

Viral dispersal across the air-sea interface

Heads in the clouds: marine viruses disperse bidirectionally along the natural water cycle

Janina Rahlff^{1,2*}, Sarah P. Esser¹, Julia Plewka¹, Mara Elena Heinrichs³, André Soares¹,
Claudio Scarchilli⁴, Paolo Grigioni⁴, Heike Wex⁵, Helge-Ansgar Giebel^{3,#},
Alexander J. Probst^{1,†}

¹Group for Aquatic Microbial Ecology, University of Duisburg-Essen, Department of Chemistry, Environmental Microbiology and Biotechnology (EMB), Universitätsstraße 5, 45141 Essen, Germany

²Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Department of Biology and Environmental Science, Linnaeus University, SE-39231 Kalmar, Sweden

³Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University of Oldenburg, Carl-von-Ossietzky-Straße 9-11, 26129 Oldenburg, Germany

⁴Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), via Anguillarese 301, 00123 Rome, Italy

⁵Department of Experimental Aerosol and Cloud Microphysics, Leibniz Institute for Tropospheric Research (TROPOS), 04318 Leipzig, Germany

#

Present address: Institute for Chemistry and Biology of the Marine Environment (ICBM), Center for Marine Sensors (ZfMarS), Carl von Ossietzky University of Oldenburg, Schleusenstraße 1, 26382 Wilhelmshaven, Germany

† Present addresses: Centre of Water and Environmental Research (ZWU), University of Duisburg-Essen, Universitätsstraße 5, Essen, Germany
Research Center One Health Ruhr, University Alliance Ruhr,
Environmental Metagenomics, University of Duisburg-Essen, Germany

*Corresponding author: Janina.rahlff@uol.de

Abstract

Marine viruses have thoroughly been studied in seawater, yet their dispersal from neuston ecosystems at the air-sea interface towards the atmosphere remains a knowledge gap. Here, we show that 6.2 % of the studied virus population were shared between air-sea interface ecosystems and rainwater. Virus enrichment in the 1-mm thin surface microlayer and sea foams happened selectively, and variant analysis proved virus transfer to aerosols and rain. Viruses detected in rain and aerosols showed a significantly higher percent G/C base content compared to marine viruses, and a genetically distinct rain virome supports that those viruses could inhabit higher air masses. CRISPR spacer matches of marine prokaryotes to “foreign” viruses from rainwater prove regular virus-host encounters at the air-sea interface. Our findings on aerosolization and long-range atmospheric dispersal implicate virus-mediated carbon turnover

Viral dispersal across the air-sea interface

44 in remote areas, viral dispersal mechanisms relevant to human health, and involvement of
45 viruses in atmospheric processes like ice-nucleation.

46

47 **Key words:** viromics, foam, rain, atmosphere, sea-surface microlayer, neuston, ice-nucleating
48 particles, CRISPR

49

50 Marine viruses represent the most abundant biological entities in the oceanic water
51 column ¹ where they contribute to microbial diversity ², supply hosts with metabolic genes ³
52 and influence carbon cycling by inducing host cell lysis (the “viral shunt”)⁴. While viruses have
53 been studied in many marine habitats including the surface water ⁵ and deep-sea sediments ⁶,
54 their presence at the air-sea interface, where microorganisms modulate gas and organic matter
55 exchange processes ⁷⁻⁹, remains mostly enigmatic. Like many micro- and macroorganisms ^{10,11},
56 viruses accumulate in the thin (<1 mm) uppermost layer of aquatic ecosystems, the surface
57 microlayer (SML, reviewed by Cunliffe, et al. ¹²; Engel, et al. ¹³), where they belong to a pool
58 of organisms collectively referred to as neuston ¹⁴. The enrichment of the virioneuston in the
59 SML is mediated by bubble transport from the underlying water ^{15,16} and likely maintained by
60 viral attachment to particles ¹⁷ and a dependency on abundant prokaryotic hosts ¹⁸. In
61 freshwater, bacteriophages residing in the SML can form autochthonous communities ¹⁹ but
62 comparatively little viromics studies have been conducted for marine SML (reviewed by Rahlff
63 ²⁰). Sea foams float as (extended) patches on the sea surface (Supplementary Video), forming
64 deposits at the shoreline and being microbial habitats that contrast the SML ^{21,22}. Based on
65 satellite data, foams (whitecaps) cover up to 6 % of oceanic surface area and are expected to
66 become more frequent with climate change ²³. During storms, foams can flood beaches ¹⁵ and
67 massively pollute coastal areas like recently in Turkey ²⁴. Furthermore, sea foams can contain
68 pathogenic bacteria ²⁵ and their easy spread might be an important step for the dispersal and
69 aerosolization of its inhabiting microbes ²⁶ and potentially viruses. Foams can effectively
70 concentrate viruses ²⁷ which survive more than three hours of drying and sunlight when caught
71 in foams ²⁸. Virus-like particles (VLP) can reach a 300 x higher abundance in foams compared
72 to surrounding waters ²⁴.

73

74 Interest in studying viruses in the skin-like layer between ocean and atmosphere arises from
75 therein appearing human pathogenic viruses ²⁹, the potential of SML viruses to get airborne ³⁰,
76 to selectively enrich in aerosols ³¹, and to disperse over long distances to eventually promote
77 turnover of algal blooms in remote regions ³². Virus aerosolization from the SML was

Viral dispersal across the air-sea interface

78 previously studied ^{15,30,33}, but investigations pursuing metagenomic approaches to explore the
79 virionuston and its aerosolization in the field are lacking. A recent review highlighted the need
80 to quantify marine aerobiota, to characterize the spatial-temporal dimensions of dispersal, and
81 to understand acclimations of marine microorganisms to atmospheric conditions ³⁴. Once
82 airborne, viruses could even fulfill other functions as recently suggested ³⁵: Airborne marine
83 viruses could serve as ice-nucleating particles (INP), a function already described for many
84 microorganisms ^{36,37}, and act as catalysts to mediate freezing at temperatures warmer than
85 - 10 °C. INP exist in the SML ³⁸⁻⁴⁰ and have an important role in cloud formation, cloud albedo
86 and precipitation and thus are key in climate regulation dynamics ⁴¹. Viruses and bacteria can
87 be found in clouds ⁴²⁻⁴⁵, where the latter might grow selectively ⁴⁶ and as INPs, trigger their
88 own precipitation ^{47,48}. Precipitation could be an underestimated source of microorganisms to
89 Earth's surface, for example it contributed to as much as 95% of atmospheric bacterial
90 deposition at a Korean site ⁴⁶. So far, research on viruses included in wet precipitation was
91 mainly focused on viruses relevant to human health, such as enteric or adenoviruses ⁴⁹⁻⁵¹. Reche,
92 et al. ⁵² reported that 10⁷ bacteria and 10⁹ viruses per m² per day deposit above the atmospheric
93 boundary layer, with marine origin making stronger contributions than terrestrial sources. This
94 rate can be one order of magnitude higher for bacteria ⁵³ and perhaps also for viruses. Rain
95 events related to a hurricane decreased marine viral diversity and abundance as well as
96 introduced new taxa in the western Gulf of Mexico ⁵⁴. Furthermore, stormwater runoff changed
97 the viral community composition of inland freshwaters and a stormwater retention pond ^{55,56}.
98 However, the transmission cycle of neustonic viruses from the sea surface via aerosols to wet
99 precipitation remains speculative and the degree of viral exchange between these habitats and
100 along the natural water cycle is unknown.

101
102 To fill these knowledge gaps, we analyzed 55 metagenomes including air-sea interface habitats
103 (sea foams and SML), subsurface water (SSW) from 1 m depth, aerosols, and precipitation
104 (rain, snow) collected in a coastal region of the Skagerrak in Tjärnö, Sweden (Fig. 1A). We
105 explored the potential of marine viruses to become aerosolized and being returned to Earth via
106 wet deposition. We hypothesized that aerosols and rain included marine viruses, and that sea
107 foams have a particular role in the contribution to viral aerosolization due to their high content
108 of organic material, high air content, and direct contact with the atmosphere. Virus-host
109 infections histories were investigated by studying the prokaryotic adaptive immunity, namely
110 the clustered regularly interspaced short palindromic repeats (CRISPR) system. Extraction and
111 matches of CRISPR spacers, short sequences obtained from cell-invading mobile genetic

Viral dispersal across the air-sea interface

112 elements, to viral protospacers allowed to reveal virus-host relationships across ecosystem
113 boundaries.

114

115 **Results**

116

117 **Cell and VLP abundance, enrichments, and correlations thereof reveal tight virus-host** 118 **coupling in the neuston**

119

120 Marine, aerosol, and rain samples were collected around Tjärnö Marine Laboratory, Sweden
121 including 11 stations in coastal waters of the Skagerrak, where air-sea interface samples (SML,
122 foam) and a reference depth were sampled (Fig. 1A). Prokaryotic, small phototrophic
123 eukaryotic and VLP counts were measured to assess virus-host ratios and correlations in the
124 neuston (SML, foam) compared to the underlying plankton in the SSW. Across all stations,
125 flow cytometry revealed viral abundance between 5.0×10^7 - 1.8×10^8 , 1.3×10^7 - 3.4×10^7 and
126 1.4×10^7 - 2.0×10^7 VLPs mL⁻¹ in floating sea foams, the SML and SSW, respectively (Extended
127 Data Table 1, Extended Data Fig. 1). This increasing VLP gradient towards the atmosphere was
128 supported by microscopic analysis as shown for station 4 (Fig. 1B a-c, Fig. S1), and VLPs in
129 sea foams often adhered to particulate matter (Fig. 1B d). Counts of VLP in precipitation
130 samples (rainwater) ranged between 3.7×10^4 - 3.4×10^5 VLPs mL⁻¹. Enrichment factors (EF)
131 for VLPs in the SML over SSW varied between 0.7 (depletion) and 1.9 (enrichment) (Extended
132 Data Table 1). Total cell numbers of prokaryotes were 1.3×10^6 - 3.8×10^6 , 7.0×10^5 - $1.1 \times$
133 10^6 as well as 7.0×10^5 - 8.7×10^5 cells mL⁻¹ in floating sea foams (Fig. 1B e), the SML and
134 SSW, respectively (Extended Data Fig. 1, Extended Data Table 1). EFs for prokaryotes
135 fluctuated between 0.9 and 1.3. Across the five precipitation samples, 2.3×10^3 - 1.8×10^4
136 prokaryotic cells mL⁻¹ were detected. Virus-host ratios (host=prokaryotes) based on flow
137 cytometry data were highest in foams (range = 25.3 - 48.4), followed by the SML (range = 15.5
138 - 34.2) and SSW (range = 19.3 - 26.7). Virus-host ratios in precipitation samples showed the
139 strongest variation and ranged between 7.1 and 127.8 (Extended Data Table 1). The highest
140 virus-host ratios in the SML were detected on days where VLP EFs were ≥ 1.8 and prokaryotic
141 EFs ≥ 1.1 at the same time. Total cell numbers of small phototrophic eukaryotes ranged between
142 3.4×10^3 - 1.8×10^4 , 1.8×10^3 - 6.2×10^3 , 2.1×10^3 - 5.0×10^3 cells mL⁻¹ in sea foams, SML
143 and SSW, respectively.

144

Viral dispersal across the air-sea interface

145 Within the SML, the number of small phototrophic eukaryotes and prokaryotes was
146 significantly correlated with VLP abundance (Pearson's $\text{corr} = 0.74$, $t = 3.13$, $p = 0.014$, $\text{df} =$
147 8 , $n = 10$ and Pearson's $\text{corr} = 0.70$, $t = 2.75$, $p = 0.025$, $\text{df} = 8$, $n = 10$ respectively, Extended
148 Data Fig. 2 A&B), while the plankton/SSW correlations with the same variables were not
149 significant (Extended Data Table 2, Extended Data Fig. 2 C & D). In addition, absolute numbers
150 of small phototrophic eukaryotes and their EFs were significantly correlated with absolute
151 numbers and EF of prokaryotes for the neuston, inferring a common transfer mechanism of
152 these cell types to the air-sea interface (Extended Data Table 2, Fig. S2). EFs of VLP were
153 significantly correlated to EFs of prokaryotes (Spearman's $\rho = 0.62$, $p = 0.006$, $n = 10$) but
154 not to EFs of small phototrophic eukaryotes (Extended Data Fig. 2 E & F) probably indicating
155 that enrichments of viruses in the SML are primarily dependent on host cell availability and
156 that most SML viruses are prokaryotic viruses.

