








Needle in a Whey-Stack: PhRACS as a Discovery Tool for Unknown Phage-Host Combinations

Eoghan Casey,^{a,b}  Brian McDonnell,^a Kelsey White,^{a,b} Panagiota Stamou,^{b,c} Tadhg Crowley,^{b,c} Ian O'Neill,^{a,b}  Katherine Lavelle,^{a,b} Stephen Hayes,^a Gabriele A. Lugli,^d Silvia Arboleya,^e  Kieran James,^b  Marco Ventura,^d Ines Martinez,^f Miguel Gueimonde,^e Fabio dal Bello,^f Ken Nally,^{b,c} Jennifer Mahony,^{a,b}  Douwe van Sinderen^{a,b}

^aSchool of Microbiology, University College Cork, Cork, Ireland

^bAPC Microbiome Ireland, University College Cork, Cork, Ireland

^cFlow Cytometry Platform, APC Microbiome Ireland, University College Cork, Cork, Ireland

^dLaboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy

^eMicrobiota, Food and Health Group, Department of Biochemistry and Microbiology of Dairy Products, Instituto de Productos Lácteos de Asturias, Consejo Superior de Investigaciones Científicas, Villaviciosa, Spain

^fSacco Srl, Cadorago (Co), Italy

Eoghan Casey, Brian McDonnell, and Kelsey White contributed equally to this work. Author order was determined alphabetically.

ABSTRACT The field of metagenomics has rapidly expanded to become the go-to method for complex microbial community analyses. However, there is currently no straightforward route from metagenomics to traditional culture-based methods of strain isolation, particularly in (bacterio)phage biology, leading to an investigative bottleneck. Here, we describe a method that exploits specific phage receptor binding protein (RBP)-host cell surface receptor interaction enabling isolation of phage-host combinations from an environmental sample. The method was successfully applied to two complex sample types—a dairy-derived whey sample and an infant fecal sample, enabling retrieval of specific and culturable phage hosts.

IMPORTANCE PhRACS aims to bridge the current divide between *in silico* genetic analyses (i.e., phageomic studies) and traditional culture-based methodology. Through the labeling of specific bacterial hosts with fluorescently tagged recombinant phage receptor binding proteins and the isolation of tagged cells using flow cytometry, PhRACS allows the full potential of phageomic data to be realized in the wet laboratory.

KEYWORDS metagenome, virome, phageome, cytometry, fluorescent, RBP, fecal, whey, fluorescence

For over a century, culture-dependent methods have been employed as the primary method for (bacterio)phage detection, isolation, and propagation (1–3). The recent proliferation of virome/phageome studies (4–7) has supplanted such traditional methods of phage identification and has led to a vast collection of presumed phage sequences associated with unknown hosts. However, functional characterization of phages identified by a virome approach will (in most cases) require identification of a permissive host with subsequent propagation to high virion levels.

Following virome/phageome sequence read generation, bioinformatic approaches may predict bacterial hosts of putative phages associated with an identified contiguous sequence (contig), though such predictions are not necessarily robust or specific (8). Host predictions are mainly based on (i) detection of genetic identity with putative hosts, where phages may have acquired genetic elements from a previous host (success rate in host species prediction for 820 phages, ~38.5%) (8), (ii) CRISPR spacer

Editor Gustavo H. Goldman, Universidade de Sao Paulo

Copyright © 2022 Casey et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Douwe van Sinderen, d.vansinderen@ucc.ie.

The authors declare a conflict of interest. Ines Martinez and Fabio dal Bello are employees of Sacco Srl.

Received 8 November 2021

Accepted 19 November 2021

Published 4 January 2022

analysis (success rate, ~21.3%) (8), and (iii) phage and host kmer profile analysis, i.e., an assessment of viral usage of oligonucleotide strings of a specific length, profiles of which tend to shift toward those of the host over time (success rate, ~17.2%, using 4-mer analysis [8]). A recent advancement in kmer analysis (HostPhinder), profiling 16-mers of phages and sequenced hosts, showed a substantially improved prediction of host species and genus at 74% and 81%, respectively, of analyzed phages ($n = 2,196$) (9). However, these predictions provide information at the genus level (or at best the species level) but cannot identify the specific strain on which target bacteriophages may propagate.

