



HAL
open science

Dynamics of the viral community on the cheese surface during maturation and persistence across production years

Thomas Paillet, Quentin Lamy-Besnier, Clarisse Figueroa, Marie-Agnès Petit,
Eric Dugat-Bony

► To cite this version:

Thomas Paillet, Quentin Lamy-Besnier, Clarisse Figueroa, Marie-Agnès Petit, Eric Dugat-Bony. Dynamics of the viral community on the cheese surface during maturation and persistence across production years. 2024. hal-04416953

HAL Id: hal-04416953

<https://hal.inrae.fr/hal-04416953v1>

Preprint submitted on 25 Jan 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Dynamics of the viral community on the cheese surface during maturation and persistence
2 across production years

3

4 Thomas Paillet,^a Quentin Lamy-Besnier,^b Clarisse Figueroa,^{a*} Marie-Agnès Petit,^b Eric
5 Dugat-Bony^{a#}

6 ^aUniversité Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 91120 Palaiseau, France

7 ^bUniversité Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78352 Jouy-en-Josas,
8 France

9

10 Running Head: Viral dynamics and persistence on the cheese surface

11

12 #Address correspondence to Eric Dugat-Bony, eric.dugat-bony@inrae.fr.

13 *Present address: Université de Paris Cité, INSERM, IAME, UMR 1137, 75018 Paris, France

14 **ABSTRACT**

15 The surface of smear-ripened cheeses constitutes a dynamic microbial ecosystem resulting from the
16 successive development of different microbial groups. Recent studies indicate that a viral community,
17 mainly composed of bacteriophages, coexists with cellular microorganisms in this ecosystem, but its
18 ecological significance remains to be elucidated. In this work, we studied a French smear-ripened
19 cheese by both viral metagenomics and 16S metabarcoding approaches to assess both the dynamics of
20 phages and bacterial communities on the cheese surface during the ripening period, and their
21 persistence in ready-to-eat cheeses over the years of production. We observed a clear transition of the
22 phage community structure during ripening with a decreased relative abundance of viral species
23 (vOTUs) associated with *Lactococcus* phages, which were replaced by vOTUs associated with phages
24 infecting ripening bacteria such as *Brevibacterium*, *Glutamicibacter*, *Pseudoalteromonas* and *Vibrio*.
25 The dynamics of the phage community was strongly associated with bacterial successions observed on
26 the cheese surface. Finally, a core of abundant vOTUs were systematically detected in ready-to-eat
27 cheeses produced at different dates spanning more than 4 years of production, indicating long-term
28 persistence of the main phages in the cheese production environment. Together, these findings offer
29 novel perspectives on the ecology of bacteriophages in smear-ripened cheese and emphasize the
30 significance of incorporating bacteriophages in the microbial ecology studies of fermented foods.

31 **IMPORTANCE**

32 Smear-ripened cheeses are microbial ecosystems made up of various microorganisms including
33 bacteria, yeasts and also viruses such as bacteriophages, which infect and regulate bacterial
34 populations. In this work, a French smear-ripened cheese was used to study how these viruses and
35 bacteria interact over time and during cheese production. It revealed that the composition of the
36 bacteriophage community shifts during the ripening process, aligning with the bacterial successions
37 observed on the cheese surface between lactic acid bacteria and ripening bacteria. Additionally, the
38 vast majority of these bacteriophages were found consistently in cheese products made over a 4-years
39 period, showing that they represent a persistent component of the cheese-making environment. This
40 research highlights the importance of considering these bacteriophages when studying the microbial
41 life of fermented foods like cheese.

42

43 INTRODUCTION

44 Due to their unique ripening process, involving frequent washes with saline and/or alcoholic solutions,
45 smear-ripened cheeses host a peculiar and diverse microbiota composed of lactic acid bacteria (LAB,
46 mainly starter cultures added at the beginning of the process for milk acidification), yeasts and salt-
47 tolerant bacteria belonging to the Actinomycetota, Bacillota and Pseudomonadota phyla (1). The
48 surface microbiota of smear-ripened cheeses is considered to be responsible for the typical flavour and
49 organoleptic properties of this type of cheese (2). From the past two decades, numerous studies have
50 been conducted to describe the composition of this microbiota using isolation-based methods (3, 4),
51 molecular fingerprinting (5–8), and more recently amplicon-based metagenomics also commonly
52 referred to as metabarcoding (9–12).

53 Time series studies also enabled to reveal microbial successions occurring on the surface of smear-
54 ripened cheeses during the maturation process (8, 13). Lactic acid bacteria (LAB), usually originating
55 from starter cultures, grow first in the milk and represent the dominant microorganisms in the curd.
56 Yeasts, *e.g. Debaryomyces hansenii* and *Geotrichum candidum*, which exhibit acid tolerance and
57 metabolize lactate, subsequently colonize the cheese surface, resulting in its deacidification. With the
58 pH increase, the establishment of a diverse bacterial community is progressively observed. The most
59 common bacterial taxa detected at the end of ripening on smear-ripened cheese belong to coryneform
60 bacteria (*e.g.* species of the *Glutamicibacter*, *Brevibacterium*, *Corynebacterium* or *Brachybacterium*
61 genera), *Staphylococcus* species and halophilic or halotolerant gram-negative bacteria (*e.g.* species of
62 the *Psychrobacter*, *Halomonas*, *Pseudoalteromonas*, *Hafnia*, *Vibrio*, *Pseudomonas* or *Proteus* genera)
63 (14).

64 Microbial interactions are key biotic factors determining the structure and functioning of cheese
65 microbial ecosystems and, ultimately, affecting cheese quality and safety (15, 16). In many natural
66 ecosystems, bacteriophage infections shapes the composition of bacterial populations (17) and recent
67 work suggests the same applies to fermented foods (18, 19). In the dairy industry, the impact of LAB
68 phages is well documented because their lytic activity can disturb the milk acidification step, causing

69 delay in the production and even total loss of production (20, 21). Consequently, the most studied
70 dairy phages are *Lactococcus*, *Streptococcus*, *Lactobacillus* and *Leuconostoc* phages, infecting the
71 main starter cultures (22). Viral metagenomics has been recently employed to characterize the
72 bacteriophage communities in dairy samples, including whey (23) and cheeses (24, 25). These
73 investigations have demonstrated that such viral communities are not restricted to LAB phages, but
74 encompass a diverse array of phages that could potentially also infect non-inoculated and ripening
75 bacteria during cheese production.

76 However, only few studies report the isolation of such virulent phages from cheese infecting ripening
77 bacteria, *i.e.* *Propionibacterium freudeunreichii* and *Brevibacterium aurantiacum* (26–28). We also
78 isolated in a previous work five new virulent phages, targeting *Glutamicibacter arilaitensis*,
79 *Brevibacterium aurantiacum*, *Psychrobacter aquimaris* and *Leuconostoc falkenbergense*, from the
80 surface of a French smear-ripened cheese suggesting that predation is likely to occur on such
81 ecosystems for most of the dominant bacteria (29). However, little is known about their ecology. In
82 this study, our aim was to enhance our understanding of the temporal distribution of bacteriophages
83 and their bacterial hosts on the surface of a French smear-ripened cheese across two distinct time
84 scales. Initially, cheeses were obtained directly from the production facility at five distinct ripening
85 stages over 28 days, to analyze phage dynamics throughout a production cycle (Figure 1, dynamic
86 study). Subsequently, ready-to-eat cheeses of the same brand and variety were sampled in 2017, 2019,
87 and 2022 to assess the long-term persistence of phages in this ecosystem (persistence study).

88

89 **RESULTS**

90 **Composition of the cheese surface virome.**

91 The sequence assembly obtained from all the studied samples (15 from the dynamic study and 9 from
92 the persistence study) led to the production of a metavirome composed of 331 vOTUs >2 Kb (Table
93 1). The vast majority of these contigs (284) were detected in samples from both the dynamic and
94 persistence studies.

