environmental microbiology reports

Environmental Microbiology Reports (2015) **7**(6), 979–989 doi:10.1111/1758-2229.12345

Deep sequencing of amplified *Prasinovirus* and host green algal genes from an Indian Ocean transect reveals interacting trophic dependencies and new genotypes

Camille Clerissi,[1,2] Yves Desdevises,[1,2] Sarah Romac,[3,4] Stéphane Audic,[3,4] Colomban de Vargas,[3,4] Silvia G Acinas,[5] Raffaella Casotti,[6] Julie Poulain,[7] Patrick Wincker,[7] Pascal Hingamp,[8] Hiroyuki Ogata[9] and Nigel Grimsley[1,2,*]

[1]*Observatoire Océanologique, Sorbonne Universités, UPMC Univ Paris 06, Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France.*

[2]*Biologie Intégrative des Organismes Marins, CNRS, UMR 7232, Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France.*

[3]*Sorbonne Universités, UPMC Univ Paris 06, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France.*

[4]*Equipe Evolution du Plancton et Paleo-Ocean, CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France.*

[5]*Department of Marine Biology and Oceanography, Institute of Marine Science (ICM), CSIC, Pg Marítim de la Barceloneta 37-49, Barcelona, Spain.*

[6]*Stazione Zoologica, Anton Dohrn, Villa Comunale, 80121, Naples, Italy*

[7]*CEA, Institut de Génomique, Génoscope, 2 Rue Gaston Crémieux, BP5706, Evry, 91057, France.*

[8]*CNRS, Université Aix-Marseille, Laboratoire Information Génomique et Structurale (UMR 7256), Mediterranean Institute of Microbiology (FR 3479), 13288, Marseille, France.*

[9]*Institute for Chemical Research, Kyoto University, Kyoto 611-0011, Japan.*

Summary

High-throughput sequencing of *Prasinovirus* DNA polymerase and host green algal (Mamiellophyceae) ribosomal RNA genes was used to analyse the diversity and distribution of these taxa over a ~10 000 km latitudinal section of the Indian Ocean. New viral and host groups were identified among the different trophic conditions observed, and highlighted that although unknown prasinoviruses are diverse, the cosmopolitan algal genera *Bathycoccus*, *Micromonas* and *Ostreococcus* represent a large proportion of the host diversity. While *Prasinovirus* communities were correlated to both the geography and the environment, host communities were not, perhaps because the genetic marker used lacked sufficient resolution. Nevertheless, analysis of single environmental variables showed that eutrophic conditions strongly influence the distributions of both hosts and viruses. Moreover, these communities were not correlated, in their composition or specific richness. These observations could result from antagonistic dynamics, such as that illustrated in a prey–predator model, and/or because hosts might be under a complex set of selective pressures. Both of these reasons must be considered to interpret environmental surveys of viruses and hosts, because covariation does not always imply interaction.

Received 27 January, 2015; revised 8 October, 2015; accepted 8 October, 2015. *For correspondence. E-mail nigel.grimsley@obs-banyuls.fr; Tel. +33468887396; Fax +33468887398.

Introduction

Microbes are the most abundant organisms in the sea, where they shape the structure and function of ecosystems (Azam *et al.*, 1983), but they are still one order of magnitude less abundant than microbe-infecting viruses (Suttle, 2005). Viruses are thus important players in microbial mortality and strongly influence biogeochemical cycles and the structure of host communities (Proctor and Fuhrman, 1990; Gustavsen *et al.*, 2014). Marine microbes and their associated viruses are thought to have high dispersal capacities because of their abundance, (Finlay, 2002; Angly *et al.*, 2006), although community composition might differ according to environmental conditions (Angly *et al.*, 2006; Martiny *et al.*, 2006).

However, little is known concerning the environmental factors that best explain their distribution and whether host and virus communities are correlated. To answer these questions, this study focuses on the genus *Prasinovirus*, a member of the *Phycodnaviridae* family (Wilson *et al.*, 2009) that infect an abundant and

widespread picoeukaryotic algal class referred to as the Mamiellophyceae (Marin and Melkonian, 2010). Known *Prasinovirus* host species include the three dominant genera *Bathycoccus*, *Micromonas* and *Ostreococcus*, infected respectively by *Bathycoccus* viruses (BpV), *Micromonas* viruses (MpV) and *Ostreococcus* viruses (OV). Several species have been described for each host genus (Marin and Melkonian, 2010; Piganeau *et al*., 2011b) that might be adapted to different environments. For example, *Ostreococcus* species might contain different ecotypes adapted to different light intensities (Rodriguez *et al*., 2005). Prasinoviruses are large, double-stranded DNA viruses and form a monophyletic group within the *Phycodnaviridae* family (Bellec *et al*., 2009). They are also abundant and widespread (Short and Short, 2008; Bellec *et al*., 2010; Park *et al*., 2011; Hingamp *et al*., 2013; Zhong and Jacquet, 2014). Previous studies suggested that both *Prasinovirus* and Mamiellophyceae have high dispersal capacities (Slapeta *et al*., 2006; Bellec *et al*., 2010) and that occurrence of genotypes is related to environmental conditions (Lepère *et al*., 2009; Bellec *et al*., 2010). However, culture-dependent methods were mainly used to study these groups so far, with no overview at the scale of communities.

The occidental part of the Indian Ocean was chosen for this analysis. This large region is affected disproportionally by global warming, because modelling and recent observations revealed a substantial temperature increase in the upper 700 m of the Indian Ocean (Lee *et al*., 2015), driving El Niño/Southern Oscillation cycles and climate change. Warm waters arriving on the Equatorial Currents from around Malaysia and Western Australia drive the warm Agulhas current southward along the East African coast, that in turn meets colder water from the South Atlantic and Benguela currents in an upwelling area. Thus this region provided contrasting conditions well suited for our objectives: (i) to describe the diversity of prasinoviruses and Mamiellophyceae at a community scale using a culture-independent sequencing approach, (ii) to disentangle the influence of the geographical and the environmental variables and (iii) to determine whether host and viral communities are correlated. We hypothesized that dispersal capacities of these communities are not limited within this oceanic region, but that compositions are highly constrained by the environment. Furthermore, *Prasinovirus* distribution might be strongly correlated to host communities, because their own replication depends on the cellular machinery.

Fig. 1. Locations of sampling sites. Numbers in station names are in chronological order. Seawater samples were collected on the schooner *Tara* at two depths: surface (SUR) and deep chlorophyll maximum (DCM). Free *Prasinovirus* particles and Mamiellophyceae were sampled using 0.1 and 0.8 μm filters respectively. ●, *Prasinovirus*; ○, Mamiellophyceae. Arrows indicate known water currents (adapted from Boebel *et al*., 1998).

