Virus-host coexistence in phytoplankton through the genomic lens

Sheree Yau1,2,+, Marc Krasovec1, L. Felipe Benites1, Stephane Rombauts3,4, Mathieu Groussin5, Emmelien Vancaester3,4, Jean-Marc Aury6, Evelyne Derelle1,7, Yves Dessevives1, Marie-Line Escande1, Nigel Grimsley1, Julie Guy6, Hervé Moreau1, Sophie Sanchez-Brosseau1, Yves van de Peer1,8,9, Klaas Vandepoele1,9, Sebastien Goubriere6, Gwenaël Piganeau1*

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 ml^-1 in the photosynthetic zone (3), a large proportion of which are prasinoviruses, a monophyletic group known to infect Mamiellphyceae, 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-infesting 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 microagal 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 Mamiellphyceae 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.
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 × t + 7.559, R² = 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 theses 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.

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⁻¹, after which cultures recovered and cell concentrations reached that of uninfected controls (fig. S1). These cultures subsequently showed no visible lysis when rechallenged 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. 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 contain 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.
The S1 SOC with the table below listing their gene identifiers (prefixed by Ostme30g), significant identity to other SOC loci. Magenta blocks show unique genes deleted in the genomic map of the 58-kb deletion (blue outline). Gray blocks are genes with 

\[
\text{Blocks denoting sense and antisense matches, respectively. (C) Zoomed-in view of sequences with significant identity to the rest of the SOC, shown by red and blue }
\]

\[
\text{sequences, respectively. (C) Zoomed-in view of the genomic map of the 58-kb deletion (blue outline). Gray blocks are genes with significant identity to other SOC loci. Magenta blocks show unique genes deleted in the S1 SOC with the table below listing their gene identifiers (prefixed by Ostme30g), putative functional description, and gene length, bp, base pairs; RNase, ribonuclease.}
\]

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 ±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 new viral particles that survive in the environment at a rate \( v_S \). 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

\[
\lambda_{\text{free}} = \frac{1}{2} \left\{ (A + B) + \sqrt{(A - B)^2 + 4a_RA_R e_R e_S} \right\}
\]

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 \( e_S a_R / (\lambda_{\text{free}} - A) \), 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

\[
\lambda_{\text{inf}} \approx B
\]

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

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

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
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 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.

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
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 incongruencies 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,
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 in Mamiellales. 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 $e_R > e_S$. 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 pouchetti (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 Na2SiO3, 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-μm 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 × 107) in exponential growth phase were harvested by centrifugation (8000g for 20 min), resuspended in 150 μl of TE buffer [10 mM tris-HCl (pH 8), 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 protease K buffer [10 mM tris HCl (pH 8), 0.5 M EDTA (pH 8), 1% lauroyl sarcosinate, and proteinase K (1 mg 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× 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 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 [20x SSC, 3 M NaCl, and 0.3 M Na$_3$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 185 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 $\alpha^{32}$Pcytidine 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 μg ml$^{-1}$)] 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). 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).

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). 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 sequences as part of the MOMETSP. The resulting transcriptomes are referenced in table S2 with their respective MOMETSP 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 GenomeThresher (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/).

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/) 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$^{-5}$; 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.

Phylogenetic 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,
have the same properties than the original population of 2161 genes, were higher than 0.972, showing that these 100 jackknife subsamples were faithful representations of each genome in terms of GC contents. We further verified that these concatenates are 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 Dollo 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^{-3}; 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 node trees. 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 (±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 ~0.5 to ~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 microalga**

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°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 ml^{-1}. In addition, we passed the RCC2590 culture through a 0.2-µm 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 µl 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°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°C, diluting as necessary in filtered seawater (<0.02 µm), staining for 10 min at 80°C with SYBR Green I (Molecular Probes) (0.5 x 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°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. Where *S* algae encounters no virus or that a viral particle encounters no *S* algae 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/eaya2587/DC1

View request a protocol for this paper from Bio-protocol.

**REFERENCES AND NOTES**


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 BIDPIC platform for access to the cytometers and microscopes. We acknowledge A. Cadouald and F. Sanchez for performing an additional resistance evolution experiment of 24 independent lines of S3 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, DECOVR ANR-12-BSV7-0009, and PHYTHESS ANR-13-JS06-0003. E.V. is funded by the French National Research Agency under the French National Research Agency grant no. H2020-MSCA-ITN-2015-475752. Author contributions: All molecular biology experiments on microalgae and virus were performed at O.B.Y. (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 Genoscape (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.Y., N.G., H.M.G., and E.D.). M.K.Y. and S.Y. completed experimental evolution experiments. M.-L.E. performed TEM. Comparative genome analysis was performed by E.V. and C.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 WK000000000 and MN686676, 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. The genomes and their GenBank accession numbers used in this study are as follows: B. prasinum (GCA_00220235.1), M. commoda RCC299 (GCA_000990985.2), P. pusilla CCMP1545 (GCA_000151265.1), O. tauri RCC4221 (CAID01000001.2 to CAID01000020.2, CR954200.2, and CR954199.2), O. lucimarinus CCE9901 (GCA_000990985.1), BpV1 (HM004432.1), BpV2 (HM004431.1), BpV1 (HM004429.1), BpV1 (AF386611.1), BpV2 (AF600414.1), BpV5 (EU304328.2), BpV6 (JN225873.1), BpV1 (HM004431.1), BpV2 (KP874736.3), BpV3 (HQ633061.1), BpV4 (JF974316.1), BpV1 (HQ632827.1), BpV6 (HQ633059.1), BpV7 (KP874737.1), and OmV1 (KP874735.1). Osteococcus sp. RCCB09 is available from the JGl 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 article may be requested from the authors.

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

Virus-host coexistence in phytoplankton through the genomic lens

Sheree Yau, Marc Krasovec, L. Felipe Benites, Stephane Rombauts, Mathieu Groussin, Emmelien Vaes, 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