95 The most abundant phages detected in the dataset were identified as *Lactococcus* phages from the 949
96 and 936 groups (*Audreyjarvisvirus* and *Skunavirus* genera, respectively) and *Glutamicibacter* phage

97 Montesquieu (Table 2). Of importance, the five virulent phages previously isolated from the same
98 cheese, namely Montesquieu, Voltaire, Rousseau, D'Alembert and Diderot were also detected in this
99 metagenomics survey. Remarkably, the vOTUs with no match (BLAT identity \times coverage $<$ 30%) to
100 any known dairy phages (Table S1) represented only 3% of the relative abundance in the dataset,
101 showing that most of the dominant phages present in this ecosystem have already close-relatives that
102 have been isolated and characterized.

103 **The composition of the viral community present on the cheese surface evolves through the**
104 **ripening process.**

105 The effect of each production step on the cheese virome composition was assessed by computing
106 Bray-Curtis dissimilarity (BC). The PERMANOVA test indicated a significant effect of this variable
107 (p -value = 0.001, $R^2=0.658$) and the principal coordinate analysis revealed that the first axis clearly
108 discriminates samples from the first 3 washes (NaCl) and samples from the two last washes (NaCl +
109 alcoholic liquor) (Figure 2A). The second axis helps discriminating samples from the third wash (W3)
110 and samples from the two first (W1 and 2). Regarding the viral diversity, as estimated by the Shannon
111 index, a slight decrease was noted from W3 onwards. However, this decrease was not statistically
112 significant, suggesting that there were no major alterations of the viral diversity related to the
113 production step ($p>0.05$, Kruskal-Wallis test; Figure 2B).

114 In order to visualize the abundance of the different vOTUs in the samples, we kept only those with a
115 normalized relative abundance above 5×10^{-5} in average (168 contigs) and represented their distribution
116 across samples through a heatmap (Figure 2C). Three blocks of vOTUs were detected. The first one
117 (top of the figure) was characterized by contigs whose abundance do not vary with the ripening and
118 that corresponded mainly to virulent *Lactococcus* phages belonging to the 936 group (*Skunavirus*
119 genus). The second block (bottom of the figure) contained contigs whose abundance decreased with
120 ripening and that corresponded to other *Lactococcus* phages, close to the P335 group containing both
121 temperate and virulent members (here qualified as ex-temperate phages). Finally, the third block
122 (middle of the figure) was essentially composed of a few vOTUs that increased in relative abundance
123 with ripening. These mainly corresponded to phages targeting ripening bacteria such as
124 *Brevibacterium*, *Glutamicibacter*, *Pseudoalteromonas* and *Vibrio*, and non-starter lactic acid bacteria

125 (NSLAB) such as *Leuconostoc*. These few vOTUs correspond to uncharacterized phages, except
126 *Glutamicibacter* phage Montesquieu, *Brevibacterium* phage Rousseau and *Leuconostoc* phage Diderot
127 which we previously isolated from the same type of cheese.

128 As observed by the principal coordinate analysis of the Bray-Curtis dissimilarity (Figure 2A), the
129 virome composition was similar between samples from W1 and W2, and between samples from W4
130 and W5 indicating the presence of two main viral communities, according to these two ripening
131 stages. Samples from W3 exhibited a composition in between samples from W1-2 and W4-5 reflecting
132 a transition stage captured in those samples.

133 **Phage community shift during ripening follows changes in bacterial composition.**

134 We then applied the DESeq2 method to identify differentially abundant viral contigs between the two
135 stable stages represented by W1-2 and W4-5 samples (Figure 3A). Interestingly, two groups emerged,
136 with very readable outlines: vOTUs corresponding to phages infecting starter cultures, such as
137 *Lactococcus* and *Streptococcus* phages, had a negative log₂ fold change meaning their relative
138 abundances significantly decreased during the ripening process. Conversely, vOTUs of virulent
139 phages targeting NSLAB and ripening bacteria (e.g. *Brevibacterium*, *Glutamicibacter*,
140 *Pseudoalteromonas* and *Vibrio*) were significantly more abundant in W4-5 samples compared to W1-2
141 samples (positive log₂ fold change) reflecting a higher population level for such phages on the cheese
142 surface in later ripening stages. Among them were two vOTUs with high similarity to the genomes of
143 *Glutamicibacter* phage Montesquieu (2017-3_NODE2; 47,703 bp; 100% identity × coverage by
144 BLAT), and *Brevibacterium* phage Rousseau (2022-5_NODE4; 41,077 bp; 54% identity × coverage
145 by BLAT but 98% identity at the nucleotidic level over a portion of >9 kb). In this group, we also
146 detected a vOTU partially related to *Brevibacterium* phage AGM1 (SA2-0_NODE10; 37148 bp; 3.5%
147 identity × coverage by BLAT but 88% identity at the nucleotidic level over a portion of 1324 nt),
148 which was isolated from a Canadian washed-rind cheese.

149 The composition of the bacterial community of the cheese surface also varied through the ripening
150 process (Figure 3B). As for the virome composition, we observed a clear transition in the bacterial
151 community structure from W3 onwards. *Lactococcus* was the dominant genus in samples from W1
152 and W2, and was progressively replaced by typical surface aerobic bacteria such as members of

153 *Psychrobacter*, *Vibrio*, *Glutamicibacter* and *Pseudoalteromonas* genera. Based on plate counts (Figure
154 3 C-E), this results should be mainly attributed to the growth of aerobic bacteria (~ 2 logs increase
155 between W1 and W5, from ~ 10^8 to ~ 10^{10} CFU/g) since lactic acid remained stable over time (~ 10^8
156 CFU/g) during the whole kinetic. This shift indicates that changes observed on the phage community
157 structure is strongly associated with bacterial successions on the cheese surface.

158 **The dominant fraction of the cheese virome persists across production years.**

159 The effect of the production year (2017, 2019 and 2022) on the composition of the virome of the
160 cheese surface (ready-to-eat cheeses, meaning after packaging and storage) was assessed by
161 computing Bray-Curtis dissimilarity (Figure 4A). The PERMANOVA test indicated a significant
162 effect of this factor (p-value of 0.002, $R^2 = 0.814$) suggesting structural variations of the phage
163 community across production years. The principal coordinate analysis revealed that the first axis
164 clearly discriminates 2019 samples from 2017 and 2022 samples. The second axis helps discriminating
165 samples from years 2017 and 2022.

166 The heatmap visualization showed that the vast majority of the most abundant contigs were shared
167 among the 3 production years and that only a few low-abundant contigs were detected specifically in
168 one or two production years (blue zones on the graph) (Figure 4C). This result is congruent with the
169 stability of the bacterial community composition over the three sampling campaigns (Figure 4B).

170 We then evaluated the proportion of vOTUs that were shared across production year. When using the
171 complete dataset (no filtering on relative abundance, Figure 4D), we observed that 56.6% of the 297
172 vOTUs present in at least one of the three years were shared in all production years (75.4% in two
173 different production years). Interestingly, samples from year 2022 had a higher number of unique
174 vOTUs (48, 32.1% of the total) than samples from years 2017 and 2019 (3 and 11, respectively). We
175 next applied the same analysis only on the most abundant vOTUs (average normalized relative
176 abundance $> 5 \times 10^{-5}$, 99 vOTUs in total) (Figure 4E). The vast majority of the vOTUs, *i.e.* 89.9%,
177 were shared among the 3 productions years and this value rose to 97% when considering only two
178 different years. This result indicates that the cheese surface virome was mostly stable in terms of
179 composition (presence/absence) and that dominant phages persist across productions years. The
180 disparity between production years, as detected by beta-diversity analysis, may therefore primarily be

181 attributed to the presence or absence of phages in low abundance, and fluctuations in the relative
182 abundance of dominant phages.