Results and discussion

From oligotrophic to eutrophic samples

The 11 samples came from eight stations in the occidental part of the Indian Ocean (Fig. 1). Most samples were taken from the surface, but three came from the deep chlorophyll maximum; stations 58, 65 and 66. The sampling sites and the environmental variables are described in detail as supplementary information for methods. The first component of principal component analysis for the 11 samples (Fig. 2) divides them mainly according to potential temperature, oxygen and density (Table S1). Beam attenuation and backscattering coefficient of light by particles are both proxies of the particle load of seawater (e.g. Neukermans *et al*., 2012) and contributed to build the second component such as heterotrophic bacteria, which divide stations 36, 38, 39, 46, 66 from 57, 58, 65. This ordination highlighted high variability of environmental conditions, from oligotrophic (57, 58) to mesotrophic (36, 38, 39, 46, 65) and eutrophic (66). Stations 57 and 58 were located in the Mozambique channel, an oligotrophic area (Lévy *et al*., 2007; Leal *et al*., 2009), and contained low concentrations of particles and heterotrophic bacteria, which are more abundant in higher nutrient situations (e.g. Thingstad *et al*., 2008). In contrast, station 66 was particularly different from the other samples, probably because it was sampled within an area of high primary production (Villar *et al*., 2015) due to upwelling from the

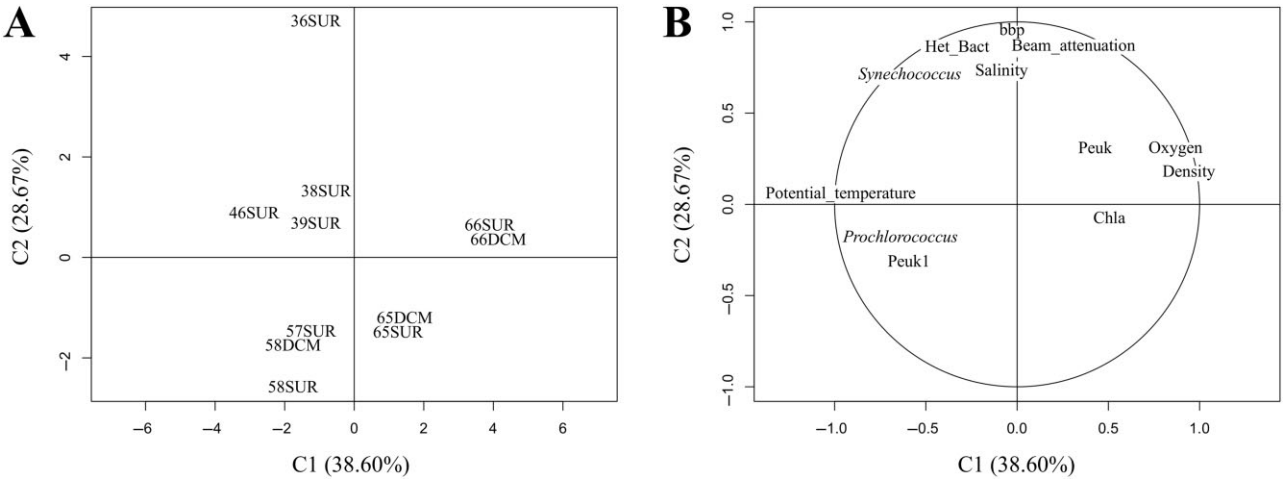

Fig. 2. Principal component analysis of the 11 samples according to the environmental variables.
A. Distances between samples.
B. Correlations between variables. Numbers in station names are in chronological order. SUR, surface; DCM, deep chlorophyll maximum. The following environmental variables were measured by the CTD (Conductivity/Temperature/Depth): salinity (g L^{-1}), potential temperature (°C; i.e. pressure-corrected temperature), density (kg/m^{-3}), oxygen (µmol kg^{-1}), chlorophyll *a* (Chla; mg Chl m^{-3}), backscattering coefficient of light by particles (bbp; 470 nm; m^{-1}) and beam attenuation (m^{-1}). Moreover, flow cytometry was used to estimate concentrations of *Prochlorococcus*, *Synechococcus*, heterotrophic bacteria (Het_Bact), picoeukaryotes (Peuk; ml^{-1}), the proportion of high-nucleic acid bacteria (HNA) and of small picoeukaryotes (Peuk1; putative Mamiellophyceae).

Benguela, South Atlantic and Agulhas currents (Fig. 1; Summerhayes *et al*., 1995; Boebel *et al*., 1998; Lutjeharms *et al*., 2000). Station 66 was characterized by motion of dense, cooler and nutrient-rich water towards the surface that increased the concentration of oxygen through enhanced photosynthetic activity. Notably, this station contained among the highest concentrations of chlorophyll *a* and photosynthetic picoeukaryotes (Table S2).

Uncultured prasinoviruses were very diverse

Although the *Prasinovirus* sequences are available for the 11 samples, the data for Mamiellophyceae concern six samples from four stations (Fig. 1). The sampling strategy is described in detail as supplementary information for methods, including the number of sequences, genotypes and Operational Taxonomic Units (Tables S3 and S4). To describe virus and host diversity of this oceanic region, phylogenetic reconstructions (Figs 3 and 4) and sequence annotations of viral DNA polymerase and host green algal RNA ribosomal (18S) genes were performed (see supplementary information for methods, Figs S1–S3, Table S5). Known host species of prasinoviruses are all species within dominant genera of the order Mamiellales (Bellec *et al*., 2009; Marin and Melkonian, 2010; reviewed in Grimsley *et al*., 2012). However, the culture-independent approach used here highlighted that although BpV and MpV were the richest groups, OV was only the seventh richest, and that unknown *Prasinovirus* contributed a high proportion of the diversity (OTU7, OTU11, OTU15 and OTU39; Fig. 3 and Fig. S2).

In contrast, the diversity of the Mamiellophyceae was consistent with previous studies; *Bathycoccus*, *Micromonas* and *Ostreococcus* were the most abundant (Fig. S3) (Not *et al*., 2004; Viprey *et al*., 2008). Notably, *Bathycoccus* and *Ostreococcus* were found in higher proportions in this region, whereas *Micromonas* dominated the eukaryotic picoplankton in the English Channel (Not *et al*., 2004) and at a Mediterranean Sea coastal site (Zhu *et al*., 2005). This composition was nevertheless realistic, because the genus *Ostreococcus* can dominate picoeukaryote communities: it is known to form blooms (O'Kelly *et al*., 2003; Treusch *et al*., 2012) and can contribute to up to 70% of the phototrophic picoeukaryotes (Countway and Caron, 2006). Moreover, phylogenetic reconstruction of Mamiellophyceae sequences also highlighted a new environmental clade related to *Crustomastix* and *Dolichomastix* [Fig. 4 box with dashed lines (OTUs were defined for a nucleotide identity of 95% instead of 97% to produce a clearer tree) and Table S6]. Remarkably, a few related sequences were found in samples from a deep-sea methane cold seep (Takishita *et al*., 2007), the sediment of a hydrothermal vent (Edgcomb *et al*., 2011), and in gut content of a bivalve (Duplessis *et al*., 2004).