Further methods are available to physically link unpropagated phages to their hosts, including phageFISH, though this requires genome sequences of both for probe design (10). Viral tagging allows identification of phages that can infect a target host by means of fluorescent labeling of phages using a DNA stain and subsequent fluorescence-activated cell sorting (FACS) (11). Limiting the applicability of this method is a requirement for a pure culture and the ability to isolate phages infecting only one known host.

The first stage of phage infection requires a host-encoded receptor(s) being bound by a phage-encoded receptor binding protein (RBP) (12), a highly specific interaction (13). This is illustrated in detail by two phages, TP901-1 and Tuc2009, both belonging to the P335 phage group infecting *Lactococcus cremoris*. In TP901-1, the RBP is represented by the BppU protein which assembles as a trimer that is connected to three BppL proteins, forming a tripod (14, 15). In Tuc2009, a third (accessory) protein (“BppA”) is incorporated into this tripod complex (15, 16). These phages recognize distinct carbohydrate moieties on the cell surfaces of their respective hosts (13) due to C-terminal divergence in BppL (16).

Previous studies have used fluorescently labeled RBPs to show specific receptor binding activity in several genera (17–20) and have been proposed as a method for specific identification of clinically relevant bacterial species (21, 22). Herein, we present an integrated method capable of overcoming limitations of *in silico* phage-host prediction methods while addressing discontinuity between metagenomics and culture-based techniques. In a significant prelation, phage RBP-activated cell sorting (PhRACS) exploits specific RBP-host receptor interaction in order to link the RBP sequence of a given virome-derived, phage-associated contig to its corresponding receptor-containing host, and to enable isolation of both from a variety of ecological niches.

RESULTS AND DISCUSSION

Proof of concept i: GFP-RBP construction and known host labeling. To investigate the validity of the PhRACS methodology (Fig. 1), two established phage-host combinations, Tuc2009/*L. cremoris* UC509.9 and TP901-1/*L. cremoris* 3107 (Table 1), were employed to specifically label a single host using fluorescently tagged RBP of its infecting phage.

His-tagged, green fluorescent protein (GFP)-RBP fusion complexes were constructed for each phage (22) as described in Materials and Methods and Tables 1 and 2. These complexes consisted of the full receptor binding protein assemblies (called tripods) of each phage (23), designated here as ^{His}GFP-UAL_{Tuc2009} and ^{His}GFP-UL_{TP901-1}. Following purification, tripods were incubated with host or nonhost cells, and their association with cell surfaces was assessed by fluorescence confocal microscopy. In this way, qualitative analysis of the affinity of this RBP-receptor interaction was performed, as previously described for *Lactobacillus casei* and *Listeria monocytogenes* (18, 19). Specific binding to susceptible hosts was observed in each phage-host combination, i.e., ^{His}GFP-UAL_{Tuc2009} exclusively binds to *L. lactis* UC509.9 cells (Fig. 2, panel A1), whereas ^{His}GFP-UL_{TP901-1} specifically interacts with *L. lactis* 3107 cells (Fig. 2, panel B2), thus establishing that GFP-tagged, phage-derived RBPs can bind specifically to susceptible host cell surfaces.

Proof of concept ii: Host retrieval by FACS. The potential for host isolation PhRACS was then investigated. Two pooled cultures were generated, consisting of (i) *Lactococcus*