183

184 **DISCUSSION**

185 Recent studies of cheese samples using viral metagenomics (24, 25) or exploration of cheese microbial
186 metagenomes (30) have revealed that the cheese environment harbours a diverse bacteriophage
187 community whose targets go beyond lactic acid starter cultures. The isolation of a few representatives
188 of these non-starter phages (27–29) suggests that phage activity occurs in this ecosystem, raising the
189 question of their overall impact on microbial successions during the ripening process. Recently, the
190 deliberate addition of one of those phages, *Brevibacterium* phage AGM9, was proven to slow down
191 the development of the orange rind color in a model system mimicking a smear-ripened cheese (31). In
192 the present study, we described the viral community of a French smear ripened cheese over time, by
193 combining two timescales: the 28-days long ripening process, as well as from ready-to-eat cheeses
194 spanning four years of production. The virome was predominantly composed of vOTUs associated
195 with a variety of *Lactococcus* phages as well as the *Glutamicibacter* phage Montesquieu, a virulent
196 phage we had previously isolated from the same cheese variety (29). This phage targets the ripening
197 bacterium *Glutamicibacter arilaitensis*. Several other phages were detected at sub-dominant levels.
198 Among them, only a few, such as the *Brevibacterium* phage Rousseau, *Leuconostoc* phage Diderot,
199 *Psychrobacter* phage d’Alembert, *Glutamicibacter* phage Voltaire, and a novel vOTU displaying
200 minimal sequence homology to *Brevibacterium* phage AGM1, have been previously documented in
201 cheese. The diversity of these sub-dominant phages is therefore not yet completely sampled and, in
202 particular, underscores the necessity to isolate a more comprehensive collection of phages from this
203 ecosystem. Notably, we identified vOTUs that partially aligned with genomic sequences from phages
204 infecting halotolerant bacteria (e.g., *Pseudoalteromonas*, *Halomonas*, *Vibrio*, and *Proteus* species).
205 Even though these bacteria are not intentionally inoculated into smear-ripened cheese, they tend to
206 dominate by the end of the ripening process (14). Their phages should therefore be looked for in future
207 isolation initiatives.

208 The dynamics of bacterial and fungal communities has been extensively studied in a wide diversity of
209 cheese products during ripening (13, 32–36), enabling to accurately describe cellular successions
210 occurring in cheese production processes (16). In this study, we present the first analysis of viral
211 dynamics throughout cheese ripening. Similar to the observed transitions in bacteria and fungi, there
212 was a distinct shift in viral composition over the course of the ripening process. Some vOTUs
213 associated with LAB starter-phages, primarily targeting *Lactococcus lactis*, were progressively
214 replaced by vOTUs specific to phages that infect ripening bacteria. When comparing phage to
215 bacterial dynamics, there was a notable relationship between phage trajectories and bacterial
216 successions on the cheese surface. Specifically, the relative abundance of phages exhibited a
217 concomitant increase with the relative abundance of their predicted bacterial hosts. This findings is of
218 importance since the increase in the relative abundance of a specific vOTU within a virome can be
219 interpreted as indicative of the active replication of the corresponding phage and further suggests the
220 presence of a predator-prey interaction within the investigated ecosystem.

221 Surprisingly, despite the changes in bacterial and viral communities' composition during the
222 production process, the relative abundance of some vOTUs remained very stable during ripening.
223 They mainly corresponded to virulent *Lactococcus* phages, belonging to the *Skunavirus* genus
224 (formerly 936 group). Among *Lactococcus* phages, this genus is by far the most frequently detected in
225 the dairy industry (37). Given that *Lactococcus lactis* predominantly proliferates during milk
226 acidification and maintains consistent concentration levels throughout the ripening period in this type
227 of cheese (13), we theorize that the maintenance of *Skunavirus* is indicative of the stability of phage
228 particles produced at the onset of cheese maturation. The remarkable stability of *Skunavirus* particles
229 in comparison to other *Lactococcus* phage groups, especially the P335 group, has been reported earlier
230 (38, 39). In contrast, we suggest that the relative decline observed in our experiment for certain
231 *Lactococcus* phages, specifically those affiliated to the P335 group, denotes their temporal instability.
232 Nevertheless, a more comprehensive examination of this phenomenon warrants dedicated
233 investigations.

234 Cheese microbial communities of washed-rind cheese are generally dominated by environmental
235 microorganisms detected in processing environments, the so-called “house” microbiota (40), which is
236 specific to each production facility and provides a microbial signature distinguishing cheeses
237 belonging to the same variety but manufactured by distinct producers (10). Recently, the analysis of
238 several Quebec’s terroir cheeses revealed that the dominant microorganisms remain stable from year
239 to year, which could be linked to typical manufacturing practices and consistency in the use of starter
240 and ripening cultures by cheesemakers (41). Here, we describe a similar observation for
241 bacteriophages. We indeed identified a core-virome composed of a large proportion of the most
242 abundant vOTUs, consistently detected across four production years. Previous work on an undefined
243 starter culture used for the production of a Swiss-type cheese, propagated for decades in the same
244 dairy environment, revealed that phages and bacteria stably coexist over time in this system and
245 suggests that this may contribute to the stable maintenance of the cheese starter culture over years
246 (42). The same may apply for phages and bacteria on the surface of smear-ripened cheese since both
247 are contaminating the cheese production environment, and are therefore likely to repeatedly
248 contaminate cheese from one production cycle to another (29, 40).

249 In conclusion, the observed dynamics of the cheese virome throughout the ripening process, coupled
250 with its relative persistence across production years, support an important role of bacteriophages in the
251 cheese microbial ecosystem. Recognizing this biotic factor is essential for a comprehensive
252 understanding of microbial successions during milk fermentation. Moreover, this knowledge may
253 offer cheesemakers novel avenues to refine and control their production processes.

254

255 **MATERIAL AND METHODS**

256 **Cheese samples.**

257 (i) Dynamic study design: French smear-ripened cheeses, all from the same production batch, were
258 collected directly from the cheese plant at five different stages during the ripening process (Figure 1).
259 These stages, labeled W1 to W5, correspond to distinct washing steps. The initial three washes utilized
260 a NaCl solution, while the final two employed a NaCl solution supplemented with increasing

261 concentrations of alcoholic liquor. At each stage, three distinct cheeses were sampled and designated
262 as replicates A, B, and C. Cheeses were immediately stored at 4°C after sampling and processed
263 within 48h.

264 (ii) Persistence study design: ready-to-consume cheeses of the same type and same brand as the
265 previously described cheeses were purchased in a local supermarket in December 2017, November
266 2019 and February 2022, spanning >4 years of production. Three different cheeses, with the same
267 production date, were sampled each year and used as replicates. Cheeses were immediately stored at
268 4°C after sampling and processed within 48h.

269 For both dynamic and persistence studies, cheese samples were analysed immediately after reception
270 at the lab. Using sterile knives, the rind, approximately 2-3 mm thick, was carefully separated from the
271 core. It was then blended and processed for microbial counts, viral DNA extraction for metavirome
272 analysis, and microbial DNA extraction for extensive amplicon sequencing targeting the 16S rRNA
273 gene, which will be subsequently referred to as 16S metabarcoding.

274 **Microbiological analysis.**

275 Bacteria and yeasts were enumerated by plating serial dilutions (10^{-1} to 10^{-7}) of one gram of cheese
276 rind mixed in 9 mL of physiological water (9 g/L NaCl) on three different culture media. Brain Heart
277 Infusion Agar (BHI, Biokar Diagnostics) supplemented with 50 mg/L amphotericin (Sigma Aldrich,
278 Saint-Louis, MO, USA) was used to count total aerobic bacteria after 48 h of incubation at 28°C. Man,
279 Rogosa and Sharpe Agar (MRS, Biokar Diagnostics, Allonne, France) supplemented with 50 mg/L
280 amphotericin was used to count lactic acid bacteria after 48 h of incubation at 30°C under anaerobic
281 conditions. Yeasts were counted on Yeast Extract Glucose Chloramphenicol (YEGC, Biokar
282 Diagnostics, Allonne, France) after 48 h of incubation at 28°C.

