at  $-80^\circ\text{C}$  until analysis. Cell and viral particle counts were performed with the BD FACSCanto II (Becton Dickinson) where the flow rate over 1 min at a low rate was determined using triplicate measurements of BD Trucount Tube beads. Cells were counted from fixed samples using red autofluorescence and side scatter. Viral counts were performed on fixed samples after rapid thawing at  $40^\circ\text{C}$ , diluting as necessary in filtered seawater ( $<0.02\ \mu\text{m}$ ), staining for 10 min at  $80^\circ\text{C}$  with SYBR Green I (Molecular Probes) ( $0.5\times$  final concentration of commercial stock), and setting the trigger to measure green fluorescence and side scatter. To measure the concentration of infective OmV2 in the RCC2590 culture medium by plaque assay, 1 ml was sampled from the cultures every 2 days and centrifuged at  $25,000g$  for 10 min at  $15^\circ\text{C}$  to remove cells, and the supernatant was serially diluted with L1 medium. Plaque assay was performed with

the diluted supernatant as described in (38) using the OmV2-susceptible *O. mediterraneus* strain S1 as the host.

### Experimental assessment of the resistant-to-susceptible phase switch

Virus-free *O. mediterraneus* lines were obtained by end-point dilution, and virus absence was confirmed by PCR using degenerate primers for the viral DNA polymerase and major capsid protein gene (table S3). We started 204 clonal cultures from one single cell in 1-ml microwells and tested for virus susceptibility after 7 days. Virus susceptibility was tested using cell-free, virus-producing culture (VPC) medium.

To obtain cell-free VPC medium (i.e., separate potential virions from cell lysates), medium from growing or lysed cultures was centrifuged at 25,000g for 10 min at room temperature and the supernatant was filtered (<0.45 μm). To test whether a cell line was susceptible to lysis, VPC medium was introduced into the growing test culture, with a control culture grown in parallel inoculated with the same volume of sterile medium, and visually inspected for lysis up to 5 days after inoculation. This visual inspection detected when most cells are lysed, and thus, the ancestral cell was inferred to be S. The validity of this inference relied on two assumptions: a low switch rate between the S and R phenotypes and a low difference in growth rates between the R and susceptible phenotypes.

### Mathematical modeling of the host-virus population dynamics

The model considered a population made of S and R microalga ( $S_t, R_t$ ) and their viruses ( $V_t$ ) at time  $t$ . The changes in microalga and viruses numbers from 1 day ( $t$ ) to another ( $t + 1$ ) were modeled according to a simple representation of the algae and virus life cycles and interactions with the following assumptions: (i) S and R microalgae produce new microalgae at rates  $a_s$  and  $a_r$ , (ii) microalgae are born from a S (R) to switch to a R (S) phenotype at a fixed rate  $e_R$  ( $e_S$ ), (iii) encounters between S algae and virus occur randomly at a rate  $c$ , (iv) each infected microalga produced  $\nu$  viruses before death, and (v) viral particles survived at a rate  $s_V$ . The population dynamic of R, S, and V can then be described by following equations

$$\begin{aligned} S_{t+1} &= (1 - e_R) a_S e^{-cV_t} S_t + e_S a_R R_t \\ R_{t+1} &= e_R a_S e^{-cV_t} S_t + (1 - e_S) a_R R_t \\ V_{t+1} &= \nu_S (1 - e^{-cV_t}) S_t + s_V e^{-cS_t} V_t \end{aligned}$$

where  $e^{-cV_t}$  is the probability drawn in Poisson distributions that a S alga encounters no virus or that a viral particle encounters no S alga to infect. This system was investigated analytically and numerically using the R software.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/14/eaay2587/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. C. A. Suttle, Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
2. C. P. D. Brussaard, S. W. Wilhelm, F. Thingstad, M. G. Weinbauer, G. Bratbak, M. Heldal, S. A. Kimmance, M. Middelboe, K. Nagasaki, J. H. Paul, D. C. Schroeder, C. A. Suttle,