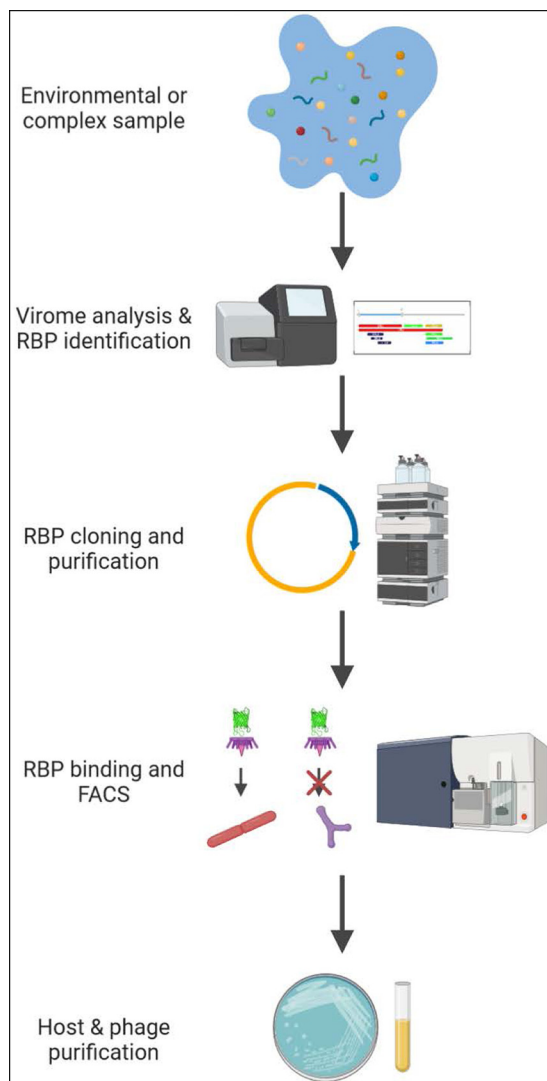


FIG 1 PhRACS workflow. Virome analysis is first performed on a complex sample, and putative RBP nucleotide sequences are identified. RBP-encoding genes are then cloned on a suitable plasmid and expressed. Incubation of GFP-tagged RBPs with the processed sample then allows isolation of the phage host from the complex sample by FACS.

Lactis 3107 wild type and chloramphenicol-resistant *L. lactis* UC509.9::pNZ44 (pool A; for assessment using $^{\text{His}}\text{GFP-UAL}_{\text{Tuc2009}}$) and (ii) chloramphenicol-resistant *L. lactis* 3107::pNZ44 and *L. lactis* UC509.9 wild type (pool B; for assessment using $^{\text{His}}\text{GFP-UL}_{\text{TP901-1}}$).

Incubation of $^{\text{His}}\text{GFP-UAL}_{\text{Tuc2009}}$ with pool A resulted in labeling of 32.2% of cells (see Fig. S1, panel A1, in the supplemental material), whereas plate counts of cells prior to incubation (Fig. S1, panel B, row 1) indicated a GFP-positive population of 56.5%. This discrepancy is likely due to the gating strategy employed, where a significant portion of the GFP-positive population was ungated (i.e., unaccounted for) due to its suboptimal purity (Fig. S1, panel A1; indicated by arrow). In total, 252,480 events (i.e., GFP-positive entities) were segregated into GM17 plus chloramphenicol. Resultant cultures were assessed for sensitivity to Tuc2009 by standard spot assay, with such sensitivity confirmed by a clearing zone on the bacterial lawn (Fig. S1, panel C1), establishing the ability to specifically retrieve viable host GFP-RBP-labeled cells by FACS.

Host retrieval was similarly demonstrated using $^{\text{His}}\text{GFP-UL}_{\text{TP901-1}}$ following incubation with pool B (40.8% GFP positive, 39.3% expected positive; Fig. S1, panel B, row 2) indicating that the GFP-tripod complex specifically binds (and thereby labels) hosts