Viral dispersal across the air-sea interface

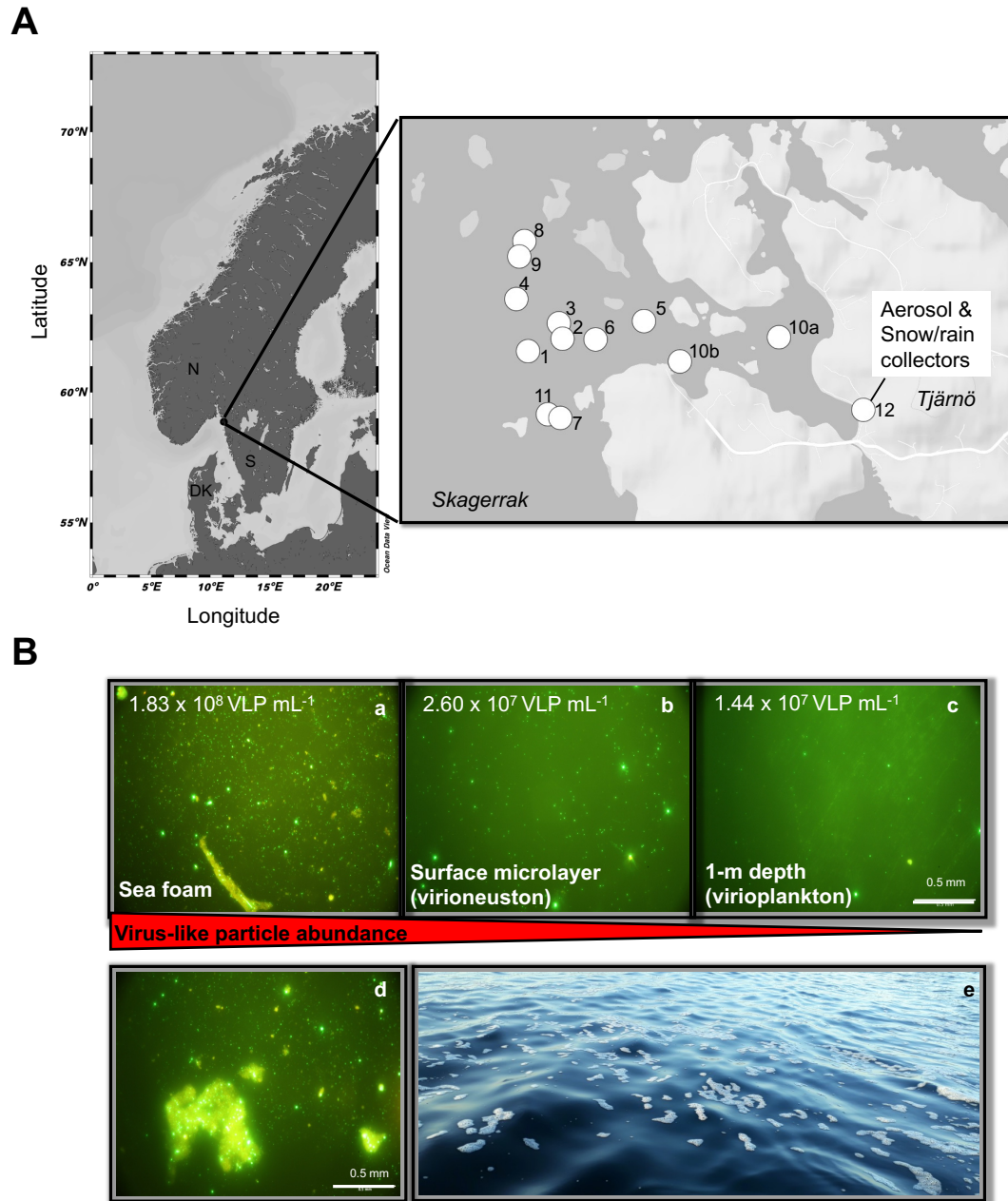


Figure 1: Map depicting sampling stations and viral enrichment in sea-surface microlayer and sea foam. Map of sampling sites was generated using Ocean Data View ¹¹⁸ and <https://maps.co>. For further details about the stations, please refer to Table S9 (A). Gradient of virus-like particles (VLP) in sea foam (a), sea-surface microlayer (b) and 1-m deep subsurface water (c) recorded in epifluorescence microscopy with VLP counts mL⁻¹ from Station 4 as obtained from flow cytometry. Virus-like particles stick to particulate matter in the foam habitat (d). Sea foams were collected as floating patches from the ocean's surface (e) (B).

Viral dispersal across the air-sea interface

1 **Correlations of cell and VLP counts with ice-nucleating particles, wind speed and** 2 **environmental parameters**

3
4 INPs, like VLPs, showed a decreasing gradient from sea foams, SML to SSW (Extended Data
5 Fig. 3), with INP in foams showing the highest concentrations over the detectable temperature
6 range. Ice activity for all samples generally started at high temperatures of ~ -4 to -6 °C,
7 comparable to observations for microorganisms in the atmosphere⁵⁷. Concentrations of INPs
8 were only significantly correlated with VLPs, small phototrophic eukaryotes and prokaryotes,
9 when all interface samples (sea foams and SML) were combined (Extended Data Table 2,
10 Extended Data Fig. 4), but were not significantly correlated with VLPs and cells in the SSW.
11 In addition, enrichment of VLPs (but not cells) in the SML was significantly correlated with
12 the enrichment of INPs (Spearman's $\rho = 0.86$, $p = 0.011$, $n = 8$ Extended Data Table 2).
13 Correlation analysis of absolute VLP counts derived from neuston and plankton as well as EFs
14 for VLPs and cells with meteorological data (light, salinity, and wind speed) did not reveal any
15 significant relationships (Extended Data Table 2). However, numbers of small phototrophic
16 eukaryotes (Pearson's $\text{corr} = 0.68$, $t = 2.59$, $p = 0.031$, $df = 8$, $n = 10$) and prokaryotes (Pearson's
17 $\text{corr} = 0.71$, $t = 2.83$, $p = 0.022$, $df = 8$, $n = 10$) from the SSW but not the SML were significantly
18 correlated with salinity, which could be explained by regular inflow of high saline water from
19 the Atlantic Ocean that probably affects deeper water layers more than the SML.
20 We applied linear models to investigate combinatory effects of environmental variables (wind
21 speed, light, and salinity) on enrichment of cells and VLPs in the SML. One linear model
22 considering the combinatory effects of wind speed and salinity on the enrichment of small
23 phototrophic eukaryotes in the SML passed the F-test with significance ($F = 5.43$, $p = 0.038$, df
24 $= 6$), and 59.6 % of the residuals were explained by this model. The model's Akaike information
25 criterion (AIC) was -2.37, which was superior to considering wind speed (AIC = 5.84) and
26 salinity (AIC = 6.40) alone. Other models testing single and combined environmental
27 parameters on the enrichments of cells and VLP in the SML were not significant.

28 29 **Aerosolization of biota and decreasing diversity from marine towards atmospheric** 30 **habitats**

31
32 Analysis of Shannon-Wiener Diversity Index of prokaryotic communities (extractions from
33 filters with 5 - 0.2 μm pore size) suggests a lower diversity of atmospheric (rain, aerosol)
34 microbiota compared to marine samples, with the difference between aerosols and SSW being

Viral dispersal across the air-sea interface

35 significant (Kruskal-Wallis with Dunn's multiple comparisons test, $p = 0.0097$, Fig. 2A).
36 Permutational multivariate analysis of variance (PERMANOVA) confirmed significant
37 differences for the non-metric multidimensional scaling (NMDS) analysis ($p < 0.001$, Fig. 2B)
38 with TukeyHSD on multivariate dispersion distances revealing significant differences between
39 aerosols and foam ($p = 0.02$), aerosols and SML ($p = 0.02$), and aerosols and SSW ($p = 0.0002$)
40 (Fig. S3). Correlation matrices showed that aerosol and rain communities are distinct from
41 marine communities and from each other (Extended Data Fig. 5A). Beta-diversity analysis
42 showed that SSW samples were mainly composed of Proteobacteria (mean relative
43 abundance \pm standard deviation = 67.8 ± 5.2 %, $n = 9$), Bacteroidetes (22.6 ± 6.0 %), and
44 Thaumarchaeota (5.1 ± 2.3 %, Fig. 2C). In general, the SML samples reflected this composition
45 although two samples deviated by containing a major percentage of Planctomycetes ($13.3 \pm$
46 28.7 %, $n = 9$) and Cyanobacteria (4.6 ± 7.2 %, $n=9$). Sea foams were like the SML but
47 additionally contained WOR-2 (2.7 ± 3.2 %, $n=3$) and an increasing proportion of Bacteroidetes
48 (37.7 ± 11.6 %, $n = 3$). Aerosols also contained Proteobacteria (43.4 ± 13.9 %, $n = 8$),
49 Bacteroidetes (21.0 ± 21.6 %) and Planctomycetes (20.8 ± 13.0 %). The snow sample contained
50 a relative abundance of 97.3 % Cyanobacteria (*Rivularia* sp.). The mean relative abundance of
51 Proteobacteria (49.9 ± 5.1 %) and Bacteroidetes (20.7 ± 2.4 %) in rain was comparable to
52 aerosols, but in contrast to sea surface water and aerosols, Actinobacteria (6.3 ± 9.4 %) and
53 Cyanobacteria (13.6 ± 5.1 %) were more abundant (Fig. 2C).

54
55 Detection of the same bacterial *ribosomal protein S3* genes in marine, aerosol and precipitation
56 samples suggests their aerosolization from the sea surface, e.g., for Proteobacteria
57 (*Oceanospirillum maris*, *Loktanella vestfoldensis*, *Candidatus Pelagibacter*), Bacteroidetes
58 (*Crocinitomix catalasitica* and *Bacteroides fragilis*), Cyanobacteria (*Crinalium epipsammum*
59 and *Oscillatoria* sp.) and Planctomycetes (*Pirellula staleyi*) (Fig. 2C&D). Interestingly,
60 aerosols and precipitation contained Cyanobacteria such as *Rivularia* sp. (max. 14.1 % in rain,
61 97.7 % in snow) and Proteobacteria such as *Sphingobium japonicum* (max. 21.8 %) or
62 *Methyloferula stellata* (max. 21.4 %), which could not be found in any of the local marine
63 samples (relative abundance = 0). Rain contained Actinobacteria (*Cryocola* sp., max. 20.3 %)
64 and Bacteroidetes such as *Mucilagibacter paludis* (max. 20.1 %) that were only scarcely
65 detected in marine samples (< 0.2 %) and thus probably originated from other sources.

Viral dispersal across the air-sea interface

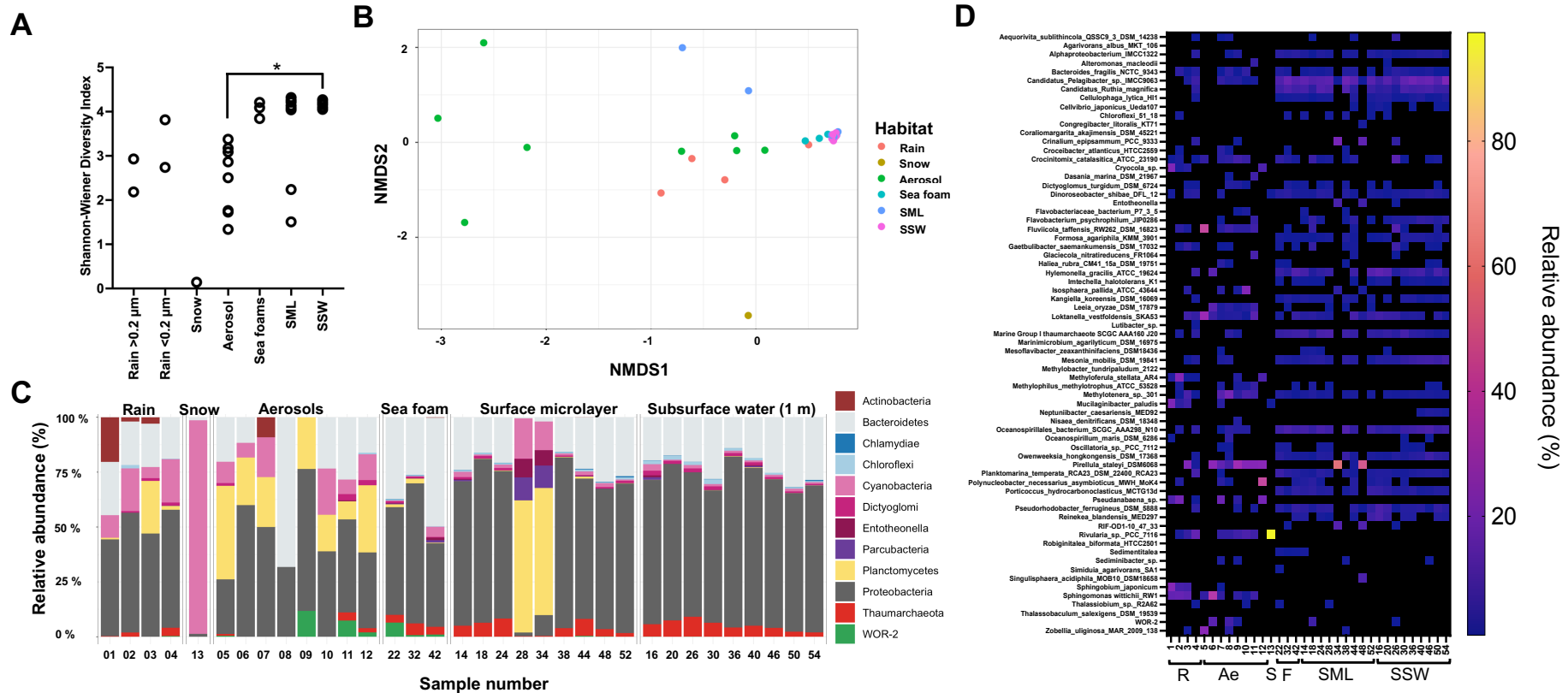


Figure 2: Diversity and relative abundance of marine and airborne prokaryotes based on relative abundance in rain, snow, aerosols, sea foam, surface microlayer (SML) and subsurface water (SSW). Diversity depicted by Shannon-Wiener index with * = $p < 0.05$ in Dunn's multiple comparison test (post hoc analysis after Kruskal-Wallis test) (A), non-metric multidimensional scaling plot based on Bray-Curtis dissimilarity (stress = 0.082) (B), and stacked bar chart on beta-diversity at the phylum level (C). In C, relative abundance is based on read-normalized coverage on scaffolds carrying the ribosomal protein 3 gene (*rps3*) as explained in the main text. Black areas represent accumulated taxonomic units of minor abundance. Seawater samples show result of > 0.2 μm samples, whereas rain contains >0.2 μm (#1+#3) and viromes (#2+#4). The heat map shows relative abundance of prokaryotic taxa across different ecosystems (D). Black colored fields indicate < 1% relative abundance; (R=rain, S=snow, Ae=aerosols, F=sea foams, SML=surface microlayer, SSW=subsurface water. Sample number is in accordance with Table S11;)