Most unknown prasinoviruses might infect Dolichomastigales

Only representatives of BpV, MpV and OV are so far available in culture (Cottrell and Suttle, 1995; Derelle

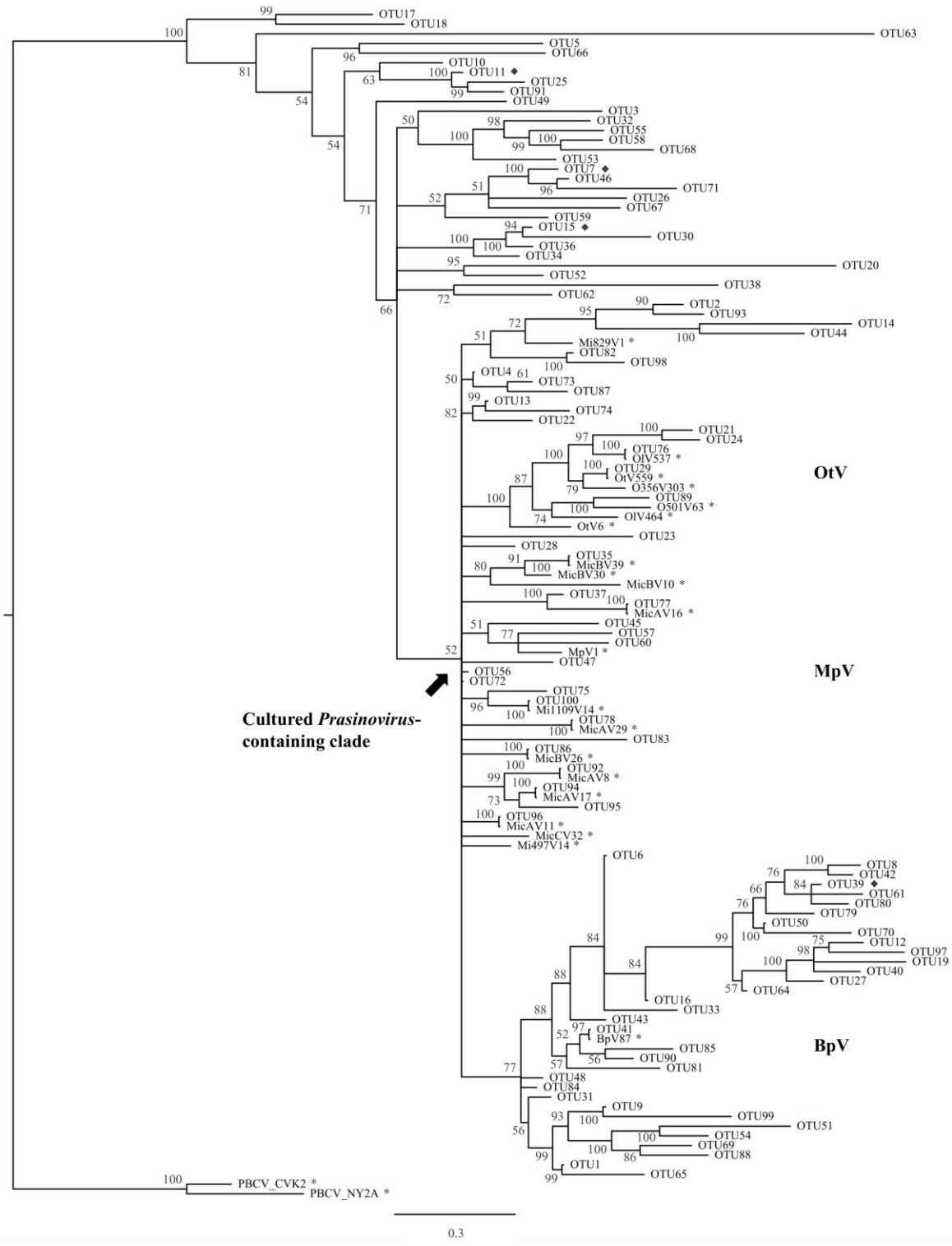

Fig. 3. Phylogenetic tree of environmental OTUs and 23 reference sequences of *Prasinovirus* and *Chlorovirus*, reconstructed using Bayesian inference. PCR amplifications, sequencing and sequence cleaning were performed such as described in Clerissi *et al.* (2014a). OTUs are defined for a nucleotide identity of 90%. Phylogenetic reconstructions were based on DNA sequences that were partitioned according to codon position, and the estimation of model parameters was unlinked across partitions. Bayesian analysis was carried out with MrBayes 3.2 (Ronquist *et al.*, 2012), with four chains of 2 000 000 generations, trees sampled every 1000 generations and burnin value set to 20% of the sampled trees. The tree was rooted using the chloroviruses. Numbers are posterior probabilities (%) reflecting clade support. Twenty-three reference sequences representing 475 *Prasinovirus* and *Chlorovirus* isolates for an OTU cutoff of 90% are indicated by an asterisk. Four abundant but unknown OTUs are indicated by a lozenge. The cultured *Prasinovirus*-containing clade is indicated by an arrow.