TABLE 1 Bacterial strains and plasmids used in this study

Bacterial strain or plasmid	Relevant characteristic(s)	Reference
Bacterial strains		
<i>L. lactis</i>		
NZ9000	MG1363 containing <i>nisRK</i> genes	64
3107	Host for TP901-1	65
UC509.9	Host for Tuc2009	66
SC34	Host for BB4_2	This study
<i>Leuconostoc</i>		
LMD55	Host for LMD_1_2	This study
<i>Escherichia coli</i>		
BL21 (DE3)	Protein expression strain	New England Biolabs
Undefined starter culture		
MUO12		This study
Plasmids		
pNZ8048	Standard <i>L. lactis</i> expression vector, Cm ^r , nisin inducible	64
pGFP8048	pNZ8048 derivative harboring GFP, contains a His tag cloned in frame	This study
pNZ44	Standard <i>L. lactis</i> expression vector, Cm ^r	67
pNZ8048-His-GFP-UAL _{Tuc2009}	pNZ8048 encoding Tuc2009 <i>bppU</i> , <i>bppA</i> , <i>bppL</i> as an operon with an N-terminally His-tagged GFP fused to the N terminus	This study
pNZ8048-His-GFP-UL _{TP901-1}	pNZ8048 encoding TP901-1 <i>bppU</i> and <i>bppL</i> as an operon with an N-terminally His-tagged GFP fused to the N terminus	This study
pGFP8048-His-GFP-RBP _{LMD}	pGFP8048 encoding Φ LMD_1_2 <i>orf22</i>	This study
pHTP9-423phi1Rv1	pHTP9 encoding 423phi1Rv1 conformation	This study
pHTP9-RBP _{BB4_2}	pHTP9 encoding BB4_2 _{<i>orf04</i>}	This study

among a mixed population in both cases (Fig. S1, panel A2, and panel B, row 2). Sorted cells were collected and showed sensitivity to TP901-1 in the same manner as observed for Tuc2009 (Fig. S1, panel C2).

Proof of concept iii: Limit of host retrieval. To further assess the specificity of this interaction and to define the limit of host retrieval, serial dilutions of a fresh overnight (o/n) culture of 3107::pNZ44 to a theoretical target concentration of 1 CFU/mL were generated. These dilutions were mixed with nontarget cells (UC509.9) in various ratios (as per Fig. S1, panel B, rows 2 to 9) for assessment using GFP_{UL-TP901-1}. As expected, the observed percentage of target GFP-positive population decreases to 0.066 (Fig. S1, panel A5; Fig. S1, panel B, row 5). However, nonspecific background fluorescence at approximately 0.02% to 0.04% remained in samples with the lowest levels of 3107::pNZ44 target cells (Fig. S1, panel B, rows 6 to 9). Despite this, cell sorting below this limit should theoretically still occur, and indeed retrieval of culturable target hosts at all tested target-nontarget cell ratios was demonstrated by subsequent reinfection using TP901-1 (Fig. S1, panels C3 to C9), despite the presence of background levels of fluorescence (Fig. S1, panel B, rows 6 to 9).

Phageome analysis i: Dairy niche. To assess whether PhRACS can retrieve an unknown phage-host combination from an industrial ecological niche, a dairy-derived whey sample was employed. Undefined mesophilic starter cultures used in cheese production represent complex communities composed of a large number of lactococcal strains with additional *Leuconostoc* species (24).

Phageome sequencing was first performed on a whey sample derived from a quark manufactured by an undefined mesophilic culture (termed MUO12). Assignment of taxonomic lineage to obtained sequence reads revealed a predicted *Leuconostoc* phage read abundance of approximately 21.6% within the sample. Following sequence assembly, a contig was identified that exhibited high nucleotide identity (98% across 97% of the contig) to *Leuconostoc* phage Φ LMD_1_2, which has been reported to infect *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides* strains (25). Kmer analysis of Φ LMD_1_2 using HostPhinder predicted *Ln. pseudomesenteroides* as the host for this phage (coverage, $1.4e-01$) (9), and this contig was selected for further analysis.