- D. Vaqué, K. E. Wommack, Global-scale processes with a nanoscale drive: The role of marine viruses. *ISME J.* **2**, 575–578 (2008).
3. P. Hingamp, N. Grimsley, S. G. Acinas, C. Clerissi, L. Subirana, J. Poulain, I. Ferrera, H. Sarmiento, E. Villar, G. Lima-Mendez, K. Faust, S. Sunagawa, J.-M. Claverie, H. Moreau, Y. Desdevises, P. Bork, J. Raes, C. de Vargas, E. Karsenti, S. Kandels-Lewis, O. Jaillon, F. Not, S. Pesant, P. Wincker, H. Ogata, Exploring nucleocytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* **7**, 1678–1695 (2013).
4. M. Tragin, D. Vaultot, Green microalgae in marine coastal waters: The Ocean Sampling Day (OSD) dataset. *Sci. Rep.* **8**, 14020 (2018).
5. F. Zhu, R. Massana, F. Not, D. Marie, D. Vaultot, Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* **52**, 79–92 (2005).
6. Z. Erez, I. Steinberger-Levy, M. Shamir, S. Doron, A. Stokar-Avihail, Y. Peleg, S. Melamed, A. Leavitt, A. Savidor, S. Albeck, G. Amitai, R. Sorek, Communication between viruses guides lysis-lysogeny decisions. *Nature* **541**, 488–493 (2017).
7. D. D. Dunigan, L. A. Fitzgerald, J. L. Van Etten, Phycodnaviruses: A peek at genetic diversity. *Virus Res.* **117**, 119–132 (2006).
8. R. Thyrhaug, A. Larsen, T. F. Thingstad, G. Bratbak, Stable coexistence in marine algal host-virus systems. *Mar. Ecol. Prog. Ser.* **254**, 27–35 (2003).
9. J. L. Van Etten, L. C. Lane, R. H. Meints, Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.* **55**, 586–620 (1991).
10. R. Thomas, N. Grimsley, M. Escande, L. Subirana, E. Derelle, H. Moreau, Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. *Environ. Microbiol.* **13**, 1412–1420 (2011).
11. J. Fricke, M. Sieber, L. Becks, Eco-evolutionary dynamics in a coevolving host-virus system. *Ecol. Lett.* **19**, 450–459 (2016).
12. K. D. Weynberg, M. J. Allen, W. H. Wilson, Marine prasinoviruses and their tiny plankton hosts: A review. *Viruses* **9**, 43 (2017).
13. L. Subirana, B. Péguin, S. Michely, M.-L. Escande, J. Meilland, E. Derelle, B. Marin, G. Piganeau, Y. Desdevises, H. Moreau, N. H. Grimsley, Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist* **164**, 643–659 (2013).
14. E. Derelle, C. Ferraz, M.-L. Escande, S. Eychenié, R. Cooke, G. Piganeau, Y. Desdevises, L. Bellec, H. Moreau, N. Grimsley, Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLOS ONE* **3**, e2250 (2008).
15. M. Krasovec, A. Eyre-Walker, N. Grimsley, C. Salmeron, D. Pecqueur, G. Piganeau, S. Sanchez-Ferandin, Fitness effects of spontaneous mutations in picoeukaryotic marine green algae. *G3 (Bethesda)* **6**, 2063–2071 (2016).
16. E. Derelle, C. Ferraz, S. Rombauts, P. Rouzé, A. Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saëys, J. Wuys, K. Jabbari, C. Bowler, O. Panaud, B. Piégu, S. G. Ball, J.-P. Ral, F.-Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, H. Moreau, Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11647–11652 (2006).
17. A. Z. Worden, J.-H. Lee, T. Mock, P. Rouzé, M. P. Simmons, A. L. Aerts, A. E. Allen, M. L. Cuvelier, E. Derelle, M. V. Everett, E. Foulon, J. Grimwood, H. Gundlach, B. Henrissat, C. Napoli, S. M. McDonald, M. S. Parker, S. Rombauts, A. Salamov, P. Von Dassow, J. H. Badger, P. M. Coutinho, E. Demir, I. Dubchak, C. Gentemann, W. Eikrem, J. E. Gready, U. John, W. Lanier, E. A. Lindquist, S. Lucas, K. F. X. Mayer, H. Moreau, F. Not, R. Otilar, O. Panaud, J. Pangilinan, I. Paulsen, B. Piégu, A. Poliakov, S. Robbens, J. Schmutz, E. Toulza, T. Wyss, A. Zelensky, K. Zhou, E. V. Armbrust, D. Bhattacharya, U. W. Goodenough, Y. Van de Peer, I. V. Grigoriev, Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268–272 (2009).
18. H. Moreau, B. Verhelst, A. Couloux, E. Derelle, S. Rombauts, N. Grimsley, M. Van Bel, J. Poulain, M. Katinka, M. F. Hohmann-Mariotti, G. Piganeau, P. Rouzé, C. Da Silva, P. Wincker, Y. Van de Peer, K. Vandepoel, Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* **13**, R74 (2012).
19. R. Blanc-Mathieu, M. Krasovec, M. Hebrard, S. Yau, E. Desgranges, J. Martin, W. Schackwitz, A. Kuo, G. Salin, C. Donnadieu, Y. Desdevises, S. Sanchez-Ferandin, H. Moreau, E. Rivals, I. V. Grigoriev, N. Grimsley, A. Eyre-Walker, G. Piganeau, Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci. Adv.* **3**, e1700239 (2017).
20. S. Robbens, E. Derelle, C. Ferraz, J. Wuys, H. Moreau, Y. Van de Peer, The complete chloroplast and mitochondrial DNA sequence of *Ostreococcus tauri*: Organelle genomes of the smallest eukaryote are examples of compaction. *Mol. Biol. Evol.* **24**, 956–968 (2007).
21. K. D. Weynberg, M. J. Allen, K. Ashelford, D. J. Scanlan, W. H. Wilson, From small hosts come big viruses: The complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ. Microbiol.* **11**, 2821–2839 (2009).