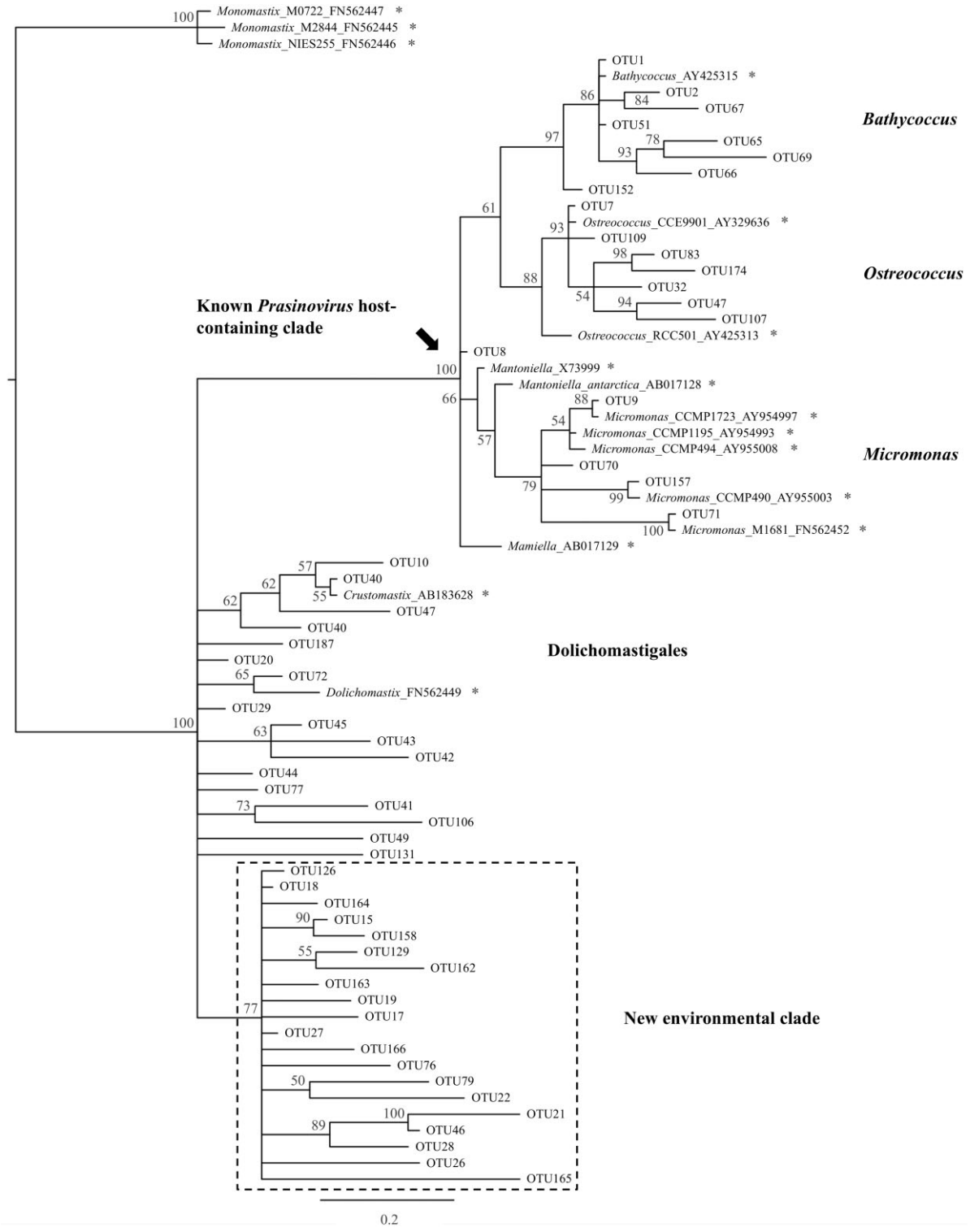

Fig. 4. Phylogenetic tree of environmental OTUs and 16 reference sequences of Mamiellophyceae, reconstructed using Bayesian inference. PCR amplifications of V9 region of the 18S were conducted using the PCR primers 1389f (5′-TTG TAC ACA CCG CCC-3′) and 1510r (5′-CCT TCY GCA GGT TCA CCT AC-3′). Amplicons were sequenced using Illumina, sequences were cleaned and chimeras were removed using USEARCH (Edgar, 2010). Phylogenetic reconstructions were based on DNA sequences, with an evolutionary model selected via Akaike information criterion and jModelTest v2 (Darriba *et al.*, 2012). Bayesian analysis was carried out with MrBayes similarly to *Prasinovirus*. The tree was rooted using *Monomastix* strains. Numbers are posterior probabilities (%) reflecting clade support. Sixteen reference sequences representing Mamiellophyceae diversity (Marin and Melkonian, 2010) for an OTU cutoff of 97% are indicated by an asterisk. The known *Prasinovirus* host-containing clade is indicated by an arrow and a new environmental clade is outlined in a box with dashed lines.

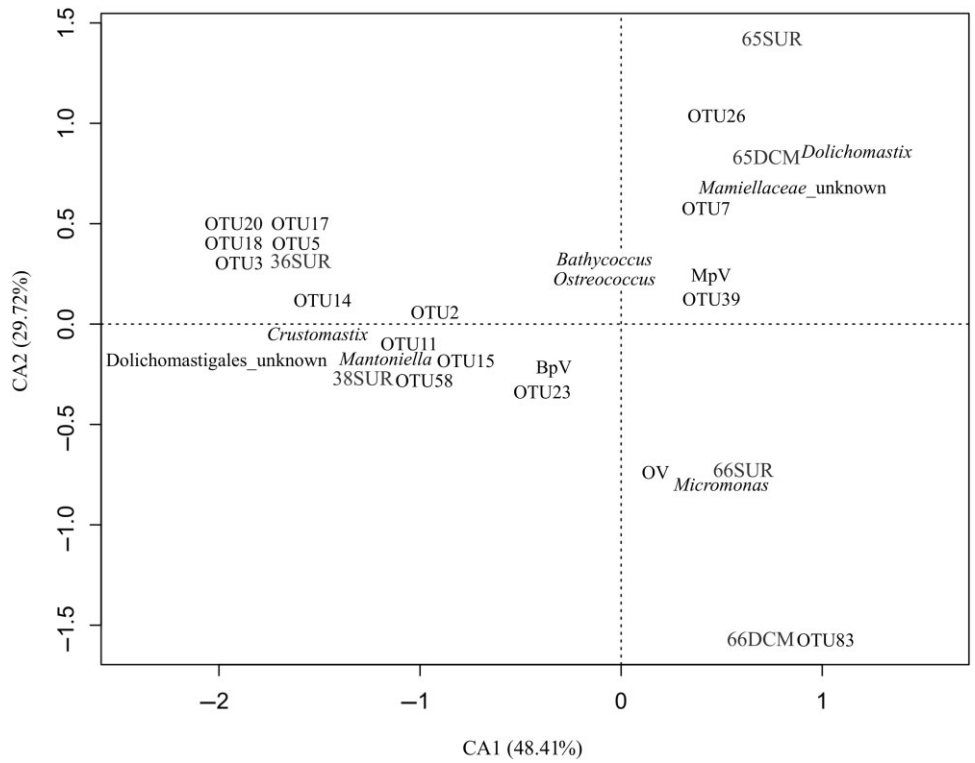

Fig. 5. Correspondence analysis of the relative abundance matrix for *Prasinovirus* and Mamiellophyceae. Clustering analyses with reference sequences were computed to annotate *Prasinovirus* OTUs and Mamiellophyceae genotypes at the genus level.

et al., 2008; 2015; Bellec *et al*., 2009; Weynberg *et al*., 2009; 2011). This lack of virus cultures for other genera might be biased, because mostly coastal areas were sampled using cultures of coastal algal strains, whereas *Mamiella*, *Crustomastix* and *Dolichomastix* were more commonly represented in oligotrophic waters (Viprey *et al*., 2008). Because unknown *Prasinovirus* genotypes were very rich in our dataset (particularly OTU7, OTU11, OTU15 and OTU39; Fig. S2), the prediction of host identities was carried out.