TABLE 2 Oligonucleotides used in this study

Oligonucleotide	Sequence (5'-3') ^a	Comment ^b	Restriction enzyme ^c
GFPwHis	AGCAGGCATGCGT CACCATCATCATCAT AGTAAAG GAGAAGAACTTTTC	F primer of gfp + flanked by a His tag	SphI
GFPRevlinkBppU	TATAAAATGTTCTGTTCATACAGAAACCACCAGAACCCACC TTTTGAGCTCATCCAT	R primer of gfp + flanked by a linker and region complementary to BppU N terminus	
BppUF	GGTTCGTGATGACAGAAACATTTTATA	F primer BppL Tuc2009/TP901-1	
BppLTucR	AGCAGCAGCTAGTCTAATTCCTCGATAAAGTTTTAC	R primer of BppL Tuc2009	SpeI
BppLTPR	AGCAGCAGCTAGTCTAATTCCTCGATAAAGTTTTAC	R primer of BppL TP901-1	SpeI
GFP8048F	AGCAGGCATGCGT CACCATCATCATCAT AGTAAAGG AGAAGAACTTTTC	F primer of gfp + flanked by a His tag	SphI
GFP8048R	AGCAGCAGCTAGTGGTACCAGATCGAATTCACCAGAACCCACCA CCAGAACCCACTTTGTAGAGCTCATCCAT	R primer of gfp + flanked by flexible linker with MCS (EcoRI, EcoRV, KpnI, SpeI, XbaI)	SpeI
LMD_22F	AAAAAAGAAATTCACATTAACAATAACAGAAATTAGTAT	F primer LMD_1_22	EcoRI
LMD_22R	AAAAAAGGTACCTTAGTCTGGAGTTGGCTTAACCCAC	R primer LMD_1_22	KpnI
GFP423phi1F	TCAGCAAAGGGCTGAGGGCAACAATCTGCTTGTAAAC	F primer targeting 423phi1_Rv1 flanked by pHTP9-specific overhang	NA
GFP423phi1R	TCAGCGGAAGCTGAGGCTACGCGGTAATCCAAACAGCC	R primer targeting 423phi1_Rv1 flanked by pHTP9-specific overhang	NA
BB4-RBPF	TCAGCAAAGGGCTGAGGATGACAAATTAATAAATTCACGTT	F primer BB4_2_04 flanked by pHTP9-specific overhang	NA
BB4-RBPR	TCAGCGGAAGCTGAGGCTAATTTATACGTTTCCAAATAT	R primer BB4_2_04 flanked by pHTP9-specific overhang	NA

^aRestriction sites are underlined. Linker and overhang sequences are italicized. His tag sequences are indicated in boldface type.

^bMCS, multiple cloning sites.

^cNA, not available.

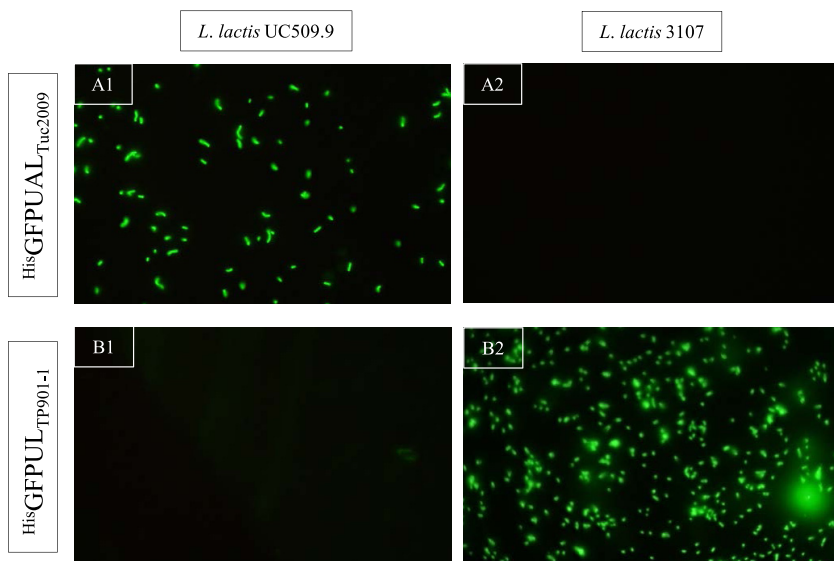


FIG 2 (A1) Interaction between recombinant $^{\text{His}}\text{GFP-UAL}_{\text{Tuc2009}}$ and Tuc2009 host *L. cremoris* UC509.9. (A2) Interaction between recombinant $^{\text{His}}\text{GFP-UAL}_{\text{Tuc2009}}$ and Tuc2009 nonhost *L. cremoris* 3107. (B1) Interaction between recombinant $^{\text{His}}\text{GFP-UL}_{\text{TP901-1}}$ and TP901-1 nonhost *L. cremoris* UC509.9. (B2) Interaction between recombinant $^{\text{His}}\text{GFP-UL}_{\text{TP901-1}}$ and TP901-1 host *L. cremoris* 3107.