22. H. Moreau, G. Piganeau, Y. Desdevises, R. Cooke, E. Derelle, N. Grimsley, Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J. Virol.* **84**, 12555–12563 (2010).
23. E. Derelle, A. Monier, R. Cooke, A. Z. Worden, N. H. Grimsley, H. Moreau, Diversity of viruses infecting the green microalga *ostreococcus lucimarinus*. *J. Virol.* **89**, 5812–5821 (2015).
24. S. Yau, C. Hemon, E. Derelle, H. Moreau, G. Piganeau, N. Grimsley, A viral immunity chromosome in the marine picoeukaryote, *Ostreococcus tauri*. *PLOS Pathog.* **12**, e1005965 (2016).
25. M. Krasovec, A. Eyre-Walker, S. Sanchez-Ferandin, G. Piganeau, Spontaneous mutation rate in the smallest photosynthetic eukaryotes. *Mol. Biol. Evol.* **34**, 1770–1779 (2017).
26. L. A. Meyers, J. J. Bull, Fighting change with change: Adaptive variation in an uncertain world. *Trends Ecol. Evol.* **17**, 551–557 (2002).
27. Y. I. Wolf, E. V. Koonin, Genome reduction as the dominant mode of evolution. *Bioessays* **35**, 829–837 (2013).
28. C. Clerissi, N. Grimsley, Y. Desdevises, Genetic exchanges of inteins between prasinoviruses (*Phycodnaviridae*). *Evolution* **67**, 18–33 (2013).
29. A. Monier, J.-M. Claverie, H. Ogata, Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *BMC Genomics* **8**, 456 (2007).
30. W. N. Chaudhry, M. Pleška, N. N. Shah, H. Weiss, I. C. McCall, J. R. Meyer, A. Gupta, C. C. Guet, B. R. Levin, Leaky resistance and the conditions for the existence of lytic bacteriophage. *PLOS Biol.* **16**, e2005971 (2018).
31. J. T. Lennon, S. A. M. Khatana, M. F. Marston, J. B. H. Martiny, Is there a cost of virus resistance in marine cyanobacteria? *ISME J.* **1**, 300–312 (2007).
32. S. E. Heath, S. Collins, Mode of resistance to viral lysis affects host growth across multiple environments in the marine picoeukaryote *Ostreococcus tauri*. *Environ. Microbiol.* **18**, 4628–4639 (2016).
33. J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge University Press, 1982).
34. T. Phillippi, J. Seger, *Trends Ecol. Evol.* **4**, 41–44 (1989).
35. K. D. A. Mojica, C. P. D. Brussaard, Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiol. Ecol.* **89**, 495–515 (2014).
36. S. Sharoni, M. Trainic, D. Schatz, Y. Lehahn, M. J. Flores, K. D. Bidle, S. Ben-Dor, Y. Rudich, I. Koren, A. Vardi, Infection of phytoplankton by aerosolized marine viruses. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6643–6647 (2015).
37. S. P. Brown, L. Le Chat, M. De Paeppe, F. Taddei, Ecology of microbial invasions: Amplification allows virus carriers to invade more rapidly when rare. *Curr. Biol.* **16**, 2048–2052 (2006).
38. K. Tarutani, K. Nagasaki, M. Yamaguchi, Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. *Appl. Environ. Microbiol.* **66**, 4916–4920 (2000).
39. C. Winter, T. Bouvier, M. G. Weinbauer, T. F. Thingstad, Trade-offs between competition and defense specialists among unicellular planktonic organisms: The “Killing the Winner” hypothesis revisited. *Microbiol. Mol. Biol. Rev.* **74**, 42–57 (2010).
40. M. Moniruzzaman, L. L. Wurch, H. Alexander, S. T. Dyrhman, C. J. Gobler, S. W. Wilhelm, Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nat. Commun.* **8**, 16054 (2017).
41. S. M. Short, C. M. Short, Quantitative PCR reveals transient and persistent algal viruses in lake ontario canada. *Environ. Microbiol.* **11**, 2639–2648 (2009).
42. L. Gallot-Lavallée, G. Blanc, A glimpse of nucleocytoplasmic large DNA virus biodiversity through the eukaryotic genomics window. *Viruses* **9**, 17 (2017).
43. E. V. Koonin, T. G. Senkevich, V. V. Dolja, The ancient virus world and evolution of cells. *Biol. Direct* **1**, 29 (2006).
44. F. Sanchez, S. Geffroy, M. Norest, S. Yau, H. Moreau, N. Grimsley, Simplified transformation of *Ostreococcus tauri* using polyethylene glycol. *Genes (Basel)* **10**, 399 (2019).
45. S. Gnerre, I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, D. B. Jaffe, High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011).
46. N. Grimsley, B. Péquin, C. Bachy, H. Moreau, G. Piganeau, Cryptic sex in the smallest eukaryotic marine green alga. *Mol. Biol. Evol.* **27**, 47–54 (2010).
47. C. Clerissi, N. Grimsley, H. Ogata, P. Hingamp, J. Poulain, Y. Desdevises, Unveiling of the diversity of Prasinoviruses (*Phycodnaviridae*) in marine samples by using high-throughput sequencing analyses of PCR-amplified DNA polymerase and major capsid protein genes. *Appl. Environ. Microbiol.* **80**, 3150–3160 (2014).