283 **Viral DNA extraction and metavirome analysis.**

284 Extraction of the viral fraction from cheese rind was performed according to protocol P4 detailed in
285 (24) comprising a filtration step and a chloroform treatment. DNA was extracted from the viral
286 particles according to the protocol described in the same study and sent to Eurofins Genomics for high
287 throughput sequencing using the Illumina NovaSeq platform (2×150 bp paired-end reads,
288 approximately 10 million reads per sample).

289 All the details about the tools, versions and parameters used in the following pipeline are available in
290 scripts deposited in the GitLab repository (https://forgemia.inra.fr/eric.dugat-bony/cheese_virome).
291 Briefly, raw reads were quality filtered using Trimmomatic v0.39 (43). Then a single assembly was
292 computed for the collection of triplicate reads from each sample with Spades v3.15.3 (44), using either
293 the complete dataset of trimmed reads available or after subsampling the dataset to 1.5 million,
294 150,000 or 15,000 trimmed reads per sample. We noted that some abundant contigs were assembled
295 into longer, nearly complete contigs, after subsampling. Contigs of length >2kb from all assemblies
296 were selected and clustered following an approach adapted from (45). Succinctly, a pairwise alignment
297 was first performed for all contigs using BLAT (46). Then, contigs with a self-alignment > 110% of
298 contig length, corresponding to chimeras, were removed. Remaining contigs were clustered at the
299 species level (90% identity × coverage) and the longest contig within each cluster was selected as the
300 representative sequence. The final contig dataset consisted in 3122 dereplicated contigs.

301 Viral contigs were selected using a combination of three detection tools: VIBRANT v1.2.1 (47),
302 VirSorter2 v2.2.4 (48) and CheckV v0.8.1 (49). The ones retained in the final virome were those
303 meeting at least one of the following criteria: declared “complete”, “high” or “medium” quality by
304 either VIBRANT or CheckV, declared “full” by VirSorter2. The bacterial host of the 332 viral contigs
305 was predicted using iPHoP (50). Finally, all sequences were compared by BLAT to an in-house
306 database consisting of genome sequences from 32 common dairy phages (listed in Table S1) in order
307 to identify potential related phages with known taxonomy, verified host and lifestyle (30% identity ×
308 coverage minimal cutoff). When appropriate, the host genus predicted by iPHoP was replaced by the
309 genus of the bacterial host of the closest relative phage identified by the BLAT search. One contig,
310 corresponding to the genome of the phage PhiX174 which is routinely used as control in Illumina
311 sequencing runs to monitor sequencing quality, was discarded resulting in a final dataset of 331 viral
312 contigs.

313 In order to evaluate the relative abundance of each viral contig in each metavirome sample, trimmed
314 reads were mapped against the viral contigs using bwa-mem2 v2.2.1 (51) and counted with Msamtools
315 v1.0.0 profile with the options --multi=equal --unit=ab -nolen
316 (<https://github.com/arumugamlab/msamtools>). The output files from all samples were joined into an

317 abundance table and processed using the R package phyloseq v1.38.0 (52). For each experimental
318 dataset (dynamic study, persistence study), read counts were rarefied to the minimum depth observed
319 in one individual sample. Then, read counts were normalized by contig length and transformed to
320 relative abundances, in order to enable both comparison of the viral community structure between
321 samples and comparison of the abundance level between contigs. Making the assumption that one
322 viral contig corresponds to one species (which is wrong each time several contigs belonging to the
323 same phage genome are present), Bray-Curtis dissimilarity index, as computed by the distance
324 function from the R package phyloseq, was used to compare viral communities between samples and
325 the effect of different variables on their structure was assessed using permutational analysis of
326 variance as computed by the adonis2 function from the R package vegan v2.6-2.

327 For the dynamic dataset, differential analysis was performed on raw counts using the DESeq function
328 implemented in the DESeq2 package v1.38.0 (53). Indeed, this function already includes a
329 normalization step (by the median of ratios method). Contigs were considered differentially abundant
330 if adjusted pvalue > 0.01, log2 fold change > 3 or < -3 and average raw counts > 1500.

331 **Microbial DNA extraction and 16S metabarcoding profiles.**

332 Total DNA extraction from the cheese surface was performed as previously described (10). PCR
333 amplification of the V3-V4 regions of the 16S rRNA gene was performed with the primers V3F (5'-
334 ACGGRAGGCWGCAG-3') and V4R (5'-TACCAGGGTATCTAATCCT-3')
335 carrying the Illumina 5'-CTTCCCTACACGACGCTCTTCCGATCT-3' and the 5'-
336 GGAGTTCAGACGTGTGCTCTTCCGATCT-3' tails, respectively. The reaction was performed
337 using 10 ng of extracted DNA, 0.5 μ M primer, 0.2 mM dNTP, and 2.5 U of the MTP Taq DNA
338 polymerase (Sigma-Aldrich, USA). The amplification was carried out using the following program:
339 94°C for 60 s, 30 cycles at 94°C for 60 s, 65°C for 60 s, 72°C for 60 s, and a final elongation step at
340 72°C for 10 min. The resulting PCR products were sent to the @BRIDGE platform (INRAE, Jouy-en-
341 Josas, France) for library preparation and sequencing. Briefly, amplicons were purified using a
342 magnetic beads CleanPCR (Clean NA, GC biotech B.V., The Netherlands), the concentration was
343 measured using a Nanodrop spectrophotometer (Thermo Scientific, USA) and the amplicon quality
344 was assessed on a Fragment Analyzer (AATI, USA) with the reagent kit ADNdb 910 (35-1,500 bp).

345 Sample multiplexing was performed by adding tailor-made 6 bp unique indexes during the second
346 PCR step which was performed on 50–200 ng of purified amplicons using the following program:
347 94°C for 10 min, 12 cycles at 94°C for 60 s, 65°C for 60 s, 72°C for 60 s, and a final elongation step at
348 72°C for 10 min. After purification and quantification (as described above), all libraries were pooled
349 with equal amounts in order to generate equivalent number of raw reads for each library. The DNA
350 concentration of the pool was quantified on a Qubit Fluorometer (ThermoFisher Scientific, USA) and
351 adjusted to a final concentration between 5 and 20 nM for sequencing. The pool was denatured (NaOH
352 0.1N) and diluted to 7 pM. The PhiX Control v3 (Illumina, USA) was added to the pool at 15% of the
353 final concentration, as described in the Illumina procedure, and the mixture was loaded onto the
354 Illumina MiSeq cartridge according to the manufacturer's instructions using MiSeq Reagent Kit v3 (2
355 × 250 bp paired-end reads).

356 Paired-end reads were analysed using FROGS v3.2 (54), according to the standard operating
357 procedure. Briefly, operational taxonomic units (OTUs) were built using Swarm with an aggregation
358 distance of 1 and the --fastidious option (55), and each OTU that accounted for <0.005% of the total
359 dataset of sequences was discarded, as previously recommended (56). Lastly, the OTU's affiliation
360 was checked using the EzBiocloud database v52018 (57). The abundance table was processed using
361 the R package phyloseq v1.38.0 (52).

362 **Data availability.**

363 Raw sequencing data for cheese virome and bacterial microbiome were deposited at the sequence read
364 archive (SRA) of the NCBI (<https://www.ncbi.nlm.nih.gov/sra/>) as part of bioprojects PRJNA984302
365 (dynamic study) and PRJNA984735 (persistence study).

366

367 **ACKNOWLEDGMENTS**

368 T.P. is the recipient of a doctoral fellowship from the French Ministry of Higher Education, Research
369 and Innovation (MESRI) and the MICA department of the French National Research Institute for
370 Agriculture, Food and Environment (INRAE). For the metabarcoding analysis, this work has benefited
371 from the facilities and expertise of @BRIDGE (Université Paris-Saclay, INRAE, AgroParisTech,
372 GABI, 78350 Jouy-en-Josas, France). We are also grateful to the INRAE MIGALE bioinformatics

373 facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi:
374 10.15454/1.5572390655343293E12) for providing computing and storage resources.