HHpred (26), which has been utilized to identify RBPs of phages infecting various genera (18, 19, 27–31), was employed to interrogate the gene products of ΦLMD_{1_2} to identify the RBP-encoding gene of this phage. Analysis of the product of $\text{orf22}_{\text{LMD}_{1_2}}$ revealed structural similarity to RBPs of lactococcal phages 1358 (probability = 99.85%) and p2 (probability = 99.8%) (Fig. 3, panel A) denoting this as the most likely candidate, and this gene was selected for cloning.

To test the robustness of the method, a second phageome analysis was performed on a whey sample from a separate MUO12-based fermentation. Here, most (90.6%) reads were predicted to originate from multiple *L. lactis* phages, all belonging to the *Skunavirus* (previously 936) phage group. Following sequence assembly, a contig (termed BB4_2) containing a partial phage genome was identified that exhibited high nucleotide identity (96% across 73% of the contig) with *L. lactis* phage 645 (32), with $\text{orf04}_{\text{BB4_2}}$ being identified as the likely RBP-encoding gene of this phage. Aligning $\text{ORF04}_{\text{BB4_2}}$ against previously sequenced 936 phage RBPs revealed it to be a member of the group III RBPs which are known to bind to either cell wall polysaccharide (CWPS) type B or C lactococcal hosts (33).

Host retrieval i: Dairy niche. Cloning of $\text{orf22}_{\text{LMD}_{1_2}}$ into pGFP8048 allowed production of an ~60-kDa protein representing the RBP of phage LMD_1_2 (Fig. 3, panel B), which was then purified (to give $^{\text{His}}\text{GFP-RBP}_{\text{LMD}}$) and assessed for its ability to bind selectively to its cognate host cell receptor in a mixed starter culture. To achieve this, the whole *L. lactis* and *Leuconostoc* complements of the commercial starter were separately cultivated in selective media, and these cultures were pooled. $^{\text{His}}\text{GFP-RBP}_{\text{LMD}}$ (or protein storage buffer [PB] as a negative control) was then added to this pooled culture as described in Materials and Methods. As expected, fluorescence microscopy revealed specific labeling of a subpopulation of the mixed culture (Fig. 3, panels C1 to C3), while the culture containing control remained unlabeled (Fig. 3, panels C4 to C6).

FACS was then performed in triplicate to isolate labeled cells. Dot plots for one FACS replicate (of a total of three performed) are presented, showing separation of 0.6% GFP-positive cells from the total community (Fig. 4, panel A) compared to 0% for the PB control (Fig. 4, panel B), enabling sorting of the GFP-positive population. This separation was consistent across all replicates with an average of 0.8% ($\pm 0.2\%$) GFP-