**Acknowledgments:** We are grateful to the GenoToul bioinformatics platform for access to the computer cluster, the GeT sequencing platform for the PacBio sequencing, the Gordon and Betty Moore Foundation for sequencing transcriptomes, and the BIOPIIC platform for access to the cytometers and microscopes. We acknowledge A. Cadoual and F. Sanchez for performing an additional resistance evolution experiment of 24 independent lines of 53 inoculated with OmV2 to confirm systematic resistance acquisition in *O. mediterraneus*. We thank A. Penrose and Y. Gourbiere for help in the manual annotation of *O. mediterraneus* and visual inspection of TEM screens. We would like to apologize for the relevant references we could not include because of space limitations. **Funding:** This research was funded by the ANR grants REVIREC ANR-12-BSV7-0006-01, DECOVIR ANR-12-BSV7-0009, and PHYTNES ANR-13-JSV6-0005. E.V. is funded by BOF project GOA 01G01715, and L.F.B. is funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. H2020-MSCA-ITN-2015-675752. **Author contributions:** All molecular biology experiments on microalga and virus were performed at OOB (S.Y., M.K., L.F.B., S.S.-B., E.D., N.G., and H.M.). Illumina genome sequencing and assembly were performed at the Genoscope (J.G. and J.-M.A.). The genome annotation was completed at VIB (E.V. and S.R.) and corrected manually at OOB (M.K., S.Y., N.G., H.M., G.P., and E.D.). M.K. and S.Y. completed experimental evolution experiments. M.-L.E. performed TEM. Comparative genome analysis was performed by E.V. and K.V., with help from L.F.B. and Y.D. on OmV2 phylogeny. GC content evolution was performed by M.G. S.G. completed mathematical model conception and analysis. S.Y., S.G., and G.P. wrote the manuscript. All authors contributed to manuscript editing. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The genome sequence for *O. mediterraneus* RCC2590 and OmV2 can be found under GenBank accession numbers WMKK00000000 and MN688676, respectively. The complete gene annotations of RCC2590 can be accessed on ORCAE (<https://bioinformatics.psb.ugent.be/orcae/>). MMETSP sequence data can be accessed from [www.imicrobe.us/#/projects/104](http://www.imicrobe.us/#/projects/104). The genomes and their GenBank accession numbers used in this study are as follows: *B. prasinos* (GCA\_002220235.1), *M. commoda* RCC299 (GCA\_000090985.2), *M. pusilla* CCMP1545 (GCA\_000151265.1), *O. tauri* RCC4221 (CAID01000001.2 to CAID01000020.2, CR954200.2, and CR954199.2), *O. lucimarinus* CCE9901 (GCA\_000092065.1), BpV1 (HM004432.1), BpV2 (HM004430.1), MpV1 (HM004429.1), OtV1 (FN386611.1), OtV2 (FN600414.1), OtV5 (EU304328.2), OtV6 (JN225873.1), OIV1 (HM004431.1), OIV2 (KP874736.1), OIV3 (HQ633060.1), OIV4 (UF974316.1), OIV5 (HQ632827.1), OIV6 (HQ633059.1), OIV7 (KP874737.1), and OmV1 (KP874735.1). *Ostreococcus* sp. RCC809 is available from the JGI Genome Portal under project ID 16233. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 4 June 2019  
Accepted 9 January 2020  
Published 1 April 2020  
10.1126/sciadv.aay2587