First, a canonical correspondence analysis (CCA) highlighted that two Mamiellophyceae OTUs were correlated to the distribution of *Prasinovirus*: OTU28 and OTU126 (*P*-value = 0.005). These two OTUs belong to the robust clade described above using the phylogenetic analysis (Fig. 4). A BLASTn search against the National Center for Biotechnology Information (NCBI) nucleotide collection suggested that they are most similar to *Crustomastix stigmatica* (Table S7), and these sequences came mostly from stations 36 and 38 where they represent ~14% of genotypes compared with an average of 2% in other samples.

Secondly, because *Prasinovirus* are mainly genus specific (Clerissi *et al*., 2012; Bellec *et al*., 2014), a co-distribution analysis was computed using genus annotation for Mamiellophyceae and the *Prasinovirus* annotation (Fig. 5, Fig. S2, Table S5). While *Ostreococcus* and *Bathycoccus* displayed a homogeneous distribution within the six samples, the correspondence analysis shows similar distributions for (i) *Micromonas* and OV in station 66, (ii) OTU7, OTU26, Mamiellaceae_unknown and *Dolichomastix* in station 65 and (iii) OTU11, OTU14, OTU15, OTU58, *Crustomastix*, *Mantoniella*_unknown and Dolichomastigales_unknown in stations 36 and 38. However, only the link between Dolichomastigales_unknown and OTU11 was significant (r = 0.99; *P*-value = 0.01). Thus both analyses suggested that uncultured *Prasinovirus* groups possibly infected Mamiellophyceae strains from the Dolichomastigales order.

The distribution of communities is influenced mainly by trophic conditions

Given the results of previous studies (Slapeta *et al*., 2006; Lepère *et al*., 2009; Bellec *et al*., 2010; Clerissi *et al*., 2014b), links with environmental conditions were expected, but not with geographical distances (locations) for both communities in this oceanic region.

First, *Prasinovirus* were correlated to both locations (Mantel test, r = 0.722, *P*-value = 0.001) and environment (Mantel test, r = 0.626, *P*-value = 0.001) (see supplementary information for methods, with details about the statis-

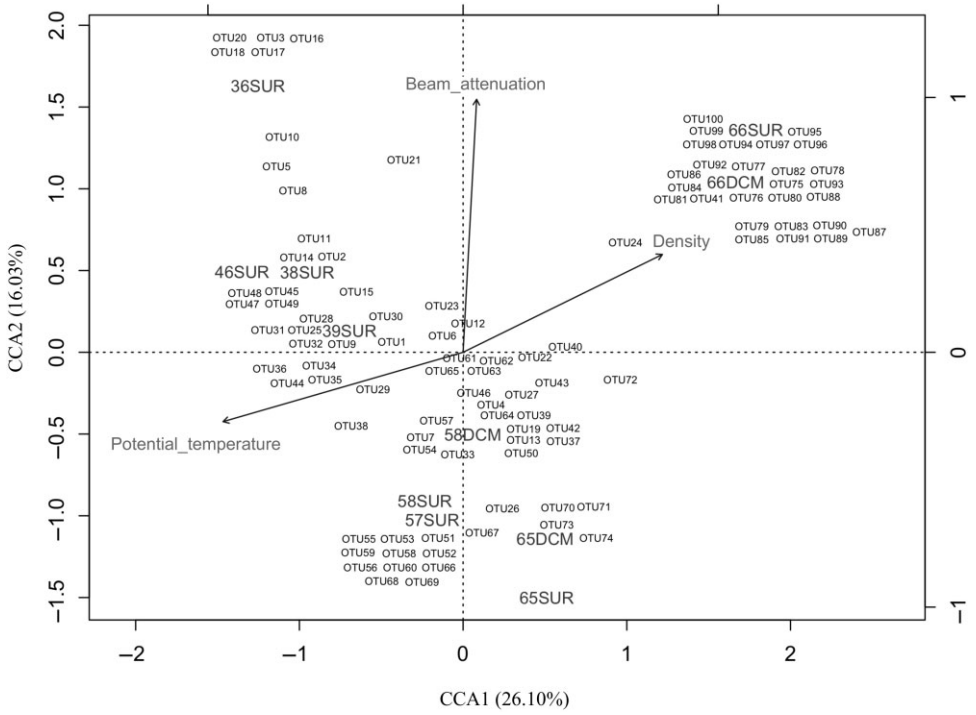

Fig. 6. Canonical correspondence analysis of the 11 samples on *Prasinovirus* assemblages constrained by environmental data. Numbers in station names are in chronological order. SUR, surface; DCM, deep chlorophyll maximum. OTUs are defined for a nucleotide identity of 90%. Only the significant variables are shown (i.e. variables that significantly explained changes in the distribution of OTU). They were selected using a forward-selection procedure associated to the canonical correspondence analysis.

tical and multivariate procedures). This spatial structure was surprising, because no links were observed between genetic distances of *Ostreococcus lucimarinus* viruses and sampling locations at a global scale (Bellec *et al*., 2010; Derelle *et al*., 2015). However, locations and environment were also correlated in our dataset (Mantel test, r = 0.521, *P*-value = 0.001), and no differences were found between the genotypic structures of *Prasinovirus* communities in the 11 samples (P-test, *P*-value = 1). These observations might indicate a key role of the environment, and that *Prasinovirus* were actually dispersed in the occidental part of the Indian Ocean.

Secondly, significant links for the Mamiellophyceae communities were not found using Mantel tests (location: r = 0.275, *P*-value = 0.141; environment: r = 0.342, *P*-value = 0.092). This lack of correlations could be the result of a low statistical power, because the dataset contains six samples, but such correlations were still significant for *Prasinovirus* communities when using the same reduced dataset (location: r = 0.852, *P*-value = 0.003; environment: r = 0.771, *P*-value = 0.004). Hence, Mamiellophyceae might be highly dispersed and homogeneously distributed in this region.