375

376 **TABLES**

377 **Table 1: summary statistics about the metavirome dataset.**

	Complete dataset (24 samples)	Dynamic study (15 samples)	Persistence study (9 samples)
Total number of vOTUs > 2 Kb	331	318	297
Number of vOTUs with relative abundance > 5×10^{-5}	157	168	99

378

379 **Table 2: Most abundant dairy phages detected on the cheese metavirome.** Phages that have been
380 previously isolated from the same cheese variety are highlighted in bold. When classification is
381 available at the International Committee on Taxonomy of Viruses (ICTV), the viral genus is denoted
382 in brackets.

Closest relatives	Relative abundance*
<i>Lactococcus</i> 949 group (<i>Audreyjarvisvirus</i>)	0.56
<i>Glutamicibacter</i> Montesquieu	0.23
<i>Lactococcus</i> 936 group (<i>Skunavirus</i>)	0.18
<i>Lactococcus</i> P335 group	1.54E-03
<i>Brevibacterium</i> Rousseau	1.08E-03
<i>Psychrobacter</i> D'Alembert	1.74E-04
<i>Leuconostoc</i> Diderot (<i>Limdunavirus</i>)	1.14E-04
<i>Glutamicibacter</i> Voltaire	1.70E-05
<i>Streptococcus</i> 987 group	7.87E-06

Lactococcus KSY1 group (*Chopinivirus*)

5.94E-07

Non dairy phages**

0.03

383 *The relative abundance value used here is not normalized by genome size, it corresponds to the
384 relative proportion of reads mapping to all vOTUs belonging to a given phage group present in our in-
385 house dairy phage database (Table S1).

386 **vOTUs without any match to our in-house dairy phage database (30% identity \times coverage minimal
387 cutoff).

388

389 **FIGURE LEGENDS**

390 **Figure 1: Schematic representation of the experimental plan followed for the dynamic study of**
391 **the virome composition during the ripening process.** Samples were respectively obtained after each
392 washing step (blue arrows indicate washes performed with a NaCl solution, red arrows indicate
393 washes performed with a NaCl solution containing alcoholic liquor).

394

395 **Figure 2: Dynamic of the cheese surface virome along the ripening process.** A) Principal
396 coordinate analysis of the Bray-Curtis dissimilarity. Samples are coloured according the ripening step.
397 B) Comparison of the viral diversity, estimated by the Shannon index, at the different ripening steps.
398 C) Heatmap representing the normalized relative abundance of the most abundant vOTUs present in
399 the dataset (168 vOTUs with an averaged normalized relative abundance value 5×10^{-5}) in the
400 different samples. The colours on the heatmap represent the log-transformed relative abundance, and
401 range from blue to red, blue indicating lower relative abundance and red indicating higher relative
402 abundance. When available, vOTU annotations were indicated such as the host genus as predicted by
403 iPHoP, the group of known phages it belongs to and its lifestyle.

404

405 **Figure 3: Relationship between phage dynamic and the composition of cheese bacterial**
406 **community.** A) Differential abundance analysis on the virome data between the two stable stages
407 represented by W1-2 and W4-5 samples. Log2 fold change values obtained for differential abundant

408 vOTUs (points) are represented and the dot colour indicate the predicted host genus according to
409 iPHoP. Grey dots: no predicted host. B) Composition of the bacterial community assessed by a
410 metabarcoding approach targeting the V3-V4 regions of the 16S rRNA gene. Three different cheeses
411 from the same batch were analysed at each sampling point. Data were aggregated at the genus level.
412 C) Yeasts counts expressed in log(CFU/g). D) Aerobic bacteria counts expressed in log(CFU/g). E)
413 Lactic acid bacteria counts expressed in log(CFU/g).

414

415 **Figure 4: Persistence of the cheese surface virome across production years.** A) Principal
416 coordinate analysis of the Bray-Curtis dissimilarity. Samples are coloured according the production
417 year (2017, 2019 and 2022). B) Composition of the bacterial community assessed by a metabarcoding
418 approach targeting the V3-V4 regions of the 16S rRNA gene. Three different cheeses from the same
419 production year were analysed. Data were aggregated at the genus level. C) Heatmap representing the
420 normalized relative abundance of the most abundant vOTUs (99 with an averaged normalized relative
421 abundance value $> 5 \times 10^{-5}$) in each sample. The colours on the heatmap represent the log-transformed
422 relative abundance, and range from blue to red, blue indicating lower relative abundance and red
423 indicating higher relative abundance. D) Venn diagram constructed from the dataset without filtering,
424 comprising 297 vOTUs. E) Venn diagram constructed from the filtered dataset comprising 99 vOTUs
425 with an averaged normalized relative abundance value $> 5 \times 10^{-5}$.

426

427 SUPPLEMENTARY MATERIAL

428 **Table S1. Common dairy phages and associated information.**

429

430 REFERENCES

- 431 1. Mounier J, Coton M, Irlinger F, Landaud S, Bonnarme P. 2017. Chapter 38 - Smear-Ripened
432 Cheeses, p. 955–996. *In* McSweeney, PLH, Fox, PF, Cotter, PD, Everett, DW (eds.), *Cheese*
433 (Fourth Edition). Academic Press, San Diego.

- 434 2. Corsetti A, Rossi J, Gobbetti M. 2001. Interactions between yeasts and bacteria in the smear
435 surface-ripened cheeses. *International Journal of Food Microbiology* 69:1–10.
- 436 3. Larpin-Laborde S, Imran M, Bonaïti C, Bora N, Gelsomino R, Goerges S, Irlinger F, Goodfellow M,
437 Ward AC, Vancanneyt M, Swings J, Scherer S, Guéguen M, Desmasures N. 2011. Surface
438 microbial consortia from Livarot, a French smear-ripened cheese. *Canadian Journal of*
439 *Microbiology* 57:651–660.
- 440 4. Lavoie K, Touchette M, St-Gelais D, Labrie S. 2012. Characterization of the fungal microflora in
441 raw milk and specialty cheeses of the province of Quebec. *Dairy Sci & Technol* 92:455–468.
- 442 5. Feurer C, Irlinger F, Spinnler H e., Glaser P, Vallaëys T. 2004. Assessment of the rind microbial
443 diversity in a farmhouse-produced vs a pasteurized industrially produced soft red-smear cheese
444 using both cultivation and rDNA-based methods. *Journal of Applied Microbiology* 97:546–556.
- 445 6. Mounier J, Monnet C, Jacques N, Antoinette A, Irlinger F. 2009. Assessment of the microbial
446 diversity at the surface of Livarot cheese using culture-dependent and independent
447 approaches. *International Journal of Food Microbiology* 133:31–37.
- 448 7. Mounier J, Gelsomino R, Goerges S, Vancanneyt M, Vandemeulebroecke K, Hoste B, Scherer S,
449 Swings J, Fitzgerald GF, Cogan TM. 2005. Surface Microflora of Four Smear-Ripened Cheeses.
450 *Appl Environ Microbiol* 71:6489–6500.
- 451 8. Rea MC, Görges S, Gelsomino R, Brennan NM, Mounier J, Vancanneyt M, Scherer S, Swings J,
452 Cogan TM. 2007. Stability of the Biodiversity of the Surface Consortia of Gubbeen, a Red-Smear
453 Cheese. *Journal of Dairy Science* 90:2200–2210.
- 454 9. Delcenserie V, Taminiau B, Delhalle L, Nezer C, Doyen P, Crevecoeur S, Roussey D, Korsak N,
455 Daube G. 2014. Microbiota characterization of a Belgian protected designation of origin cheese,
456 Herve cheese, using metagenomic analysis. *Journal of Dairy Science* 97:6046–6056.