Viral dispersal across the air-sea interface

1 **K-mer based virus-host assignments reveal *Pelagibacter* and *Porticoccus* as prevalent** 2 **hosts**

3
4 In total, 116 metagenome-assembled genomes (MAGs) could be recovered from 24 different
5 samples (Table S1 & S2), which ranged from 58.8 – 100 % completeness (median = 86.3 %) and
6 0 – 11.8 % contamination (median = 3.9 %) based on uBin⁵⁸. CheckM⁵⁹ resulted in
7 completeness and contamination scores of 18.4 – 99.5 % (median = 80.6 %) and 0 – 17.3 %
8 (median = 1.6%) for these MAGs, respectively. Most host MAGs were of bacterial origin,
9 except for three assigned to the genus *Nitrosopumilus* (Archaea). Recovering MAGs from
10 atmospheric samples was particularly challenging with only a single MAG obtained from rain
11 (genus *Pedobacter*) and from an aerosol sample (class *Planctomycetes*, order *Pirellulales*),
12 respectively. Overall, bacterial MAGs were mostly classified as Gammaproteobacteria (n = 43),
13 Alphaproteobacteria (n = 30), Bacteroidia (n = 36), and Planctomycetes (n = 4). Based on read
14 mapping and breadth, all MAGs were detected in a marine ecosystem (except for the
15 *Pedobacter* sp. MAG), rain and some additionally in aerosols. (Table S3). MAGs were matched
16 to viruses based on shared k-mer frequency patterns, revealing that 120 marine viruses matched
17 a MAG assigned to *Candidatus Pelagibacter* (Extended Data Fig. 6). Hosts of rain viruses (not
18 detectable in other sampled ecosystems) and one aerosol virus were predicted as MAGs
19 belonging to the family *Porticoccaceae* and a *Flavobacterium*.

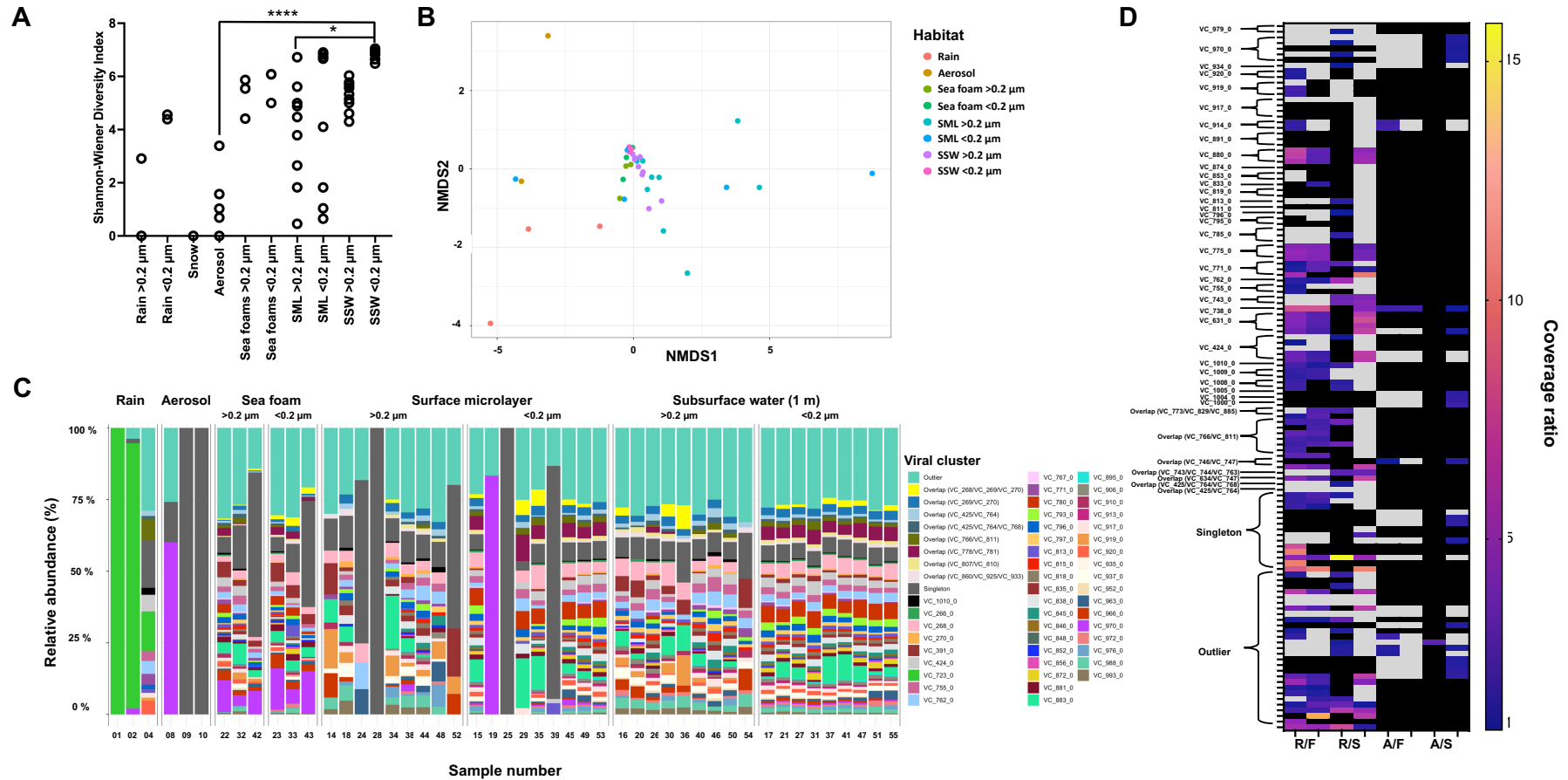
20 21 **Viral diversity and transfer from the sea surface to aerosols and rain**

22
23 Shannon-Wiener index values were significantly different for viruses between aerosols and
24 SSW (Dunn's multiple comparison test, $p < 0.0001$), but also weakly different between
25 SML > 0.2 μm fraction and SSW virome samples ($p = 0.029$) (Fig. 3A). The distinct viral
26 community of SML > 0.2 μm samples was also demonstrated by beta-diversity analysis
27 (Fig. 3 B&C). Here, PERMANOVA confirmed significant differences for the NMDS analysis
28 ($p = 0.001$), and *post hoc* pairwise analysis revealed significant differences between the SSW
29 virome and the SML 0.2 μm fraction ($p = 0.013$, Fig. S3). We investigated further on SML and
30 foam viral clusters (VC) that were detected in aerosols and rain. Rainwater contained the
31 abundant cluster VC_723_0 absent in samples from the other ecosystems (Fig. 3C) and with
32 one associated scaffold related to *Rhizobium* phage RHph_N3_2. EFs were overall higher for
33 VCs in rain (EF max. = 15.8) compared to enrichments in aerosols in reference to SML and
34 foam (EF max. = 2.8, Fig 3D). VC_880_0 (max. EF = 7.0), VC_738_0 (max. EF = 7.8), and

Viral dispersal across the air-sea interface

35 VC_771_0 (max EF = 5.4) were strongly enriched (coverage ratio > 1) in rain compared to
36 foam and/or SML but contained only viruses from this study and were unrelated to viruses from
37 public databases. Enriched in rain, overlap cluster (VC_634/VC_747) and overlap cluster
38 (VC_746/VC_747) were both related to *Pelagibacter* phage HTVC023P (max. EF = 7.0, Table
39 S4), whereas overlap cluster (VC_773/VC_829/VC_885) was related to *Flavobacterium* phage
40 vB_FspM_immuto_3-5A (max. EF = 2.6). An overlap cluster refers to genomes sharing genetic
41 overlap with other genome(s) belonging to multiple VCs. VConTACT2 detected various
42 singletons and outliers, which usually represent new viruses, and had an EF > 8 for rain over
43 sea surface ecosystems but were unrelated to any known virus. In aerosols, e.g., VC_970_0
44 (max EF = 1.2), VC_914_0 (max EF = 2.1), and VC_738_0 (max. EF = 1.9) showed slight
45 enrichments compared to marine samples but were also unrelated to any known viruses (Fig.
46 3D). Three singletons and nine outliers were additionally enriched in aerosols with outliers
47 (EF = 2.8 and 1.4) being associated with a virus related to *Methylophilales* phage Melnitz-1
48 EXVC043M and *Vibrio* phage vB_VorS-PVo5, respectively. Overall, assembled marine
49 viruses shared protein clusters with *Synechococcus*, *Rhizobium*, *Cellulophaga*, *Flavobacteria*,
50 *Vibrio* and *Pelagibacter* phages in vConTACT2⁶⁰ (Table S4, Extended Data Fig. 7). The
51 correlation matrix across all viromes and samples shows most positive correlations between
52 foam, SML, SSW, which are well-interconnected systems, and some positive correlations of
53 specific marine samples with aerosol and rain samples (Extended Data Fig. 5B). Viromes of
54 marine samples and rain samples were sometimes even negatively correlated, suggesting
55 alternative sources of rain viruses other than the sea surface.

Viral dispersal across the air-sea interface



56
57
58
59
60
61
62
63

Figure 3: Diversity and enrichment analysis of marine and airborne viruses based on relative abundance in rain, snow, aerosols, sea foam, surface microlayer (SML) and subsurface water (SSW). Alpha-diversity for all samples (except for #12, which contains 0 viruses) depicted by Shannon-Wiener index; * = $p < 0.05$, **** = $p < 0.0001$ in Dunn's multiple comparison test (post hoc analysis after Kruskal-Wallis test) (A), non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarity (stress = 0.06) (B), and stacked bar chart on beta-diversity (C). If samples only contained rare viruses (sample #7), a single virus (#3, #5, #6, #11 #13) or no viruses (#12), they were removed, and only the relative abundance of the 200 most abundant viruses assigned to viral clusters (VCs), outliers, and singletons were considered for B&C. In vConTACT2, outliers and singletons typically represent new viruses. In (C), marine samples are separated by size fraction: > 0.2 μm = prokaryote fractions, < 0.2 μm

Viral dispersal across the air-sea interface

64 = viromes; Rain sample #1 is a $> 0.2 \mu\text{m}$ sample, whereas #2 and #4 are rain viromes. Enrichment ratio of SML and foam viruses in rain and aerosols
65 (D). Shown are ratios ≥ 1 of virus coverage for rain/foam (R/F), rain/SML (R/S), aerosol/foam (A/F), aerosol/SML (A/S), where the left tick stands
66 for foam and SML virome and the right tick for foam and SML $0.2 \mu\text{m}$ fraction in the denominator. Black fields mean that the virus was absent in one
67 or both ecosystems in the respective sample. Grey areas show out of range fields (ratio between 0 and 1, indicating depletion). Sample number is
68 explained in Table S11.
69

Viral dispersal across the air-sea interface

We then investigated the aerosolization patterns of two circular viral genomes of similar length and carrying viral hallmark genes (terminases, portal protein) across the different habitats and stations (Fig. 4A+B). Virus_1 (39.7 kb, percent G/C content = 46.7 %, no RefSeq match in vConTACT2) was constantly of lower abundance in seawater samples compared to Virus_2 (35.1 kb, percent G/C content = 35.1 %, no RefSeq match in vConTACT2) across different stations. However, Virus_1 was consistently abundant in sea foams and was additionally found in three aerosol samples and two rain samples. Instead, Virus_2 was absent from the atmosphere despite its abundance in surface water. Virus_1 was linked to a *Porticoccus* MAG based on k-mer patterns (Extended Data Fig. 6) and single-nucleotide polymorphism (SNP) analysis revealed multiple SNP overlaps for this virus between a foam, aerosol, and rain sample, supporting its transfer from the sea surface to aerosols and rainwater (Fig. 4C&D, Table S5). Cross-mapping of reads from all samples against all 1813 viral scaffolds revealed shared viral populations between ecosystems (Fig. 6A). Sea foams, SML and SSW shared 837 viruses, whereas 15 viruses were present in all studied ecosystems. Overlaps between aerosols and rain samples must be treated with caution, because some rain could have reached the aerosol filter membrane during sampling and filter exchange (Table S6), although we tried to rule out the second possibility by subtracting reads from handling controls. Precipitation had the highest number of unique viruses (109), followed by foams (25), SSW (18), SML (7) and aerosols (6). Uniqueness means that no other habitat had 90 % identical reads with 75 % scaffold coverage of at least 1 x for that virus. A percentage of 6.2 % (111) of all viruses was shared between seawater including foam and precipitation. Interestingly, the rain sample pooled from Feb. 14th to 22nd of Feb. 2020 (event 2) defined most of this overlap compared to a sample from Feb. 7th and 9th February (event 1): Based on read-mapping, rain events 1 and 2 were associated with 22 versus 85 viruses assembled from marine samples as well as 112 versus 44 viruses assembled from rain, respectively. Based on read-mapping, event 1 delivered 38 marine prokaryotic MAGs (min. 90 % genome covered with reads), whereas 79 marine MAGs were found in the rain sample belonging to event 2 (Fig. 5). To explain these differences by tracking to potential sources, backward trajectories (TJs) for air masses were calculated. They showed that during event 2, air masses spent, during the first four days before arriving at the site, on average 72 % of their time over the sea and loading conditions (loading of air masses with generic particles) were fulfilled on average 35 % of the TJ. On the other hand, for event 1, air masses spent less time above the sea (64 %) and loading conditions were fulfilled, on average, only 10 % of the TJ points (Fig. 5).