However, to further decipher the influence of environmental variables on both communities, CCAs were computed with a forward-selection procedure. This analysis highlighted (i) that potential temperature, density and beam attenuation constrained *Prasinovirus* distribution in the 11 samples (*P*-value = 0.005) (Fig. 6; a similar trend was observed for the reduced dataset of six samples, Fig. S4) and (ii) that potential temperature influenced Mamiellophyceae in the six samples (*P*-value = 0.015). Because potential temperature and density tend to separate station 66 from the other samples for both analyses, the eutrophic conditions of station 66 seem to highly constrain communities of this host–virus system.

Few links between Prasinovirus *and Mamiellophyceae communities*

Because *Prasinovirus* entirely depend on hosts for their replication, a strong correlation between both communities was expected, but links were significant neither for community compositions (r = 0.397, *P*-value = 0.172) (Table 1) nor for specific richness (Spearman correlation, ρ = 0.6, *P*-value = 0.242). This lack of correlation can be explained by at least three hypotheses: (i) a poor resolution of membership content of both viral and host communities according to different unknown biases (DNA extraction, polymerase chain reaction, sequencing), (ii) a non-corresponding taxonomic threshold between viruses and hosts and (iii) antagonistic oscillations between hosts and viruses.

Table 1. Mantel test correlations.

	Eleven samples		Six samples		
	Environment	Location	Mamiellophyceae	Environment	Location
Prasinovirus	**0.626**	**0.722**	0.397	**0.771**	**0.852**
Mamiellophyceae	N.A.	N.A.	–	0.342	0.275
Environment	–	**0.521**	0.342	–	**0.775**
Location	**0.521**	–	0.275	**0.775**	–

Prasinovirus and Mamiellophyceae OTUs are defined for a nucleotide identity of 90 and 97% respectively. N.A., not available. Numbers indicate correlation coefficients and significant correlations (P-value < 0.05) are in bold. The distance matrices were computed using the Bray–Curtis dissimilarity for virus and host communities and the Euclidean metric for the environmental variables and the geographic coordinates after a standardization step.

A non-corresponding taxonomic threshold might result from an overestimation of *Prasinovirus* diversity and/or an underestimation of host diversity. On one hand, because the environmental diversity of prasinoviruses was not known, their phylogenetic limit was defined arbitrarily by the *Chlorovirus* sister clade (see supplementary information for methods). In addition, it is possible that the thresholds used to define virus and host OTUs did not correspond to the taxonomic interaction and that not all were able to infect Mamiellophyceae. On the other hand, some evidence suggests that host diversity is underestimated when using the 18S as genetic marker (Piganeau *et al*., 2011a), especially because strains with identical sequences display different susceptibilities to prasinoviruses (Clerissi *et al*., 2012).

Antagonistic oscillations between hosts and viruses are also a plausible source of noise for correlation analyses. Indeed, viruses might shape the structure of host communities via the top-down elimination of different members (Thingstad and Lignell, 1997; Winter *et al*., 2010). They can terminate blooms of hosts and be present when hosts are not (Bratbak *et al*., 1993; Schroeder *et al*., 2003). As a consequence, an increasing abundance of viral genotypes is expected to be associated with a decrease of their specific hosts. However links are not necessarily linear and can be complex because host ranges vary widely, for example (Winter *et al*., 2010). Because free viral particles were sampled independently of host cells (fraction below 0.2 μm for viruses), it is tempting to speculate that the antagonistic dynamics observed is a likely hypothesis to explain the lack of correlations between *Prasinovirus* and Mamiellophyceae communities in this study. In particular, OV were mainly found in station 66 with *Micromonas* (Fig. 5). Their occurrence suggests a bloom of the genus *Ostreococcus* before an algal succession dominated by *Micromonas*.

Lastly, while viruses mainly depend on the presence of hosts and on factors involved in their decay, hosts must face not only bottom-up (nutrients) and top-down factors (viruses and grazers such as ciliates and flagellates), but also sideways controls such as competition for nutrients against other algae and heterotrophic bacteria (e.g. Thingstad *et al*., 2008). Thus, host occurrence depends on a complex set of selective pressures, and this might explain absence of correlations for Mamiellophyceae communities with viruses and environments in this study.

To conclude, *Prasinovirus* and Mamiellophyceae communities were compared in the west part of the Indian Ocean, and the results suggest that trophic conditions influenced their distribution. Until now, known *Prasinovirus* were characterized mainly in samples from eutrophic waters, but here we showed that related communities also occur in nutrient-limited waters and that unknown genotypes possibly infect Dolichomastigales.

In addition, geographic barriers seemed inexistent for viruses and hosts in this region, and taxa represented in each sample probably arose from growth of adapted genotypes before further dispersal. Our analysis also highlighted that host–virus interactions in natural environments can be difficult to study because these partners may follow complex antagonistic dynamics. Hence, future projects should focus on temporal analyses of specific sites or use a unique sampling strategy that describes both viruses and hosts (e.g. cell sorting using flow cytometry or sampling through 0.8 μm filters).

Finally, the link between *Prasinovirus* communities and the environment suggested the presence of different propagation strategies, such as described for OtV2, a virus that infects the low-light adapted *Ostreococcus tauri* strain and that contains specific genes certainly acquired laterally (Weynberg *et al*., 2011). This observation leads to exciting new questions from an evolutionary point of view: do *Prasinovirus* genomes contain adaptive genes to promote infections of their hosts in different trophic conditions? If so, are they acquired by lateral transfers from hosts or other viruses during coinfection events?

Acknowledgements

We thank the Genophy team in Banyuls-sur-Mer, especially Sheree Yau, for stimulating discussions. We are also grateful to Martha Clokie, Ramon Massana, Marcelino Suzuki, Sonja

Fagervold, Joseph M Gasol, Christopher Quince and Michel Krawczyk for useful comments and advice. We particularly want to thank Marc Picheral, Sarah Searson, Grigor Obolensky, Gaby Gorsky, Lars Stemmann, Vincent Taillandier, François Roullier, Josephine Ras, Léo Berline, Céline Dimier, Annick Bricaud and Emmanuel Boss for environmental variables. We thank the Wikimedia commons for image 'BlankMap-World-162E-flat' that we modified to use as a background for Fig. 1 (http://commons.wikimedia.org/wiki/File:BlankMap-World-162E-flat.svg#mediaviewer/File:BlankMap-World-162E-flat.svg). This work was supported by an 'Agence Nationale de Recherche' grant 'TARA-GIRUS' ANR-09-PCS-GENM-218 (coordinator H. Ogata). It was partially supported by JSPS KAKENHI (Grant Number 26430184) to HO and OCEANOMICS (ANR-11-BTBR-0008). Camille Clerissi benefited from a doctoral fellowship from the AXA Research Fund. We are keen to thank commitment of the following people and sponsors who made the *Tara Oceans* expedition possible: CNRS, EMBL, Genoscope/CEA, VIB, Stazione Zoologica Anton Dohrn, UNIMIB, ANR (projects POSEIDON, BIOMARKS, PROMETHEUS), FWO, BIO5, Biosphere 2, Agnès b., the Veolia Environment Foundation, Region Bretagne, World Courier, Illumina, Cap L'Orient, the EDF Foundation EDF Diversiterre, FRB, the Prince Albert II de Monaco Foundation, the Italian Research for the Sea (Flagship Project RITMARE) to R.C., Etienne Bourgois, the *Tara* schooner and its captain and crew. *Tara Oceans* would not exist without the continuous support of the participating 23 institutes (see http://oceans.taraexpeditions.org). This is contribution number 0032 of the *Tara Oceans* Expedition 2009–2012.