- 457 10. Dugat-Bony E, Garnier L, Denonfoux J, Ferreira S, Sarthou A-S, Bonnarme P, Irlinger F. 2016.
458 Highlighting the microbial diversity of 12 French cheese varieties. *International Journal of Food*
459 *Microbiology* 238:265–273.
- 460 11. Quigley L, O’Sullivan O, Beresford TP, Ross RP, Fitzgerald GF, Cotter PD. 2012. High-throughput
461 sequencing for detection of subpopulations of bacteria not previously associated with artisanal
462 cheeses. *Applied and Environmental Microbiology* 78:5717–5723.
- 463 12. Wolfe BE, Button JE, Santarelli M, Dutton RJ. 2014. Cheese rind communities provide tractable
464 systems for in situ and in vitro studies of microbial diversity. *Cell* 158:422–433.
- 465 13. Irlinger F, Monnet C. 2021. Temporal differences in microbial composition of Époisses cheese
466 rinds during ripening and storage. *Journal of Dairy Science* 104:7500–7508.
- 467 14. Kothe CI, Bolotin A, Kraïem B-F, Dridi B, Renault P. 2021. Unraveling the world of halophilic and
468 halotolerant bacteria in cheese by combining cultural, genomic and metagenomic approaches.
469 *International Journal of Food Microbiology* 358:109312.
- 470 15. Irlinger F, Mounier J. 2009. Microbial interactions in cheese: implications for cheese quality and
471 safety. *Current Opinion in Biotechnology* 20:142–148.
- 472 16. Mayo B, Rodríguez J, Vázquez L, Flórez AB. 2021. Microbial Interactions within the Cheese
473 Ecosystem and Their Application to Improve Quality and Safety. 3. *Foods* 10:602.
- 474 17. Brown TL, Charity OJ, Adriaenssens EM. 2022. Ecological and functional roles of bacteriophages
475 in contrasting environments: marine, terrestrial and human gut. *Current Opinion in*
476 *Microbiology* 70:102229.

- 477 18. Ledormand P, Desmasures N, Dalmasso M. 2020. Phage community involvement in fermented
478 beverages: an open door to technological advances? *Critical Reviews in Food Science and*
479 *Nutrition* 1–10.
- 480 19. Paillet T, Dugat-Bony E. 2021. Bacteriophage ecology of fermented foods: anything new under
481 the sun? *Current Opinion in Food Science* 40:102–111.
- 482 20. Garneau JE, Moineau S. 2011. Bacteriophages of lactic acid bacteria and their impact on milk
483 fermentations. *Microb Cell Fact* 10 Suppl 1:S20.
- 484 21. Pujato S a., Quiberoni A, Mercanti D j. 2019. Bacteriophages on dairy foods. *Journal of Applied*
485 *Microbiology* 126:14–30.
- 486 22. Marcó MB, Moineau S, Quiberoni A. 2012. Bacteriophages and dairy fermentations.
487 *Bacteriophage* 2:149–158.
- 488 23. Muhammed MK, Kot W, Neve H, Mahony J, Castro-Mejía JL, Krych L, Hansen LH, Nielsen DS,
489 Sørensen SJ, Heller KJ, Sinderen D van, Vogensen FK. 2017. Metagenomic analysis of dairy
490 bacteriophages: extraction method and pilot study on whey samples derived from using
491 undefined and defined mesophilic starter cultures. *Appl Environ Microbiol* 83:e00888-17.
- 492 24. Dugat-Bony E, Lossouarn J, De Paepe M, Sarthou A-S, Fedala Y, Petit M-A, Chaillou S. 2020. Viral
493 metagenomic analysis of the cheese surface: A comparative study of rapid procedures for
494 extracting viral particles. *Food Microbiology* 85:103278.
- 495 25. Queiroz LL, Lacorte GA, Isidorio WR, Landgraf M, de Melo Franco BDG, Pinto UM, Hoffmann C.
496 2022. High Level of Interaction between Phages and Bacteria in an Artisanal Raw Milk Cheese
497 Microbial Community. *mSystems* 0:e00564-22.

- 498 26. Cheng L, Marinelli LJ, Grosset N, Fitz-Gibbon ST, Bowman CA, Dang BQ, Russell DA, Jacobs-Sera
499 D, Shi B, Pellegrini M, Miller JF, Gautier M, Hatfull GF, Modlin RL. 2018. Complete genomic
500 sequences of *Propionibacterium freudenreichii* phages from Swiss cheese reveal greater
501 diversity than *Cutibacterium* (formerly *Propionibacterium*) *acnes* phages. *BMC Microbiology*
502 18:19.
- 503 27. de Melo AG, Rousseau GM, Tremblay DM, Labrie SJ, Moineau S. 2020. DNA tandem repeats
504 contribute to the genetic diversity of *Brevibacterium aurantiacum* phages. *Environmental*
505 *Microbiology* 22:3413–3428.
- 506 28. Gautier M, Rouault A, Sommer P, Briandet R. 1995. Occurrence of *Propionibacterium*
507 *freudenreichii* bacteriophages in swiss cheese. *Appl Environ Microbiol* 61:2572–2576.
- 508 29. Paillet T, Lossouarn J, Figueroa C, Midoux C, Rué O, Petit M-A, Dugat-Bony E. 2022. Virulent
509 Phages Isolated from a Smear-Ripened Cheese Are Also Detected in Reservoirs of the Cheese
510 Factory. 8. *Viruses* 14:1620.
- 511 30. Walsh AM, Macori G, Kilcawley KN, Cotter PD. 2020. Meta-analysis of cheese microbiomes
512 highlights contributions to multiple aspects of quality. *Nat Food* 1:500–510.
- 513 31. de Melo AG, Lemay M-L, Moineau S. 2023. The impact of *Brevibacterium aurantiacum* virulent
514 phages on the production of smear surface-ripened cheeses. *International Journal of Food*
515 *Microbiology* 400:110252.
- 516 32. De Pasquale I, Di Cagno R, Buchin S, De Angelis M, Gobbetti M. 2014. Microbial Ecology
517 Dynamics Reveal a Succession in the Core Microbiota Involved in the Ripening of Pasta Filata
518 Caciocavallo Pugliese Cheese. *Applied and Environmental Microbiology* 80:6243–6255.

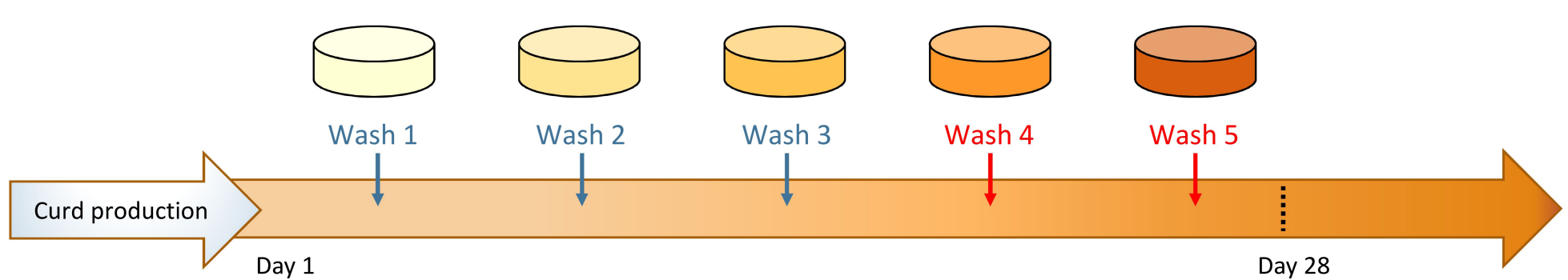
- 519 33. Flórez AB, Mayo B. 2006. Microbial diversity and succession during the manufacture and
520 ripening of traditional, Spanish, blue-veined Cabrales cheese, as determined by PCR-DGGE.
521 International Journal of Food Microbiology 110:165–171.
- 522 34. Fuka MM, Wallisch S, Engel M, Welzl G, Havranek J, Schloter M. 2013. Dynamics of Bacterial
523 Communities during the Ripening Process of Different Croatian Cheese Types Derived from Raw
524 Ewe’s Milk Cheeses. PLoS ONE 8:e80734.
- 525 35. Larpin S, Mondoloni C, Goerges S, Vernoux J-P, Guéguen M, Desmasures N. 2006. *Geotrichum*
526 *candidum* dominates in yeast population dynamics in Livarot, a French red-smear cheese. FEMS
527 Yeast Research 6:1243–1253.
- 528 36. Pangallo D, Šaková N, Koreňová J, Puškárová A, Kraková L, Valík L, Kuchta T. 2014. Microbial
529 diversity and dynamics during the production of May bryndza cheese. International Journal of
530 Food Microbiology 170:38–43.
- 531 37. Jolicoeur AP, Lemay M-L, Beaubien E, Bélanger J, Bergeron C, Bourque-Leblanc F, Doré L,
532 Dupuis M-È, Fleury A, Garneau JE, Labrie SJ, Labrie S, Lacasse G, Lamontagne-Drolet M, Lessard-
533 Hurtubise R, Martel B, Menasria R, Morin-Pelchat R, Pageau G, Samson JE, Rousseau GM,
534 Tremblay DM, Duquenne M, Lamoureux M, Moineau S. 2023. Longitudinal Study of Lactococcus
535 Phages in a Canadian Cheese Factory. Applied and Environmental Microbiology 0:e00421-23.
- 536 38. Madera C, Monjardín C, Suárez JE. 2004. Milk contamination and resistance to processing
537 conditions determine the fate of Lactococcus lactis bacteriophages in dairies. Applied and
538 Environmental Microbiology 70:7365–7371.
- 539 39. Wagner N, Brinks E, Samtlebe M, Hinrichs J, Atamer Z, Kot W, Franz CMAP, Neve H, Heller KJ.
540 2017. Whey powders are a rich source and excellent storage matrix for dairy bacteriophages.
541 International Journal of Food Microbiology 241:308–317.