Viral dispersal across the air-sea interface

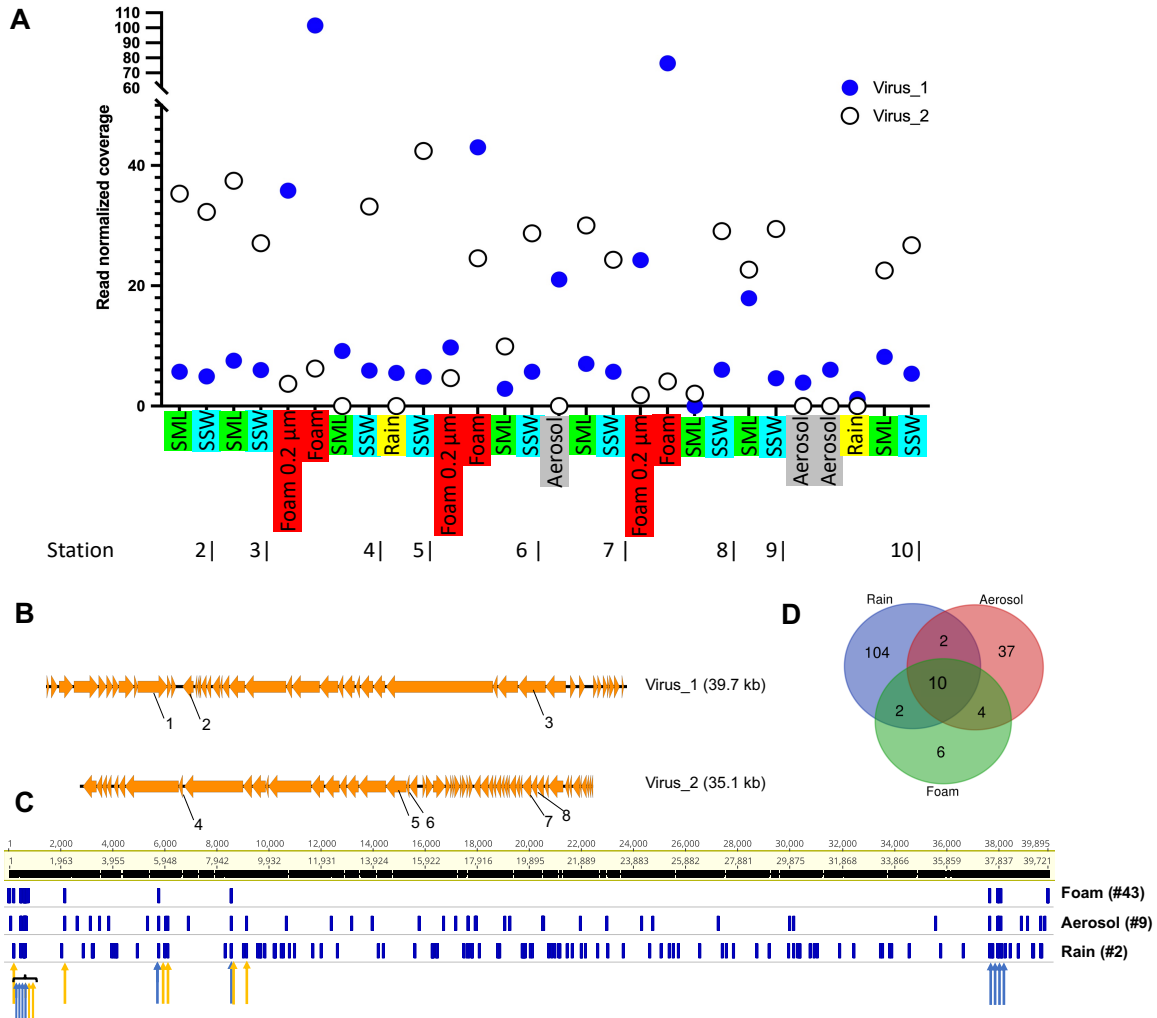


Figure 4: Succession of the coverage of two circular viral genomes across 28 metagenomes derived viromes of surface microlayer (SML), 1-m deep subsurface water (SSW) and sea foam as well as from sea foam filtered onto 0.2 μm membranes, aerosol, and rainwater samples (A). Synteny and functional annotations of the two circular viruses visualized using Easyfig⁸² and annotated with DRAM-v⁸⁰. Functional annotations are 1: Bifunctional DNA primase/polymerase, N-terminal [PF09250.12], 2: Sec-independent protein translocase protein (TatC) [PF00902.19], 3: Terminase, 4: Concanavalin A-like lectin/glucanases superfamily [PF13385.7], 5: Phage P22-like portal protein [PF16510.6], 6: Terminase-like family [PF03237.16]; Terminase RNaseH like domain [PF17288.3], 7: C-5 cytosine-specific DNA methylase [PF00145.18], 8: PD-(D/E)XK endonuclease [PF11645.9] (B). Variant analysis of Virus_1 for a sea foam, aerosol and rain sample reveals overlapping nucleotide polymorphisms. Blue and orange arrows indicate overlaps between three and two samples, respectively. For details, please see Table S4 (C). Venn diagram showing variant overlaps for Virus_1 in different ecosystems as shown in C (D).

Viral dispersal across the air-sea interface

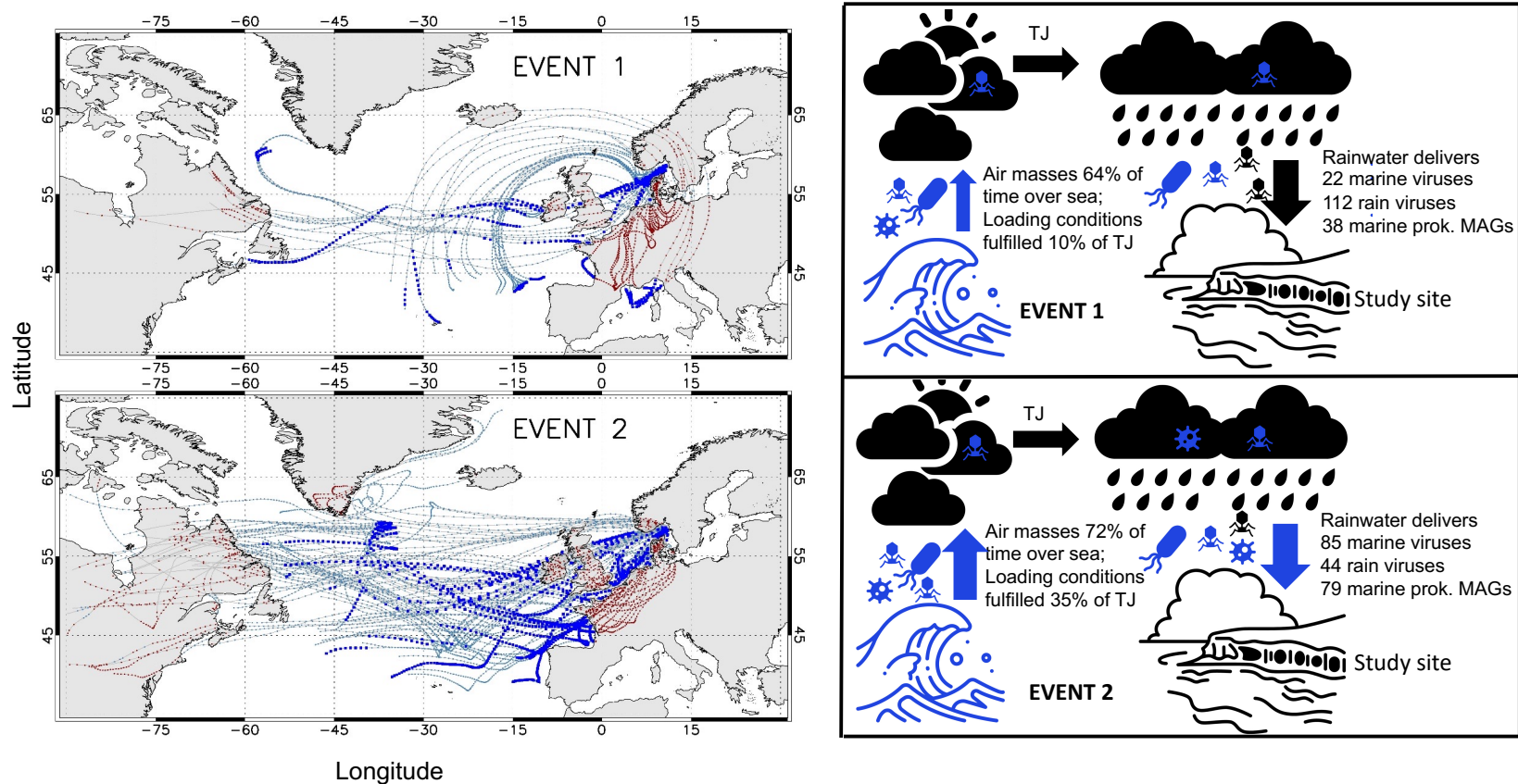


Figure 5: Backward trajectories (TJs) of two rain events leading to different deliveries of marine viruses and metagenome-assembled genomes (MAGs) at the study site. Event 1 (upper left panel) refers to a rainwater sample from 7th to 9th of Feb., and event 2 (lower left panel) to a sample collected between 14th to 22nd of Feb. 2020. Sky blue and red points highlight where backward trajectories travel above sea and land, respectively. Blue filled squares represent points where loading conditions were fulfilled (please see main text for further explanation). Panels on the right show corresponding deliveries of viruses and MAGs to the study sites. A virus was considered marine or from rain if assembled in such a sample and counted if detected based on read mapping. Marine MAGs were considered present in rainwater 0.2 μm samples if 90% of the genome was covered with reads (see Table S3).

Viral dispersal across the air-sea interface

Rain and aerosol viruses show adaptations toward atmospheric residence and are targeted by marine prokaryote adaptive immunity

To investigate if unique atmospheric viruses have genetic adaptations, we explored the content of guanine (G) and cytosine (C) bases in viral scaffolds. Viral scaffolds solely occurring in rain samples (Rain_unique, n = 109) exhibited a significantly higher percent G/C content than total viruses found in rain (Kruskal-Wallis with Dunn's multiple comparisons test, $p = 0.0002$). All viruses found in marine samples had a highly significantly lower percent G/C content compared to aerosol, total rain and unique rain viruses when compared pairwise (KW-test, $p < 0.0001$, Fig. 6B).

One very abundant circular viral genome was identified as a unique rain virus (VC_723_0, 39 kb, coverage = 189 x, percent G/C content = 59.7 %) with the closest relative of the VC being the *Rhizobium* phage RHph_N3_2. This phage carried typical phage hallmark genes like a major capsid protein, an endonuclease, a terminase and modification methylase, but also carried many hypothetical proteins (Table S7). In addition, two large viral scaffolds unique to rain (270 kb and 496 kb) and with some genes related to *Mimiviridae* and k-mer linked to a *Flavobacteriaceae* MAG, encoded for sensors of blue-light using FAD (BLUF, PFAM ID PF04940.13), a photoreceptor and for an UV-endonuclease UvdE (PF03851.15). Another 16 kb viral genome with typical phage proteins (terminase, capsid) encoded for Tellurium resistance genes TerD (PF02342.19). From metagenomic assemblies, and from one MAG of *Schleiferiaceae* bacterium MAG-54, CRISPR arrays with evidence level 3 and 4 from 18 different samples could be detected by CRISPRCasFinder⁶¹, and mostly belonged to marine ecosystems (n = 14) and rainwater (n = 4, Table S8). CRISPR spacers extracted across all samples based on consensus direct repeat (DR) sequences from recovered arrays matched protospacers of viruses from seawater, but also the unique rain viruses (Fig. 6C, Fig. S4). Interestingly, CRISPR spacers matching most viral protospacers were extracted from two dominant arrays, with one of them targeting primarily (unique) rain viruses and the other one marine viruses (Fig S4, Table S8).