Data deposition footnote

The sequence datasets have been submitted to the Sequence Read Archive of the European Nucleotide Archive under the following accession numbers 36SUR (ERR632179; ERR562665), 38SUR (ERR632184; ERR562391), 39SUR (ERR632191), 46SUR (ERR632186), 57SUR (ERR632175), 58DCM (ERR632185), 58SUR (ERR632181), 65DCM (ERR632174; ERR562488), 65SUR (ERR632195; ERR562667), 66DCM (ERR632194; ERR562660), 66SUR (ERR632169; ERR562457).

References

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., *et al*. (2006) The marine viromes of four oceanic regions. *PLoS Biol* **4:** e368.

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., and Thingstad, F. (1983) The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10:** 257–263.

Bellec, L., Grimsley, N., Moreau, H., and Desdevises, Y. (2009) Phylogenetic analysis of new Prasinoviruses (*Phycodnaviridae*) that infect the green unicellular algae *Ostreococcus*, *Bathycoccus* and *Micromonas*. *Environ Microbiol Rep* **1:** 114–123.

Bellec, L., Grimsley, N., and Desdevises, Y. (2010) Isolation of prasinoviruses of the green unicellular algae *Ostreococcus* spp. on a worldwide geographical scale. *Appl Environ Microbiol* **76:** 96–101.

Bellec, L., Grimsley, N., Derelle, E., Moreau, H., and Desdevises, Y. (2010) Abundance, spatial distribution and genetic diversity of *Ostreococcus tauri* viruses in two different environments. *Environ Microbiol Rep* **2:** 313–321.

Bellec, L., Clerissi, C., Edern, R., Foulon, E., Simon, N., Grimsley, N., and Desdevises, Y. (2014) Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. *BMC Evol Biol* **14:** 59.

Boebel, O., Duncombe Rae, C., Garzoli, S., Lutjeharms, J., Richardson, P., Rossby, T., *et al*. (1998) Float experiment studies interocean exchanges at the tip of Africa. *EOS Trans Am Geophys Union* **79:** 1–8.

Bratbak, G., Egge, J.K., and Heldal, M. (1993) Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser* **93:** 39–48.

Clerissi, C., Desdevises, Y., and Grimsley, N. (2012) Prasinoviruses of the marine green alga *Ostreococcus tauri* are mainly species-specific. *J Virol* **86:** 4611–4619.

Clerissi, C., Grimsley, N., Ogata, H., Hingamp, P., Poulain, J., and Desdevises, Y. (2014a) 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.

Clerissi, C., Grimsley, N., Subirana, L., Maria, E., Oriol, L., Ogata, H., *et al*. (2014b) *Prasinovirus* distribution in the Northwest Mediterranean Sea is affected by the environment and particularly by phosphate availability. *Virology* **466–467:** 146–157. http://dx.doi.org/10.1016/j.virol.2014.07.016i.

Cottrell, M.T., and Suttle, C.A. (1995) Genetic diversity of algal viruses which lyse the photosynthetic picoflagellate *Micromonas pusilla* (Prasinophyceae). *Appl Environ Microbiol* **61:** 3088–3091.

Countway, P.D., and Caron, D.A. (2006) Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microbiol* **72:** 2496–2506.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9:** 772.

Derelle, E., Ferraz, C., Escande, M.-L., Eychenié, S., Cooke, R., Piganeau, G., *et al*. (2008) Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLoS ONE* **3:** e2250.

Derelle, E., Monier, A., Cooke, R., Worden, A.Z., Grimsley, N.H., and Moreau, H. (2015) Diversity of viruses infecting the green microalga *Ostreococcus lucimarinus*. *J Virol* **89:** 5812–5821.

Duplessis, M.R., Dufour, S.C., Blankenship, L.E., Felbeck, H., and Yayanos, A.A. (2004) Anatomical and experimental evidence for particulate feeding in *Lucinoma aequizonata* and *Parvilucina tenuisculpta* (Bivalvia: Lucinidae) from the Santa Barbara basin. *Mar Biol* **145:** 551–561.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460–2461.

Edgcomb, V.P., Beaudoin, D., Gast, R., Biddle, J.F., and Teske, A. (2011) Marine subsurface eukaryotes: the fungal majority. *Environ Microbiol* **13:** 172–183.

Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296:** 1061–1063.

Grimsley, N.H., Thomas, R., Kegel, J.U., Jacquet, S., Moreau, H., and Desdevises, Y. (2012) Genomics of algal host-virus interactions. *Adv Bot Res* **64:** 343–381.

Gustavsen, J.A., Winget, D.M., Tian, X., and Suttle, C.A. (2014) High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities. *Front Microbiol* **5:** 703.

Hingamp, P., Grimsley, N., Acinas, S.G., Clerissi, C., Subirana, L., Poulain, J., *et al*. (2013) Exploring nucleocytoplasmic large DNA viruses in *Tara Oceans* microbial metagenomes. *ISME J* **7:** 1678–1695.

Lévy, M., Shankar, D., André, J.-M., Shenoi, S.S.C., Durand, F., and de Boyer Montégut, C. (2007) Basin-wide seasonal evolution of the Indian Ocean's phytoplankton blooms. *J Geophys Res* **112:** C12014.

Leal, M.C., Sá, C., Nordez, S., Brotas, V., and Paula, J. (2009) Distribution and vertical dynamics of planktonic communities at Sofala Bank, Mozambique. *Estuar Coast Shelf Sci* **84:** 605–616.

Lee, S.-K., Park, W., Baringer, M.O., Gordon, A.L., Huber, B., and Liu, Y. (2015) Pacific origin of the abrupt increase in Indian Ocean heat content during the warming hiatus. *Nat Geosci* **8:** 445–449.

Lepère, C., Vaulot, D., and Scanlan, D.J. (2009) Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. *Environ Microbiol* **11:** 3105–3117.