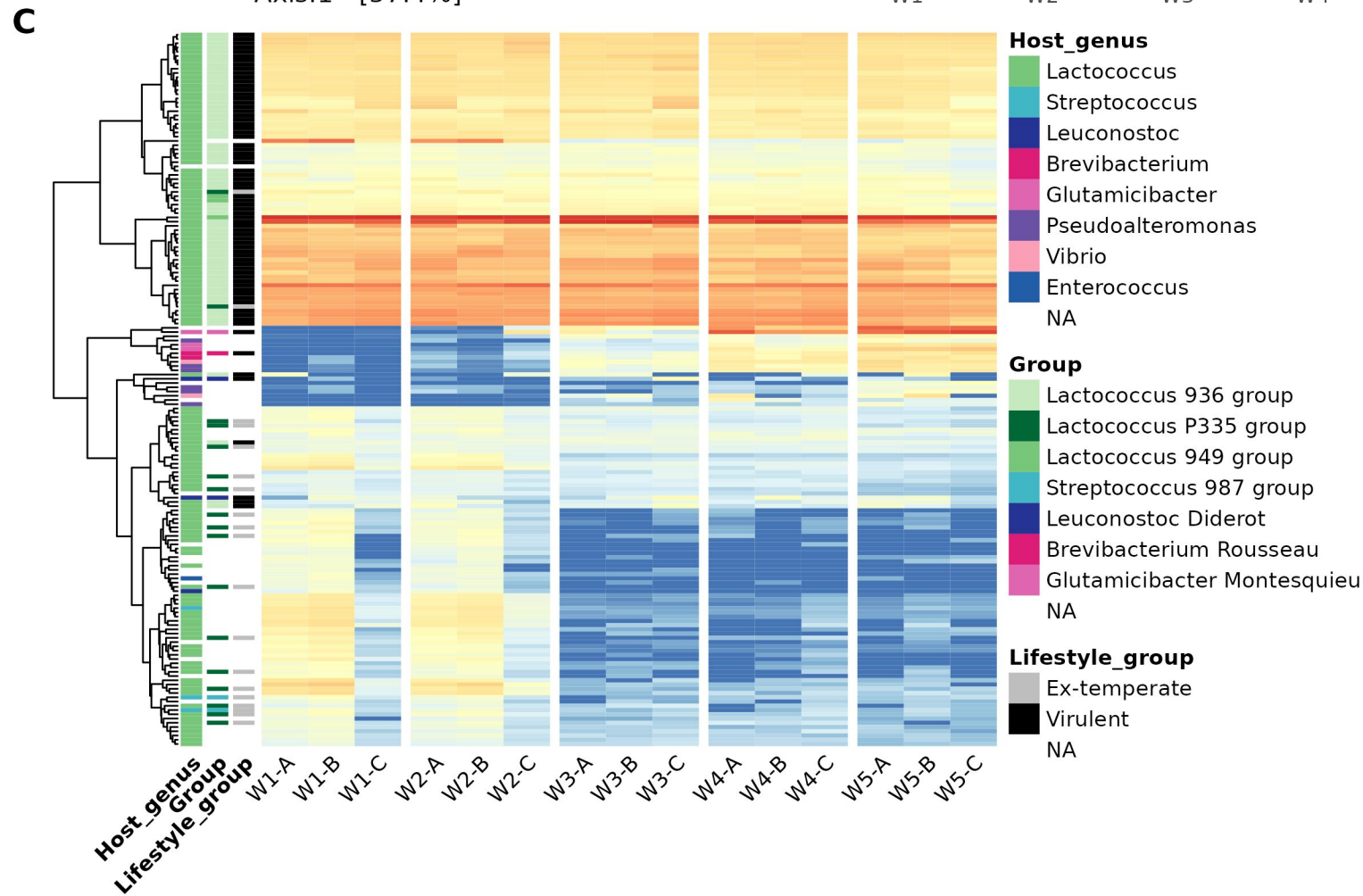
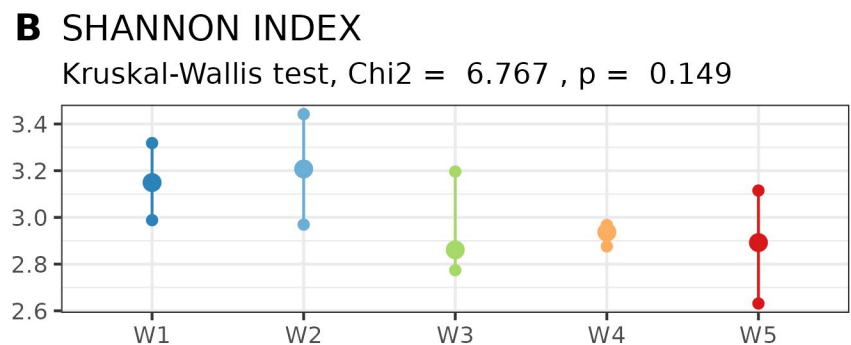
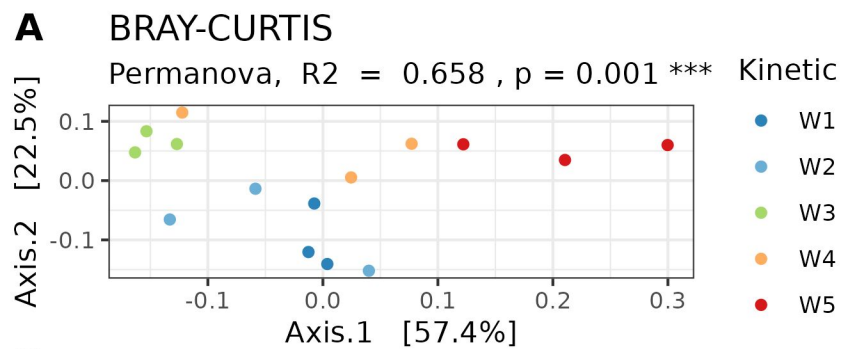
- 542 40. Bokulich NA, Mills DA. 2013. Facility-Specific “House” Microbiome Drives Microbial Landscapes
543 of Artisan Cheesemaking Plants. *Applied and Environmental Microbiology* 79:5214–5223.
- 544 41. Raymond-Fleury A, Lessard M-H, Chamberland J, Pouliot Y, Dugat-Bony E, Turgeon SL, St-Gelais
545 D, Labrie S. 2022. Analysis of Microbiota Persistence in Quebec’s Terroir Cheese Using a
546 Metabarcoding Approach. *Microorganisms* 10:1381.
- 547 42. Somerville V, Berthoud H, Schmidt RS, Bachmann H-P, Meng YH, Fuchsmann P, von Ah U, Engel
548 P. 2021. Functional strain redundancy and persistent phage infection in Swiss hard cheese
549 starter cultures. *ISME J* 1–12.
- 550 43. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
551 data. *Bioinformatics* 30:2114–2120.
- 552 44. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
553 Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.
554 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.
555 *J Comput Biol* 19:455–477.
- 556 45. Shah SA, Deng L, Thorsen J, Pedersen AG, Dion MB, Castro-Mejía JL, Silins R, Romme FO,
557 Sausset R, Jessen LE, Ndela EO, Hjelmsø M, Rasmussen MA, Redgwell TA, Leal Rodríguez C,
558 Vestergaard G, Zhang Y, Chawes B, Bønnelykke K, Sørensen SJ, Bisgaard H, Enault F, Stokholm J,
559 Moineau S, Petit M-A, Nielsen DS. 2023. Expanding known viral diversity in the healthy infant
560 gut. *Nat Microbiol* 8:986–998.
- 561 46. Kent WJ. 2002. BLAT—The BLAST-Like Alignment Tool. *Genome Res* 12:656–664.
- 562 47. Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation
563 of microbial viruses, and evaluation of viral community function from genomic sequences.
564 *Microbiome* 8:90.