Viral dispersal across the air-sea interface

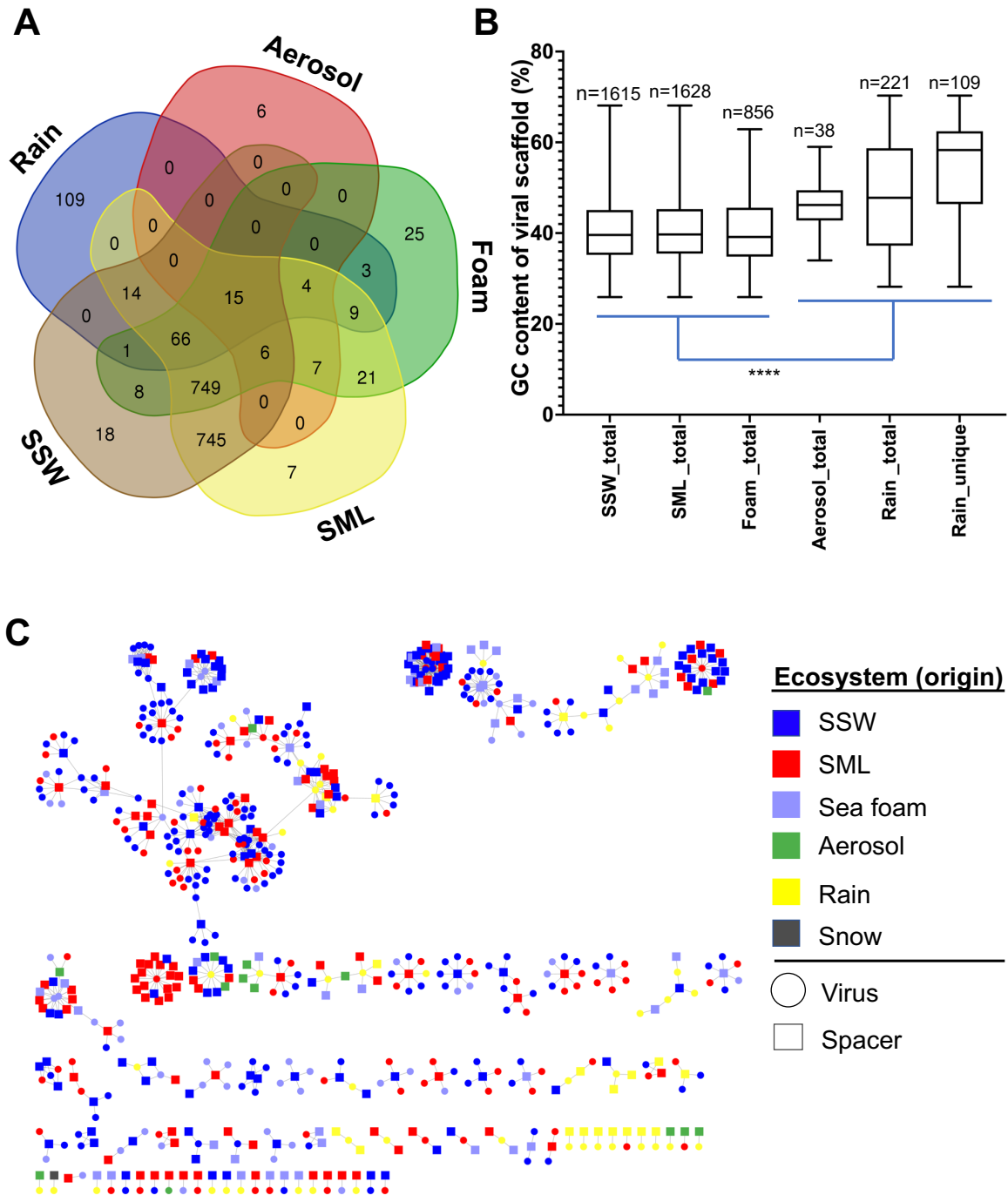


Figure 6: Overlapping occurrence of viral scaffolds, their percent G/C base content and CRISPR spacer to viral protospacer hits. Overview of shared viral scaffolds (>10 kb) length between seawater, aerosol and precipitation habitats obtained from 55 metagenomes and determined by mapping of reads. A viral genome was considered present in a sample if at least 75 % of the genome were covered with reads at least 90 % identical to the genome, in accordance with suggested viromics benchmarks⁷⁸ (A). Percent of the bases guanine (G) and cytosine (C) in viral scaffolds from rain, aerosol, foam, surface microlayer (SML) and subsurface 1-m deep water (SSW) based on read mapping. “Rain_unique” refers to viral genomes exclusively found in rain. Stars indicate significant differences after Kruskal Wallis test and Dunn’s multiple comparison test (****, $p = <0.0001$). In each pairwise comparison, the marine groups were significantly different from the atmospheric groups. Rain_total was also significantly different from Rain_unique (***, $p = <0.001$), which is not indicated to

Viral dispersal across the air-sea interface

reduce complexity of the figure (B). CRISPR spacers (origin indicated as square) matching assembled viral scaffolds (circles) derived from different habitats (C).

Discussion

The atmosphere is known to mediate rapid transport of prokaryotes and viruses from marine and terrestrial sources over thousands of kilometers^{62,63}. Aerosol and rain samples contained MAGs and genetic signatures from marine microbes, suggesting that aerosolization from the sea surface generally took place. We found that precipitation samples for instance contained cyanobacteria of *Rivularia* sp. (order *Nostocales*), but this species could not be found in marine samples from the study site inferring transport from remote areas. Especially the rainwater community related to event 2 contained a mixture of local and probably dispersed prokaryotes. This confirms their travel along the natural water cycle and that the origin of the air mass is crucial for understanding airborne microbial diversity, especially near the ocean-atmosphere interface⁶⁴.

Due to the tighter coupling of VLPs to their hosts in the neuston, one of the determining factors for a virus to get aloft is probably being associated with a host cell or other particulates, such as transparent exopolymer particles. These are prone to aerosolization, were found in cloud water and to absorb viruses^{65,66}. Viral attachment to biotic and abiotic surfaces in seawater is a common phenomenon⁶⁷ and more likely to occur in the SML¹⁷ and in foams (Fig. 1B). We found that VLPs in foams often adhered to particulate matter though $>50\ \mu\text{m}$ particles were removed prior flow cytometric measurements, probably leading to an underestimation of the measured VLP counts in particle-rich sea foams. Michaud, et al.³³ used a mesocosm experiment to show that bacterial and viral aerosolization is taxon-specific, and our data support and extend the concept of selective aerosolization from the water surface under field conditions. Two viral genomes experienced the same environmental conditions but showed a consistently distinct aerosolization behavior on a spatial-temporal scale. Virus_1 became constantly more abundant in sea foams at different stations, and SNP analysis proved its transfer to rainwater. As enrichments of viruses in rainwater over foam compared to rainwater over SML were enhanced in the $> 0.2\ \mu\text{m}$ fraction (Fig.3 D), aerosolization from sea foams as a “springboard” from sea to air represents a likely scenario. In addition, correlation analyses demonstrated independence of VLP enrichment in the SML from various environmental conditions, but findings only apply for wind speeds $< 6\ \text{m s}^{-1}$, since higher wind speeds prevent SML sampling and bacterioneuston enrichment⁶⁸. The impact of wind-induced sea spray formation on virus and host

Viral dispersal across the air-sea interface

aerosolization, needs future investigations for instance by carrying out experiments in wind-wave facilities.

Atmospheric dispersal of viruses allows the spread of “foreign” genetic material into new habitats enabling bacterial evolution and explains why similar viral genomes can be found across large geographical distances⁶⁹. The fact that rain samples shared up to 6.2 % of the virome with marine samples, generally indicates profound viral exchange between both ecosystems. This is further supported by CRISPR spacers from sea surface prokaryotes matching viruses found exclusively in rain samples. Established adaptive immunity indicates a long-lasting virus-host relationship and proves a constant genetic inflow to the sea surface by precipitation.

Previous work reported that bacterial isolates from rain are often pigmented⁷⁰, aiding their survival at high UV radiation. Rain and aerosol viruses detected in this study had a significantly higher percent G/C base content compared to marine viruses, which is a suitable adaptation for avoidance of thymine-specific damage by UV radiation in bacteria⁷¹ and occurred in recently described bacterial isolates from the stratosphere⁷². A viral cluster specific to rain was identified, and many viruses enriched in the atmosphere compared to seawater lacked representatives in public databases. Some viruses carried genes, e.g. for Tellurium resistance, which could be relevant for tolerating atmospheric conditions where Tellurium contaminations occur⁷³. As genetic adaptations cannot be established within hours, for instance shortly after aerosolization, we assume that the atmosphere contains its own virome, likely maintained across large areas and different altitudes, supplied by marine or terrestrial sources, and then distributed up to the higher troposphere or even above. The hereby detected 109 distinct rainwater viruses associated with rain event 1 and carrying genes related to atmospheric life supports this conclusion. In contrast, air masses delivering rain during event 2 were loaded over sea and contained more marine viruses and prokaryotic MAGs. We speculate that these viruses endure less well in the atmosphere than the high percent G/C-adapted rain viruses but can still be dispersed with air masses or clouds and deposited from them. Consequently, air mass and cloud migration and therein transported microbes and viruses could have strong impacts on biogeography, diversity, and evolution of microbial communities on the ground, particularly in aquatic habitats. Future work using culture-dependent experiments is required to elucidate if marine viruses remain infective when they are returned to the surface during precipitation. Since sea surface hosts possess CRISPR spacers matching genomes of unique rain viruses, ongoing infections with these viruses in the sea surface are probable. Moreover, there is a tendency that at ecosystem boundaries, different CRISPR arrays are in charge for targeting

Viral dispersal across the air-sea interface

viruses from different origins (air, sea), although we assume that these arrays have high turnover rates in this dynamic interface ecosystem. Future work is certainly required to confirm the here reported trends. In conclusion, our study shows broad dispersal of microbes and viruses along the natural water cycle and emphasizes that viruses crossing interfaces and ecosystem boundaries, e.g., with rainwater, can have a crucial role for shaping Earth's aquatic microbiomes.

Online methods

Seawater sampling and processing

Seawater sampling sites were located in the bay offshore Tjärnö, Swedish west coast in the Skagerrak (Fig.1), an area characterized by strong salinity gradients⁷⁴ (Table S9). Foams and SML were sampled from a small boat using the glass plate method⁷⁵ as previously described²². Corresponding subsurface water (SSW) from 1 m depth was collected as a reference using a syringe connected to a weighted hose. Wind speed was measured with a handheld VOLT-CRAFT AN-10 anemometer (Conrad Electronic, Hirschau, Germany), and light conditions were recorded on the boat using the Galaxy Sensors smartphone application v.1.8.10. Temperature and salinity were measured from the small boat using a portable thermosalinometer (WTWTM MultiLineTM 3420).

Water samples were stored in the dark and on ice until processing in the laboratory. Filtration equipment was treated prior to all usages with household bleach and rinsed with MilliQ water. Seawater (500 mL SML and 2 L SSW) and sea foams (200-400 mL) were sequentially vacuum filtered through 5 µm and 0.2 µm pore size Omnipore PTFE filter membranes (47 mm diameter, Merck/Sigma Aldrich, Munich, Germany). The flow-through of the 0.2 µm filter membrane was precipitated with 1 mg L⁻¹ iron-III-chloride (Alfa Aesar/Thermo Fisher Scientific, Uppsala, Sweden) for 1 hour at room temperature⁷⁶, and the flocculates were in turn filtered onto another 0.2 µm Omnipore PTFE filter membrane to obtain viruses and small prokaryotes. All filters were stored at -80° C until further processing and shipped on dry ice to the home laboratory for DNA extraction from the 0.2 µm filter and the FeCl₃ flocculates.

Aerosol and precipitation sampling

Viral dispersal across the air-sea interface

We used a land-based aerosol pump/constant flow sampler (QB1, Dadolab, Milan, Italy) with a custom-made filtration unit (SIMA-tec GmbH, Schwalmtal, Germany) to filter aerosols from the atmosphere in coastal proximity (Fig. S5). Incoming air was filtered through 0.1 μm pore sized Omnipore PTFE filter membranes (Merck/Sigma Aldrich, Munich, Germany). Filtered volumes and filtration duration varied and ranged from 19 to 61 m^3 (average volume flow 7 L min^{-1}) and from 24 to 96.5 hours, respectively (Table S5). Gas volume was normalized to mean temperature and mean air pressure from start and end of an aerosol filtration. Handling controls for aerosol samples were collected as follows: a filter membrane was briefly placed on the filter unit, and directly frozen in a falcon tube at $-80\text{ }^\circ\text{C}$. Snow and rain with a volume of 90 mL and 150 to 1050 mL, respectively were collected using funnels taped to Duran glass bottles. Rain was collected and, like seawater, filtered onto 0.2 μm pore size PTFE filter membranes, and the viral fraction was obtained as explained above. Rain collected between 14th to 22nd of Feb. 2020 was prefiltered onto 5 μm due to visible pieces (probably plant-based) in the sample. For enough DNA yield for sequencing, DNA from rain for the periods 07-09. Feb and 14th to 22nd of Feb. 2020 were pooled, respectively. The snow sample was prefiltered on 5 μm and frozen at $-80\text{ }^\circ\text{C}$. Later in the home laboratory, it was thawed at room temperature and concentrated in an Amicon® Ultra-15 centrifugal filter unit (Ultracel 100 kDa) by spinning in several steps at 3000 x g, 10 minutes at 4 $^\circ\text{C}$ before DNA extraction.