**Citation:** S. Yau, M. Krasovec, L. F. Benites, S. Rombauts, M. Groussin, E. Vancaester, J.-M. Aury, E. Derelle, Y. Desdevises, M.-L. Escande, N. Grimsley, J. Guy, H. Moreau, S. Sanchez-Brosseau, Y. van de Peer, K. Vandepoele, S. Gourbiere, G. Piganeau, Virus-host coexistence in phytoplankton through the genomic lens. *Sci. Adv.* **6**, eaay2587 (2020).

## Virus-host coexistence in phytoplankton through the genomic lens

Sheree Yau, Marc Krasovec, L. Felipe Benites, Stephane Rombauts, Mathieu Groussin, Emmelien Vancaester, Jean-Marc Aury, Evelyne Derelle, Yves Desdevises, Marie-Line Escande, Nigel Grimsley, Julie Guy, Hervé Moreau, Sophie Sanchez-Brosseau, Yves van de Peer, Klaas Vandepoele, Sebastien Gourbiere and Gwenael Piganeau

*Sci Adv* 6 (14), eaay2587.  
DOI: 10.1126/sciadv.aay2587

|                         |  |
|-------------------------|--|
| ARTICLE TOOLS           | <a href="http://advances.sciencemag.org/content/6/14/eaay2587">http://advances.sciencemag.org/content/6/14/eaay2587</a>  |
| SUPPLEMENTARY MATERIALS | <a href="http://advances.sciencemag.org/content/suppl/2020/03/30/6.14.eaay2587.DC1">http://advances.sciencemag.org/content/suppl/2020/03/30/6.14.eaay2587.DC1</a>  |
| REFERENCES              | This article cites 46 articles, 11 of which you can access for free<br><a href="http://advances.sciencemag.org/content/6/14/eaay2587#BIBL">http://advances.sciencemag.org/content/6/14/eaay2587#BIBL</a> |
| PERMISSIONS             | <a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>  |

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).