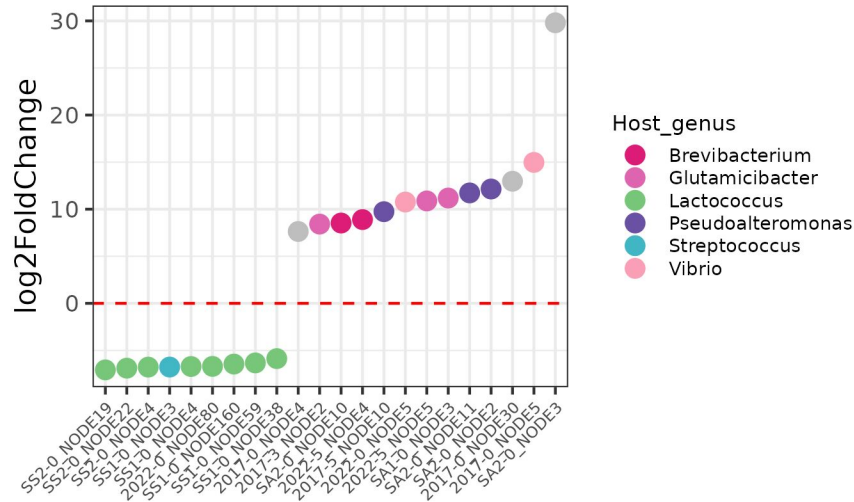
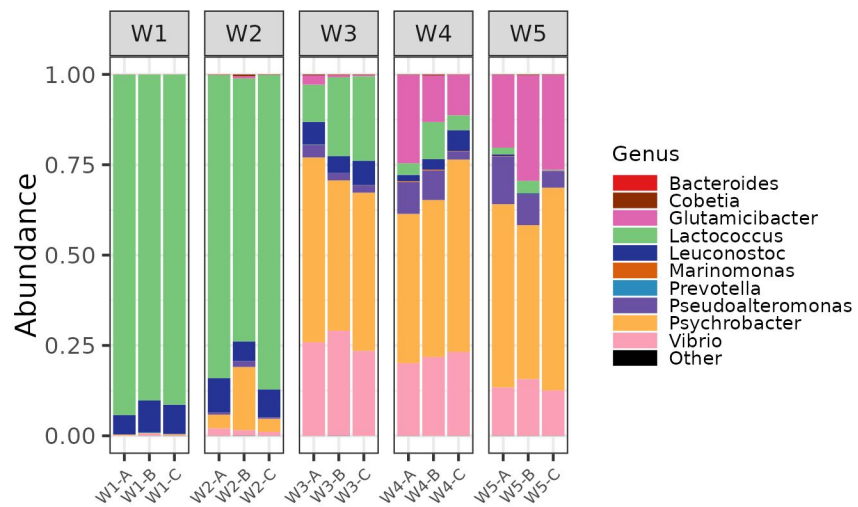
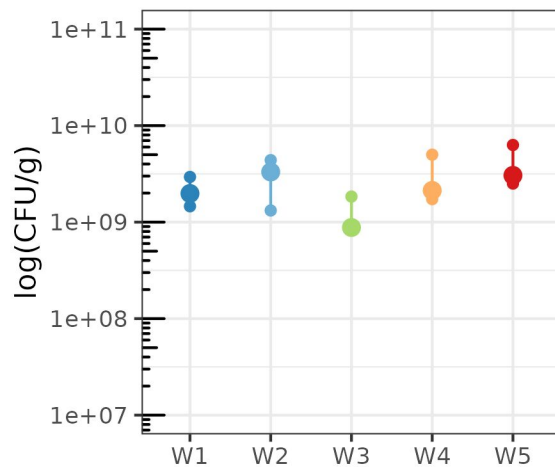
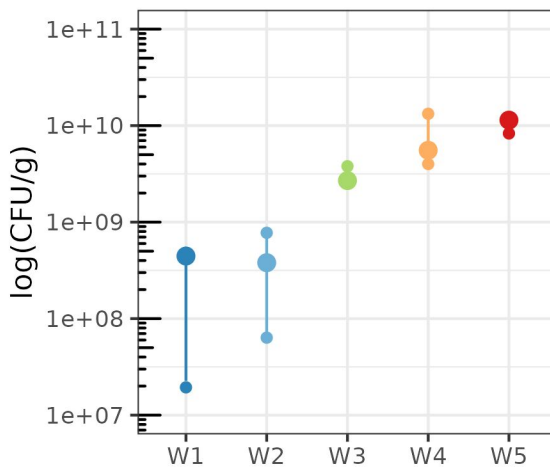
- 565 48. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa
566 MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to
567 detect diverse DNA and RNA viruses. *Microbiome* 9:37.
- 568 49. Nayfach S, Camargo AP, Schulz F, Eloie-Fadrosch E, Roux S, Kyrpides NC. 2021. CheckV assesses
569 the quality and completeness of metagenome-assembled viral genomes. 5. *Nat Biotechnol*
570 39:578–585.
- 571 50. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A. 2022. iPHoP: an
572 integrated machine-learning framework to maximize host prediction for metagenome-
573 assembled virus genomes. preprint. *Microbiology*.
- 574 51. Vasimuddin Md, Misra S, Li H, Aluru S. 2019. Efficient Architecture-Aware Acceleration of BWA-
575 MEM for Multicore Systems, p. 314–324. *In* 2019 IEEE International Parallel and Distributed
576 Processing Symposium (IPDPS).
- 577 52. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and
578 graphics of microbiome census data. *PLOS ONE* 8:e61217.
- 579 53. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
580 RNA-seq data with DESeq2. *Genome Biology* 15:550.
- 581 54. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet
582 G, Combes S, Pascal G. 2017. FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics*
583 <https://doi.org/10.1093/bioinformatics/btx791>.
- 584 55. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. 2014. Swarm: robust and fast clustering
585 method for amplicon-based studies. *PeerJ* 2:e593.

- 586 56. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG.
587 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing.
588 Nat Methods 10:57–59.
- 589 57. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a
590 taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int
591 J Syst Evol Microbiol 67:1613–1617.

592





A Differentially abundant vOTUs**B** Bacterial community composition**C** Yeasts**D** Aerobic bacteria**E** Lactic acid bacteria