Air mass paths (backward trajectories)

Transport pathways of air masses was evaluated with 5-day backward trajectories (TJs) generated using the Hybrid Single-Particle Lagrangian Integrated Trajectories (HYSPLIT) model ⁷⁷. The TJs were calculated every one hour ending at 700 m above the site for the period 1st to 29th February 2020. The European Centre for Medium-range Weather Forecasts (ECMWF) ERA5 model atmospheric reanalysis ⁷⁸ is used to initialize HYSPLIT. After five days, the uncertainty associated with trajectories is estimated between 10 and 30 % of the travel distance⁷⁹. Each TJ was then projected on the 10-m wind, total precipitation, land mask, surface pressure and cloud fraction model fields (ERA5), associating each point along the path with the nearest values of the considered model variables. The choice of the ending height (700 m) above the site is based on the analysis of *in situ* meteorological (Nordkoster A Automatic Weather Station, 58.890 $^\circ\text{N}$, 11.010 $^\circ\text{E}$ as obtained from SMHI, <https://www.smhi.se/en>) and model (ERA5) data for February 2020 (Fig. S6). To identify loading areas and air masses presumably responsible for the transport towards the site, a selection of TJs was carried out considering

Viral dispersal across the air-sea interface

those (ones) arriving above the site during precipitation sampling events 1 (7-9 February 2020) and 2 (14-16 and 20-22 February 2020). Similar to Becagli, et al. ⁸⁰, loading conditions along TJs were evaluated searching where each TJ was within the mixing layer and wind speed at surface was greater than 3 m s⁻¹.

Microbial cell counts and virus-like particle abundances

Duplicates of unfiltered seawater, foam and precipitation samples were fixed with glutardialdehyde (1% final concentration, Merck, Sweden), stored for 1 hour in the dark and subsequently stored at -80 °C. Particle-enriched foams were gravity filtered onto 50 µm filters (CellTrics®, Sysmex Partec, Muenster, Germany) before cell counts of prokaryotes and small phototrophic (autofluorescent) eukaryotes were measured by a flow cytometer (BD Accuri C6, Becton Dickinson Biosciences, Franklin Lakes, USA) according to established protocols ⁸¹⁻⁸³. Due to previously reported low coefficient of variance among SML biological replicates in flow cytometry ⁶⁸, we did not measure biological replicates. VLPs were measured as done previously ⁸⁴. The gating strategy is shown in Fig. S7. Flow cytometry results were further compared to VLPs counted under the epifluorescence microscope (Leica DMRBE Trinocular, Leica Microsystems). Enrichment factors (EFs) were calculated as previously performed ²². EF > 1 and < 1 indicate an enrichment and a depletion of measured specimens, respectively.

Ice-nucleating particles

INP were measured from the 5 µm filter membrane that was used for pre-filtration of seawater samples. Of these filters, small disks with 1 mm diameter were punched out, using biopsy punches, and each disk was immersed in 50 µL of ultrapure water in a well of a 96-well PCR tray (BrandtTech®, Essex, CT, USA). For each filter membrane, 24 punches were examined, filling one quarter of a PCR-tray. The PCR-tray was then sealed and cooled down in an ethanol bath of a thermostat with a cooling rate of 1 K min⁻¹, while a camera took pictures every 0.1 K from below. On these pictures, frozen wells can be well distinguished from unfrozen ones, and the cumulative number of frozen wells was assessed for the different samples, a clean filter and pure water. Concentrations of INP were calculated from the cumulative number of frozen droplets, based on the known amount of filtered water and Poisson statistics. A more thorough description of the measurement method and data evaluation was described earlier ⁸⁵.

Viral dispersal across the air-sea interface

Statistical analyses

Correlations between abundances of prokaryotic cells, small phototrophic eukaryotes and VLPs were investigated using R version 4.0.3.⁸⁶ within R studio⁸⁷. Pearson or Spearman correlations were applied after the Shapiro-Wilk test confirmed normal distribution of data and residuals (for linear models). Dependences of EFs on environmental variables (wind speed, light, salinity) and interactive effects of those parameters were further investigated using linear regressions, and the models were validated using adjusted R² and AIC in the R programming environment. Differences in alpha diversity and viral percent G/C base content were analyzed using a Kruskal-Wallis test with Dunn's multiple comparison as *post hoc* analysis in Graphpad Prism v.9.4.1. Ecosystem-based differences in beta diversity shown in NMDS plots were assessed using PERMANOVA (n = 999 permutations) as well as Betadispersion analyses followed by a TukeyHSD test and executed by 'adonis2' and 'betadisper' function of the R package vegan⁸⁸, respectively.

DNA extraction and sequencing of metagenomes

Genomic DNA was extracted from seawater (0.2 µm and <0.2 µm flocculated viral fraction), rain filter membranes (47 mm diameter, Merck/Sigma Aldrich, Munich, Germany), and the concentrated snow sample using the DNeasy PowerSoil Pro Kit (Qiagen, Hilden, Germany). DNA from aerosol filters (90 mm diameter, Merck/Sigma Aldrich, Munich, Germany) was extracted using DNeasy PowerMax Soil Kit (Qiagen) with a subsequent DNA precipitation step. After concentration in a speed-vac "Concentrator plus" (Eppendorf AG, Hamburg, Germany), DNA was quantified using Qubit™ dsDNA High Sensitivity Assay Kit on a Qubit™ 4 Fluorometer (Invitrogen/Thermo Fisher Scientific) and sent for metagenomic sequencing to Fulgent Genetics (CA, USA). Library preparation was done according to the Illumina DNA Prep with Enrichment Reference Guide (Document # 1000000048041 v05, June 2020).

Metagenomic analyses

Raw shotgun sequencing reads of seawater (foams, SML, SSW), aerosols and precipitation datasets were quality-trimmed using `bbduk` (<https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh>) and Sickle⁸⁹.

Viral dispersal across the air-sea interface

Sequencing controls were assembled using MetaSPAdes version 3.13⁹⁰ and used as a blueprint for read mapping⁹¹ of actual samples; any reads that mapped to the negative controls were removed from downstream analyses (https://github.com/ProbstLab/viromics/tree/master/extract_unmapped_stringent). The same procedure including handling controls was carried out for metagenomic reads of aerosol samples.

Within a snakemake workflow⁹² designed for detecting viruses and prokaryotes, quality-controlled paired-end reads were first assembled with MetaviralSPAdes⁹³ and reads were mapped back⁹¹ to the assembly. Unassembled reads were assembled using MetaSPAdes version 3.14⁹⁰, and the two assemblies were joined for downstream processing. VIBRANT v.1.2.1.⁹⁴, VirSorter v.1⁹⁵ (only category 1, 2, 4, 5 were considered) and ViralVerify⁹³ were used to identify viral scaffolds and host contamination was removed with CheckV⁹⁶. Only viral scaffolds >10 kb were kept and clustered at species level (95 % similarity) using VIRIDIC⁹⁷, and the longest or circular scaffold of each cluster was used as representative. Metagenomic reads were mapped to >10 kb viral genomes with at least 90 % identity using Bowtie2⁹¹ with settings mentioned previously⁹⁸. To show the succession of two circular viral genomes across different samples, a separate mapping was done for these two scaffolds, and SNP analysis was performed for one marine virus that got airborne using Geneious v.11.1.5⁹⁹ with default settings for variant analysis. Venn diagrams were constructed using ugent Webtool (<https://bioinformatics.psb.ugent.be/webtools/Venn/>). Complying to current viromics conventions¹⁰⁰, only scaffolds covered with 75 % of reads were considered further, and breadth was checked with Calcopo (<https://github.com/ProbstLab/viromics/tree/master/calcopo>). Mean coverage of viral scaffolds was calculated (https://github.com/ProbstLab/uBin-helperscripts/blob/master/bin/04_01calc_coverage_v3.rb), and sum-normalized based on sequencing depth. Genes on viral scaffolds were predicted using Prodigal v.2.6.3¹⁰¹ in meta mode and resulting coverages were further read-sum normalized and functionally annotated using DRAM-v¹⁰². Synteny of viral genomes was visualized using Easyfig v.2.2.5¹⁰³. Clustering of dereplicated viral genomes with a RefSeq database (release Dec. 2021)¹⁰⁴ was performed using vConTACT2 v.0.9.19^{60,105}. Information on VCs and closest relative were compiled using graphanalyzer v.1.5.1 (<https://github.com/lazzarigioele/graphanalyzer>), and networks visualized in Cytoscape v.3.9¹⁰⁶. Relative abundance of VCs was used for beta-diversity analysis. The % G/C base content of viral scaffolds counting towards a sample if detected based on read mapping, was calculated with an inhouse script (https://github.com/ProbstLab/uBin-helperscripts/blob/master/bin/04_02gc_count.rb). For

Viral dispersal across the air-sea interface

investigating aerosolization and enrichment of VCs in rain over SML and foam, the maximum (sum-normalized) coverage of a virus across an ecosystem, e.g., across all aerosol samples, was considered assuming this value represents the highest possible abundance in that ecosystem. Then coverage ratios were calculated for the pairing rain/foam (R/F), rain/SML (R/S), aerosol/foam (A/F), aerosol/SML (A/S) and foam and SML samples were distinguished between the 5-0.2 μm and virome fraction.

Prokaryotic community composition, binning of MAGs and virus-host interactions

Genes from combined scaffolds from the two assembly steps were predicted using Prodigal v.2.6.3 in meta mode ¹⁰¹, and genes were annotated using DIAMOND ¹⁰⁷ blast against FunTaxDB ⁵⁸. The genes were clustered at 99% identity using CD-HIT ¹⁰⁸, and the centroid of the cluster was used for downstream mapping. Profiling of the prokaryotic community composition was done by read mapping using Bowtie2 of individual samples against scaffolds carrying *rpS3* genes. Taxonomic assignment of *rpS3* genes was performed using USEARCH ¹⁰⁹ against the *rpS3* taxonomy database by Hug, et al. ¹¹⁰ Any unclassified and eukaryotic hits were excluded and coverages were read-sum normalized. For mean relative abundances taxonomic units of the same taxonomy were summed up. Analysis of the Shannon-Wiener Index (alpha diversity) using the 'estimate_richness' function, beta diversity, and NMDS plots based on Bray-Curtis dissimilarity were performed using phyloseq package¹¹¹ in R version v.4.0.3.⁸⁶ within R studio v.1.3.1093. Binning of MAGs was done using MetaBAT 2¹¹² and Maxbin2¹¹³ with aggregating best genomes by DasTool ¹¹⁴ and followed by manual curation in uBin v.0.9.14. ⁵⁸. MAGs were quality-checked using CheckM v.1.1.3 ⁵⁹, and taxonomic assignment was performed with the classify workflow of GTDB-tk v.1.7.0 (database release r202)¹¹⁵. Mapping to individual MAGs was performed with Bowtie2 under allowance of 2 % error rate (3 mismatches) for breadth calculation. Virus-host interactions were inferred from CRISPR-spacer matches and shared k-mer frequency patterns between assembled viruses and host MAGs. At first, CRISPRCasFinder ⁶¹ with -minDR 16 was run on sample assemblies >1 kb and MAGs to find CRISPR consensus DR sequences from arrays with \geq evidence level 3. Consensus DR sequences with 100 % similarity hits to viral scaffolds were removed from further analysis. Then DRs were used in MetaCRASST¹¹⁶ with settings -d 3 -l 60 -c 0.99 -a 0.99 -r to extract CRISPR spacer from the read files of each sample. Spacers were homopolymer and length-filtered (20-60 bp), clustered at 99% identity, BLAST was performed with a BLASTn -short algorithm ¹¹⁷ against the viral scaffolds, and filtered at 80% nucleotide

Viral dispersal across the air-sea interface

similarity. Prokaryotic MAGs (116) were compared and dereplicated using dRep¹¹⁸ at 95% average nucleotide identity. Additional MAGs that were excluded in the dRep process due to low quality in CheckM but had good contamination/completeness scores in uBin and formed their own cluster in the dRep compare mode were additionally considered for k-mer based virus-host linkages. Viral scaffolds were assigned to these MAGs using VirusHostMatcher¹¹⁹ at a d2* threshold of 0.3, as previously performed¹²⁰. Spacer-protospacer interactions and virus-host interactions based on k-mers were visualized using Cytoscape v.3.9¹⁰⁶.