Lutjeharms, J.R.E., Cooper, J., and Roberts, M. (2000) Upwelling at the inshore edge of the Agulhas Current. *Cont Shelf Res* **20:** 737–761.

Marin, B., and Melkonian, M. (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear-and plastid-encoded rRNA operons. *Protist* **161:** 304–336.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Neukermans, G., Loisel, H., Mériaux, X., Astoreca, R., and McKee, D. (2012) *In situ* variability of mass-specific beam attenuation and backscattering of marine particles with respect to particle size, density, and composition. *Limnol Oceanogr* **57:** 124–144.

Not, F., Latasa, M., Marie, D., Cariou, T., Vaulot, D., and Simon, N. (2004) A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English Channel. *Appl Environ Microbiol* **70:** 4064–4072.

O'Kelly, C.J., Sieracki, M.E., Thier, E.C., and Hobson, I.C. (2003) A transient bloom of *Ostreococcus* (Chlorophyta, Prasinophyceae) in West Neck Bay, Long Island, New York. *J Phycol* **39:** 850–854.

Park, Y., Lee, K., Lee, Y.S., Kim, S.W., and Choi, T.-J. (2011) Detection of diverse marine algal viruses in the South Sea regions of Korea by PCR amplification of the DNA polymerase and major capsid protein genes. *Virus Res* **159:** 43–50.

Piganeau, G., Eyre-Walker, A., Grimsley, N., and Moreau, H. (2011a) How and why DNA barcodes underestimate the diversity of microbial eukaryotes. *PLoS ONE* **6:** e16342.

Piganeau, G., Grimsley, N., and Moreau, H. (2011b) Genome diversity in the smallest marine photosynthetic eukaryotes. *Res Microbiol* **162:** 570–577.

Proctor, L.M., and Fuhrman, J.A. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature* **343:** 60–62.

Rodriguez, F., Derelle, E., Guillou, L., Gall, F.L., Vaulot, D., and Moreau, H. (2005) Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol* **7:** 853–859.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., *et al*. (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61:** 539–542.

Schroeder, D.C., Oke, J., Hall, M., Malin, G., and Wilson, W.H. (2003) Virus succession observed during an *Emiliania huxleyi* bloom. *Appl Environ Microbiol* **69:** 2484–2490.

Short, S., and Short, C. (2008) Diversity of algal viruses in various North American freshwater environments. *Aquat Microb Ecol* **51:** 13–21.

Slapeta, J., Lopez-Garcia, P., and Moreira, D. (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol* **23:** 23–29.

Summerhayes, C.P., Kroon, D., Rosell-Melé, A., Jordan, R.W., Schrader, H.-J., Hearn, R., *et al*. (1995) Variability in the Benguela Current upwelling system over the past 70,000 years. *Prog Oceanogr* **35:** 207–251.

Suttle, C.A. (2005) Viruses in the sea. *Nature* **437:** 356–361.

Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y., and Maruyama, T. (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extremophiles* **11:** 563–576.

Thingstad, T.F., and Lignell, R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol* **13:** 19–27.

Thingstad, T.F., Bellerby, R.G.J., Bratbak, G., Børsheim, K.Y., Egge, J.K., Heldal, M., *et al*. (2008) Counterintuitive carbon-to-nutrient coupling in an Arctic pelagic ecosystem. *Nature* **455:** 387–390.

Treusch, A.H., Demir-Hilton, E., Vergin, K.L., Worden, A.Z., Carlson, C.A., Donatz, M.G., *et al*. (2012) Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. *ISME J* **6:** 481–492.

Villar, E., Farrant, G.K., Follows, M., Garczarek, L., Speich, S., Audic, S., *et al*. (2015) Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348:** 1261447.

Viprey, M., Guillou, L., Ferréol, M., and Vaulot, D. (2008) Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10:** 1804–1822.

Weynberg, K.D., Allen, M.J., Ashelford, K., Scanlan, D.J., and Wilson, W.H. (2009) From small hosts come big viruses:

the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ Microbiol* **11:** 2821–2839.
Weynberg, K.D., Allen, M.J., Gilg, I.C., Scanlan, D.J., and Wilson, W.H. (2011) Genome sequence of *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. *J Virol* **85:** 4520–4529.
Wilson, W.H., Etten, J.L., and Allen, M.J. (2009) The *Phycodnaviridae*: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol* **328:** 1–42.
Winter, C., Bouvier, T., Weinbauer, M.G., and Thingstad, T.F. (2010) Trade-offs between competition and defense specialists among unicellular planktonic organisms: the 'Killing the Winner' hypothesis revisited. *Microbiol Mol Biol Rev* **74:** 42–57.
Zhong, X., and Jacquet, S. (2014) Contrasting diversity of phycodnavirus signature genes in two large and deep western European lakes. *Environ Microbiol* **16:** 759–773.
Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52:** 79–92.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Mamiellophyceae annotation. The clustering corresponds to the presence/absence of reference sequences in OTUs for different nucleotide identities (Jaccard index; unweighted pair group method with arithmetic mean). The name of each strain is followed by its accession number.
Fig. S2. Rank abundance of the *Prasinovirus* genotypes in the 11 samples. OTUs are defined for a cutoff of 74%.
Fig. S3. Annotation and rank abundance of the Mamiellophyceae genotypes in six samples. A–E correspond to the clades defined in Marin and Melkonian (2010). Only sequence-containing taxa are shown.
Fig. S4. Canonical correspondence analysis of the six samples on *Prasinovirus* assemblages constrained by environmental data. Numbers in station names are in chronological order. SUR, surface; DCM, deep chlorophyll maximum. OTUs are defined for a nucleotide identity of 90%. Only the significant variables are shown.
Fig. S5. Canonical correspondence analysis of the six samples on Mamiellophyceae assemblages constrained by environmental data. Numbers in station names are in chronological order. SUR, surface; DCM, deep chlorophyll maximum. OTUs are defined for a nucleotide identity of 97%. Only the significant variable is shown.
Table S1. Contributions of the environmental variables used to build the first and second components of the polymerase chain reaction (PCA).
Table S2. Geography and environmental variables of the 11 samples.
Table S3. *Prasinovirus* sequencing data.
Table S4. Mamiellophyceae sequencing data.
Table S5. Annotation of *Prasinovirus* sequences for a nucleotide identity of 74%. OTU representative sequences were compared with the NCBI database via BLASTn searches.
Table S6. Annotation of Mamiellophyceae sequences belonging to the robust but unknown clade. OTU representative sequences were compared with the NCBI database via BLASTn searches.
Table S7. Annotation of Mamiellophyceae OTUs constraining *Prasinovirus* distribution using CCA.
Appendix S1. Experimental procedures.