Data availability: ECMWF ERA5 reanalysis data are freely available on <https://cds.climate.copernicus.eu/cdsapp#!/home>. Flow cytometry data have been stored at PANGAEA database and linked to the Integrated Marine Information System (IMIS). Sequencing data for this project are stored in Bioproject PRJNA811790, physical Biosamples SAMN26419398 – SAMN26419459 at NCBI. Sequence raw reads of samples and controls can be found at sequence read archive (SRA) under accession numbers SRR18243813 – SRR18243874. MAGs (116) are stored as Biosamples SAMN26866845 – SAMN26866960 and correspond to accessions JALHYV000000000 – JALHYZ000000000, JALHZA000000000 – JALHZZ000000000, JALIAA000000000 – JALIAZ000000000, JALIBA000000000 – JALIBZ000000000, JALICA000000000 – JALICZ000000000, and JALIDA000000000 – JALIDG000000000. The viral metagenome of 1813 scaffolds is stored within Biosample SAMN27124720, accession number JALJEC000000000. Data will be released upon publication of the manuscript. For further details, please refer to Table S10.

Code availability statement: Links for relevant code are stated in the section “Online methods” or references for bioinformatic tools have been accordingly provided.

Author contributions: JR conceptualized the study, conducted data analysis, wrote the first draft of the manuscript and together with SPE conducted field sampling. JP and AJP developed the viromics pipeline and together with SPE provided bioinformatic assistance. HAG measured and analyzed cell counts and together with MEH determined VLP counts, HW provided data on ice-nucleating particles in seawater samples. AS conducted binning of prokaryotic MAGs, CS and PG calculated backward trajectories. AJP provided supervision, bioinformatic guidance, help with analysis, scripts, and resources. All authors contributed to writing and editing of the final manuscript.

Viral dispersal across the air-sea interface

Acknowledgements: We acknowledge funding for the MIDSEAS (Microbial dispersal from air to sea and snow) project by ASSEMBLE PLUS (European Union’s Horizon 2020 research and innovation program, Grant Agreement No. 730984), and by the German Aerospace Center (DLR) for the project DISPERS (50WB1922). JR received funding by the German Research Foundation (DFG RA3432/1-1). AJP received funding by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen (“Nachwuchsgruppe Dr. Alexander Probst”) and the DFG (PR1603/2-1). MEH received funding within the framework of the PhD research-training group “The Ecology of Molecules” (EcoMol) funded by German Research Foundation within the TRR 51 Collaborative Research Center Roseobacter. HAG was funded the German Research Foundation (DFG), grant/award no. 34509606: Ökologie, Physiologie und Molekularbiologie der Roseobacter-Gruppe: Aufbruch zu einem systembiologischen Verständnis einer global wichtigen Gruppe mariner Bakterien. We like to thank the Sven Lovén Centre for Marine Sciences, Tjärnö, Sweden as part of the University of Gothenburg for hosting JR and SPE during the campaign and providing excellent science support including the use of laboratory and boat facilities. Here, we especially like to thank Anna-Karin Ring, Kerstin Johannesson, Joel White. We further thank Sabrina Eisfeld for laboratory maintenance, and Amelie Assenbaum for technical assistance during INP measurements. We also thank Ken Dreger for server administration, as well as Svenja Parge and Till Bornemann for bioinformatics assistance. The data handling was partially enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Pavlin Mitev at UPPMAX is acknowledged for assistance.

References

- 1 Suttle, C. A. Viruses in the sea. *Nature* **437**, 356-361, doi:10.1038/nature04160 (2005).
- 2 Weinbauer, M. G. & Rassoulzadegan, F. Are viruses driving microbial diversification and diversity? *Environ Microbiol* **6**, 1-11, doi:10.1046/j.1462-2920.2003.00539.x (2004).
- 3 Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* **101**, 11013-11018, doi:10.1073/pnas.0401526101 (2004).
- 4 Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **49**, 781-788 (1999).
- 5 Wu, S. *et al.* Diverse and unique viruses discovered in the surface water of the East China Sea. *BMC Genomics* **21**, 441, doi:10.1186/s12864-020-06861-y (2020).
- 6 Li, Z. *et al.* Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *ISME J*, 1-13, doi:10.1038/s41396-021-00932-y (2021).

Viral dispersal across the air-sea interface

- 7 Upstill-Goddard, R. C. *et al.* Bacterioneuston control of air-water methane exchange determined with a laboratory gas exchange tank. *Global Biogeochem Cy* **17**, doi:10.1029/2003gb002043 (2003).
- 8 Rahlff, J. *et al.* Oxygen profiles across the sea-surface microlayer-effects of diffusion and biological activity. *Front Mar Sci* **6**, doi:10.3389/fmars.2019.00011 (2019).
- 9 Reinthaler, T., Sintes, E. & Herndl, G. J. Dissolved organic matter and bacterial production and respiration in the sea-surface microlayer of the open Atlantic and the western Mediterranean Sea. *Limnol Oceanogr* **53**, 122-136, doi:10.4319/lo.2008.53.1.0122 (2008).
- 10 Hardy, J. T. The sea surface microlayer: Biology, chemistry and anthropogenic enrichment. *Prog Oceanogr* **11**, 307-328, doi:10.1016/0079-6611(82)90001-5 (1982).
- 11 Rahlff, J. *et al.* Blue pigmentation of neustonic copepods benefits exploitation of a prey-rich niche at the air-sea boundary. *Sci Rep* **8**, 11510, doi:10.1038/s41598-018-29869-7 (2018).
- 12 Cunliffe, M. *et al.* Sea surface microlayers: A unified physicochemical and biological perspective of the air–ocean interface. *Progr Oceanogr* **109**, 104-116, doi:10.1016/j.pocean.2012.08.004 (2013).
- 13 Engel, A. *et al.* The ocean's vital skin: Toward an integrated understanding of the sea surface microlayer. *Front Mar Sci* **4**, doi:10.3389/fmars.2017.00165 (2017).
- 14 Naumann, E. Beiträge zur Kenntnis des Teichnannoplanktons, II. Über das Neuston des Süßwassers. *Biol. Centralblatt* **37**, 98-106 (1917).
- 15 Baylor, E. R., Baylor, M. B., Blanchard, D. C., Syzdek, L. D. & Appel, C. Virus transfer from surf to wind. *Science* **198**, 575-580, doi:10.1126/science.918656 (1977).
- 16 Blanchard, D. C. Jet Drop Enrichment of Bacteria, Virus, and Dissolved Organic Material. *Pure Appl Geophys* **116**, 302-308, doi:10.1007/Bf01636887 (1978).
- 17 Bigg, E. K., Leck, C. & Tranvik, L. Particulates of the surface microlayer of open water in the central Arctic Ocean in summer. *Mar Chem* **91**, 131-141, doi:10.1016/j.marchem.2004.06.005 (2004).
- 18 Vaqué, D. *et al.* Enhanced Viral Activity in the Surface Microlayer of the Arctic and Antarctic Oceans. *Microorganisms* **9**, 317, doi:10.3390/microorganisms9020317 (2021).
- 19 Drucker, V. V., Belykh, O. I., Gorshkova, A. S., Bondar, A. A. & Sykilinda, N. N. Autochthonous Bacteriophages in the Microbial Loop Structure of Different Biotopes of Lake Baikal. *Contemp Probl Ecol+* **12**, 143-154, doi:10.1134/s1995425519020045 (2019).
- 20 Rahlff, J. The Virioneuston: A Review on Viral–Bacterial Associations at Air–Water Interfaces. *Viruses* **11**, 191, doi:10.3390/v11020191 (2019).
- 21 Maynard, N. G. Aquatic foams as an ecological habitat. *Z Allg Mikrobiol* **8**, 119-126, doi:10.1002/jobm.19680080205 (1968).
- 22 Rahlff, J. *et al.* Sea foams are ephemeral hotspots for distinctive bacterial communities contrasting sea-surface microlayer and underlying surface water. *FEMS Microbiol Ecol* **97**, doi:10.1093/femsec/fiab035 (2021).
- 23 Jenkinson, I. R. *et al.* The rôles of plankton and neuston microbial organic matter in climate regulation. *J Plankton Res* **43**, 801-821, doi:10.1093/plankt/fbab067 (2021).
- 24 Savun-Hekimoğlu, B. & Gazioğlu, C. Mucilage Problem in the Semi-Enclosed Seas: Recent Outbreak in the Sea of Marmara. *International Journal of Environment and Geoinformatics* **8**, 402-413, doi:10.30897/ijgeo.955739 (2021).
- 25 Shetye, S. S. *et al.* Sea foam-associated pathogenic bacteria along the west coast of India. *Environ Monit Assess* **193**, 27, doi:10.1007/s10661-020-08783-4 (2021).
- 26 Rahlff, J. *et al.* Overlooked diversity of ultramicrobacterial minorities at the air-sea interface. *Atmosphere* **11**, 1214 (2020).

Viral dispersal across the air-sea interface

- 27 Morrow, A. W. Concentration of the virus of foot and mouth disease by foam flotation. *Nature* **222**, 489-490, doi:10.1038/222489a0 (1969).
- 28 Baylor, E., Baylor, M., Blanchard, D. C., Syzdek, L. D. & Appel, C. Virus transfer from surf to wind. *Science* **198**, 575-580, doi:10.1126/science.918656 (1977).
- 29 Gularte, J. S. *et al.* Human mastadenovirus in water, sediment, sea surface microlayer, and bivalve mollusk from southern Brazilian beaches. *Mar Pollut Bull* **142**, 335-349, doi:10.1016/j.marpolbul.2018.12.046 (2019).
- 30 Aller, J. Y., Kuznetsova, M. R., Jahns, C. J. & Kemp, P. F. The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols. *J Aerosol Sci* **36**, 801-812, doi:10.1016/j.jaerosci.2004.10.012 (2005).
- 31 Rastelli, E. *et al.* Transfer of labile organic matter and microbes from the ocean surface to the marine aerosol: an experimental approach. *Sci Rep* **7**, 11475, doi:10.1038/s41598-017-10563-z (2017).
- 32 Sharoni, S. *et al.* Infection of phytoplankton by aerosolized marine viruses. *Proc Natl Acad Sci U S A* **112**, 6643-6647, doi:10.1073/pnas.1423667112 (2015).
- 33 Michaud, J. M. *et al.* Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. *Nat. Commun.* **9**, 2017, doi:10.1038/s41467-018-04409-z (2018).
- 34 Alsante, A. N., Thornton, D. C. O. & Brooks, S. D. Ocean Aerobiology. *Front Microbiol* **12**, 764178, doi:10.3389/fmicb.2021.764178 (2021).
- 35 Adams, M. P. *et al.* Ice nucleation by viruses and their potential for cloud glaciation. *Biogeosciences* **18**, 4431-4444, doi:10.5194/bg-18-4431-2021 (2021).
- 36 Wilbourn, E. K. *et al.* Ice nucleation by marine aerosols over the North Atlantic Ocean in late spring. *J Geophys Res Atmos* **125**, e2019JD030913, doi:10.1029/2019jd030913 (2020).
- 37 Sellegri, K. *et al.* Surface ocean microbiota determine cloud precursors. *Sci Rep* **11**, 281, doi:10.1038/s41598-020-78097-5 (2021).
- 38 Irish, V. E. *et al.* Ice-nucleating particles in Canadian Arctic sea-surface microlayer and bulk seawater. *Atmos Chem Phys* **17**, 10583-10595, doi:10.5194/acp-17-10583-2017 (2017).
- 39 Irish, V. E. *et al.* Revisiting properties and concentrations of ice-nucleating particles in the sea surface microlayer and bulk seawater in the Canadian Arctic during summer. *Atmos Chem Phys* **19**, 7775-7787, doi:10.5194/acp-19-7775-2019 (2019).
- 40 Gong, X. D. *et al.* Characterization of aerosol particles at Cabo Verde close to sea level and at the cloud level - Part 2: Ice-nucleating particles in air, cloud and seawater. *Atmos Chem Phys* **20**, 1451-1468, doi:10.5194/acp-20-1451-2020 (2020).
- 41 Wilson, T. W. *et al.* A marine biogenic source of atmospheric ice-nucleating particles. *Nature* **525**, 234-238, doi:10.1038/nature14986 (2015).
- 42 Castello, J. D. Detection of Infectious Tomato Mosaic Tobamovirus in Fog and Clouds. *Phytopathology* **85**, 1409-1412, doi:10.1094/Phyto-85-1409 (1995).
- 43 Amato, P. *et al.* Metatranscriptomic exploration of microbial functioning in clouds. *Sci Rep* **9**, 4383, doi:10.1038/s41598-019-41032-4 (2019).
- 44 Amato, P. Clouds provide atmospheric oases for microbes. *Microbe* **7**, 119-123 (2012).
- 45 Amato, P. *et al.* Microbial population in cloud water at the Puy de Dome: Implications for the chemistry of clouds. *Atmos Environ* **39**, 4143-4153, doi:10.1016/j.atmosenv.2005.04.002 (2005).
- 46 Woo, C. & Yamamoto, N. Falling bacterial communities from the atmosphere. *Environ Microbiome* **15**, 22, doi:10.1186/s40793-020-00369-4 (2020).

Viral dispersal across the air-sea interface

- 47 Christner, B. C. *et al.* Geographic, seasonal, and precipitation chemistry influence on the abundance and activity of biological ice nucleators in rain and snow. *Proc Natl Acad Sci U S A* **105**, 18854-18859, doi:10.1073/pnas.0809816105 (2008).
- 48 Morris, C. E., Georgakopoulos, D. G. & Sands, D. C. in *Journal de Physique IV (Proceedings)*. 87-103 (EDP sciences).
- 49 Haramoto, E. *et al.* Effects of rainfall on the occurrence of human adenoviruses, total coliforms, and Escherichia coli in seawater. *Water Sci Technol* **54**, 225-230, doi:10.2166/wst.2006.473 (2006).
- 50 Inoue, K., Asami, T., Shibata, T., Furumai, H. & Katayama, H. Spatial and temporal profiles of enteric viruses in the coastal waters of Tokyo Bay during and after a series of rainfall events. *Sci Total Environ* **727**, 138502, doi:10.1016/j.scitotenv.2020.138502 (2020).
- 51 Victoria, M. *et al.* Gastroenteric virus dissemination and influence of rainfall events in urban beaches in Brazil. *J Appl Microbiol* **117**, 1210-1218, doi:10.1111/jam.12592 (2014).
- 52 Reche, I., D'Orta, G., Mladenov, N., Winget, D. M. & Suttle, C. A. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* **12**, 1154-1162, doi:10.1038/s41396-017-0042-4 (2018).
- 53 Péguilhan, R. *et al.* Rainfalls sprinkle cloud bacterial diversity while scavenging biomass. *FEMS Microbiol Ecol* **97**, fiab144, doi:10.1093/femsec/fiab144 (2021).
- 54 Woods, A. C., Walker, J. R., Jackson, C. D. & Labonte, J. M. Record-Breaking Rain Event Altered Estuarine Viral Assemblages. *Microorganisms* **10**, 729, doi:10.3390/microorganisms10040729 (2022).
- 55 Green, J. C., Rahman, F., Saxton, M. A. & Williamson, K. E. Quantifying aquatic viral community change associated with stormwater runoff in a wet retention pond using metagenomic time series data. *Aquat Microb Ecol* **81**, 19-35, doi:10.3354/ame01856 (2018).
- 56 Williamson, K. E., Harris, J. V., Green, J. C., Rahman, F. & Chambers, R. M. Stormwater runoff drives viral community composition changes in inland freshwaters. *Front Microbiol* **5**, 105, doi:10.3389/fmicb.2014.00105 (2014).
- 57 Pummer, B. G. *et al.* Ice nucleation by water-soluble macromolecules. *Atmos Chem Phys* **15**, 4077-4091, doi:10.5194/acp-15-4077-2015 (2015).
- 58 Bornemann, T. L. V., Esser, S. P., Stach, L., Burg, T. & Probst, A. J. uBin – a manual refining tool for metagenomic bins designed for educational purposes. *bioRxiv*, doi:10.1101/2020.07.15.204776 (2020).
- 59 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
- 60 Bolduc, B. *et al.* vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* **5**, e3243, doi:10.7717/peerj.3243 (2017).
- 61 Couvin, D. *et al.* CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* **46**, W246-W251, doi:10.1093/nar/gky425 (2018).
- 62 Després, V. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B: Chem Phys Meteorol* **64**, 15598 (2012).
- 63 Rahlff, J., Bornemann, T. L. V., Lopatina, A., Severinov, K. & Probst, A. J. Host-Associated Phages Disperse across the Extraterrestrial Analogue Antarctica. *Appl Environ Microbiol* **88**, e0031522, doi:10.1128/aem.00315-22 (2022).

Viral dispersal across the air-sea interface

- 64 Archer, S. D. *et al.* Air mass source determines airborne microbial diversity at the ocean–atmosphere interface of the Great Barrier Reef marine ecosystem. *ISME J.* **14**, 871-876 (2020).
- 65 van Pinxteren, M. *et al.* High number concentrations of transparent exopolymer particles in ambient aerosol particles and cloud water – a case study at the tropical Atlantic Ocean. *Atmos Chem Phys* **22**, 5725-5742, doi:10.5194/acp-22-5725-2022 (2022).
- 66 Brussaard, C. P. D., Mari, X., Van Bleijswijk, J. D. L. & Veldhuis, M. J. W. A mesocosm study of *Phaeocystis globosa* (Prymnesiophyceae) population dynamics - II. Significance for the microbial community. *Harmful Algae* **4**, 875-893, doi:10.1016/j.hal.2004.12.012 (2005).
- 67 Yamada, Y., Guillemette, R., Baudoux, A. C., Patel, N. & Azam, F. Viral Attachment to Biotic and Abiotic Surfaces in Seawater. *Appl Environ Microbiol* **86**, doi:10.1128/AEM.01687-19 (2020).
- 68 Rahlff, J. *et al.* High wind speeds prevent formation of a distinct bacterioneuston community in the sea-surface microlayer. *FEMS Microbiol Ecol* **93**, doi:10.1093/femsec/fix041 (2017).
- 69 Bellas, C. M., Schroeder, D. C., Edwards, A., Barker, G. & Anesio, A. M. Flexible genes establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems. *Nat Commun* **11**, 4403, doi:10.1038/s41467-020-18236-8 (2020).
- 70 Beall, C. M. *et al.* Cultivable, halotolerant ice nucleating bacteria and fungi in coastal precipitation. *Atmos Chem Phys* 1-33, doi:10.5194/acp-2020-1229 (2021).
- 71 Singer, C. E. & Ames, B. N. Sunlight ultraviolet and bacterial DNA base ratios. *Science* **170**, 822-825, doi:10.1126/science.170.3960.822 (1970).
- 72 Ellington, A. J., Bryan, N. C., Christner, B. C. & Reisch, C. R. Draft Genome Sequences of Actinobacterial and Betaproteobacterial Strains Isolated from the Stratosphere. *Microbiol Resour Announc* **10**, e0100921, doi:10.1128/MRA.01009-21 (2021).
- 73 Wiklund, J. A. *et al.* Widespread atmospheric Tellurium contamination in industrial and remote regions of Canada. *Environ Sci Technol* **52**, 6137-6145, doi:10.1021/acs.est.7b06242 (2018).
- 74 Wrangle, A. L. *et al.* Importance of plasticity and local adaptation for coping with changing salinity in coastal areas: a test case with barnacles in the Baltic Sea. *BMC Evol Biol* **14**, 156, doi:10.1186/1471-2148-14-156 (2014).
- 75 Harvey, G. W. & Burzell, L. A. A simple microlayer method for small samples *Limnol Oceanogr* **17**, 156-157, doi:10.4319/lo.1972.17.1.0156 (1972).
- 76 John, S. G. *et al.* A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* **3**, 195-202, doi:10.1111/j.1758-2229.2010.00208.x (2011).
- 77 Rolph, G., Stein, A. & Stunder, B. Real-time Environmental Applications and Display sYstem: READY. *Environmental Modelling & Software* **95**, 210-228, doi:10.1016/j.envsoft.2017.06.025 (2017).
- 78 Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999-2049, doi:10.1002/qj.3803 (2020).
- 79 Scarchilli, C., Frezzotti, M. & Ruti, P. M. Snow precipitation at four ice core sites in East Antarctica: provenance, seasonality and blocking factors. *Climate Dynamics* **37**, 2107-2125, doi:10.1007/s00382-010-0946-4 (2010).
- 80 Becagli, S. *et al.* Biogenic aerosol in central East Antarctic Plateau as a proxy for the ocean-atmosphere interaction in the Southern Ocean. *Sci Total Environ* **810**, 151285, doi:10.1016/j.scitotenv.2021.151285 (2022).

Viral dispersal across the air-sea interface

- 81 Marie, D., Simon, N., Guillou, L., Partensky, F. & Vaultot, D. in *In Living Color. Springer Lab Manuals* (eds R.A. Diamond & S. Demaggio) Ch. Chapter 34, 421-454 (Springer, 2000).
- 82 Giebel, H. A. *et al.* Microbial growth and organic matter cycling in the Pacific Ocean along a latitudinal transect between subarctic and subantarctic waters. *Front Mar Sci* **8**, 764383, doi:10.3389/fmars.2021.764383 (2021).
- 83 Giebel, H. A., Wolterink, M., Brinkhoff, T. & Simon, M. Complementary energy acquisition via aerobic anoxygenic photosynthesis and carbon monoxide oxidation by *Planktomarina temperata* of the *Roseobacter* group. *FEMS Microbiol Ecol* **95**, fiz050, doi:10.1093/femsec/fiz050 (2019).
- 84 Brussaard, C. P. D. *et al.* Manual of aquatic viral ecology. *Am Soc Limnol Oceanogr*, 102-109 (2010).
- 85 Wex, H. *et al.* Annual variability of ice-nucleating particle concentrations at different Arctic locations. *Atmos Chem Phys* **19**, 5293-5311, doi:10.5194/acp-19-5293-2019 (2019).
- 86 Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* (2017).
- 87 Racine, J. S. RStudio: a platform-independent IDE for R and Sweave. (2012).
- 88 Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930, doi:10.1111/j.1654-1103.2003.tb02228.x (2003).
- 89 Joshi, N. & Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. (2011).
- 90 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824-834, doi:10.1101/gr.213959.116 (2017).
- 91 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 92 Molder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res* **10**, 33, doi:10.12688/f1000research.29032.2 (2021).
- 93 Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* **36**, 4126-4129, doi:10.1093/bioinformatics/btaa490 (2020).
- 94 Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90, doi:10.1186/s40168-020-00867-0 (2020).
- 95 Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985, doi:10.7717/peerj.985 (2015).
- 96 Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*, doi:10.1038/s41587-020-00774-7 (2020).
- 97 Moraru, C., Varsani, A. & Kropinski, A. M. VIRIDIC-A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* **12**, doi:10.3390/v12111268 (2020).
- 98 Nilsson, E., Bayfield, O. W., Lundin, D., Antson, A. A. & Holmfeldt, K. Diversity and host interactions among virulent and temperate Baltic Sea *Flavobacterium* phages. *Viruses* **12**, 158, doi:10.3390/v12020158 (2020).
- 99 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
- 100 Roux, S., Emerson, J. B., Elie-Fadrosh, E. A. & Sullivan, M. B. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817, doi:10.7717/peerj.3817 (2017).

Viral dispersal across the air-sea interface

- 101 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
identification. *BMC Bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).
- 102 Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation
of microbiome function. *Nucleic Acids Res* **48**, 8883-8900, doi:10.1093/nar/gkaa621
(2020).
- 103 Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison
visualizer. *Bioinformatics* **27**, 1009-1010, doi:10.1093/bioinformatics/btr039 (2011).
- 104 Cook, R. *et al.* INfrastructure for a PHAge REference Database: Identification of
Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *Phage* **2**,
214-223, doi:10.1089/phage.2021.0007 (2021).
- 105 Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes
is enabled by gene-sharing networks. *Nat Biotechnol* **37**, 632-639,
doi:10.1038/s41587-019-0100-8 (2019).
- 106 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of
biomolecular interaction networks. *Genome Res* **13**, 2498-2504,
doi:10.1101/gr.1239303 (2003).
- 107 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using
DIAMOND. *Nature Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).
- 108 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
generation sequencing data. *Bioinformatics* **28**, 3150-3152,
doi:10.1093/bioinformatics/bts565 (2012).
- 109 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
Bioinformatics **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).
- 110 Hug, L. A. *et al.* A new view of the tree of life. *Nat Microbiol* **1**, 16048,
doi:10.1038/nmicrobiol.2016.48 (2016).
- 111 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive
analysis and graphics of microbiome census data. *PLoS One* **8**, e61217,
doi:10.1371/journal.pone.0061217 (2013).
- 112 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient
genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359,
doi:10.7717/peerj.7359 (2019).
- 113 Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning
algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**,
605-607, doi:10.1093/bioinformatics/btv638 (2016).
- 114 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication,
aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843, doi:10.1038/s41564-018-
0171-1 (2018).
- 115 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to
classify genomes with the Genome Taxonomy Database. *Bioinformatics*,
doi:10.1093/bioinformatics/btz848 (2019).
- 116 Moller, A. G. & Liang, C. MetaCRIST: reference-guided extraction of CRISPR
spacers from unassembled metagenomes. *PeerJ* **5**, e3788, doi:10.7717/peerj.3788
(2017).
- 117 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local
alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2
(1990).
- 118 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and
accurate genomic comparisons that enables improved genome recovery from
metagenomes through de-replication. *ISME J* **11**, 2864-2868,
doi:10.1038/ismej.2017.126 (2017).

Viral dispersal across the air-sea interface

- 119 Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A. & Sun, F. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res* **45**, 39-53, doi:10.1093/nar/gkw1002 (2017).
- 120 Hwang, Y., Rahlff, J., Schulze-Makuch, D., Schloter, M. & Probst, A. J. Diverse viruses carrying genes for microbial extremotolerance in the Atacama Desert hyperarid soil. *mSystems* **6**, e00385-00321, doi:10.1128/mSystems.00385-21 (2021).