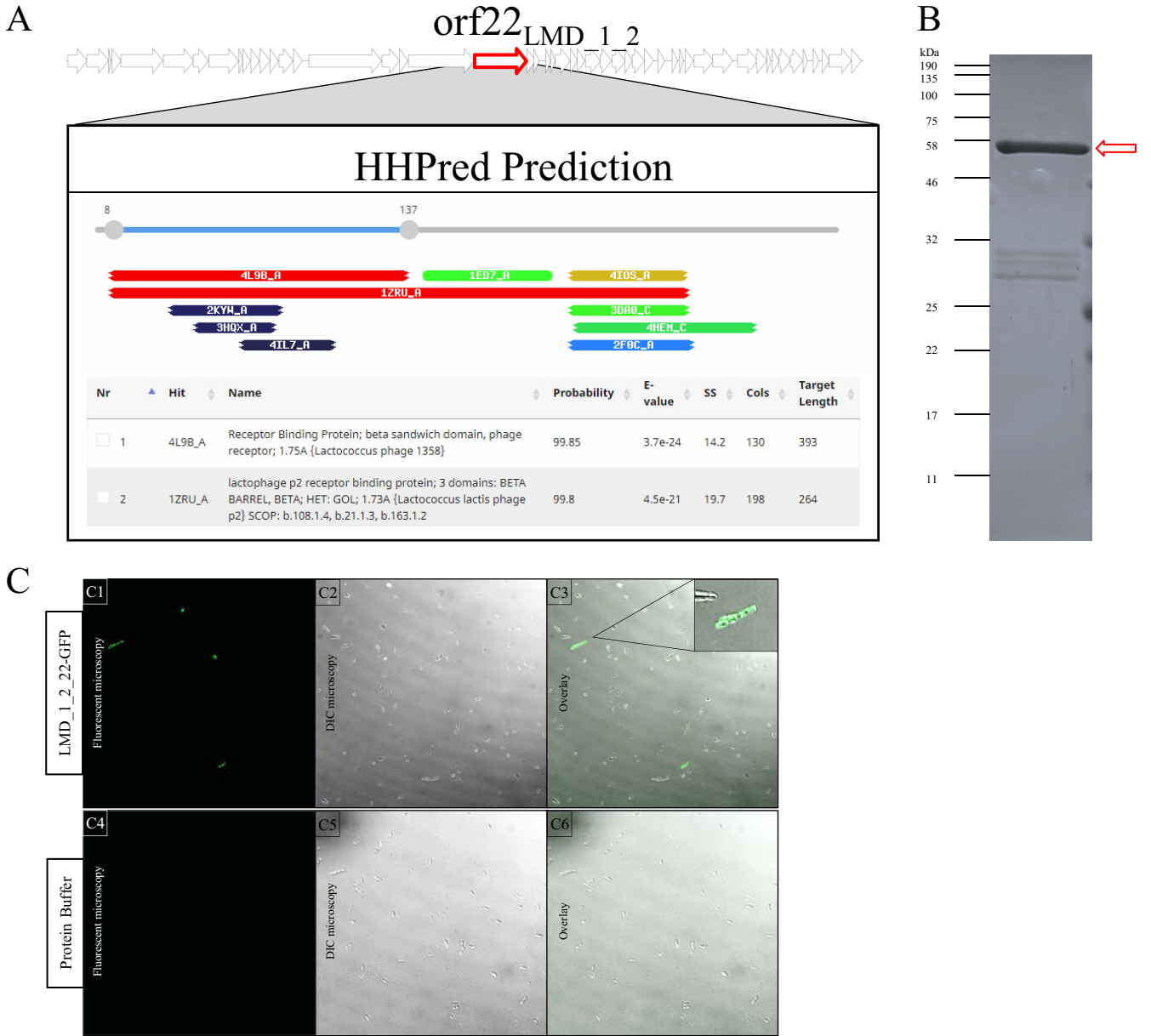


FIG 3 (A) HHpred analysis of *orf22*_{LMD_1_2}, a putative RBP, displaying similarity to lactococcal phages 1358 and p2 RBPs. (B) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of HisGFP-RBP_{LMD} fusion protein (indicated by the red arrow). (C) Putative phage-susceptible hosts within an undefined culture specifically labeled by putative GFP-tagged putative RBPs (C1 to C3), visualized using fluorescence microscopy. Protein buffer (C4 to C6) was employed as a control. DIC, differential interference contrast.

positive cells. Normalizing processed events across samples to 1,500,000 events resulted in an average of $49,591 \pm 19,129$ events sorted following incubation with HisGFP-RBP_{LMD}, while incubation with PB yielded 1.9 ± 0.4 events (Fig. 4, panel C). Plating of sorted events from PB controls yielded no growth, while plating of HisGFP-RBP_{LMD} sorted cells on MRS agar yielded colonies with two distinct morphologies, large and small, being present at 6.2×10^4 CFU/mL and 8.67×10^3 CFU/mL across three replicates, respectively (Fig. 4, panel D). 16S rRNA gene sequence analysis of selected cells of the large colony morphology revealed 99% nucleotide identity with strains of *Ln. mesenteroides*, whereas those of the smaller colony morphology exhibited 99% nucleotide identity with *L. lactis*. This coisolation may be explained by close association of these bacterial genera in milk cultures, as shown and reviewed previously (34, 35). Despite apparent coisolation, binding assays conducted using purified HisGFP-RBP_{LMD}

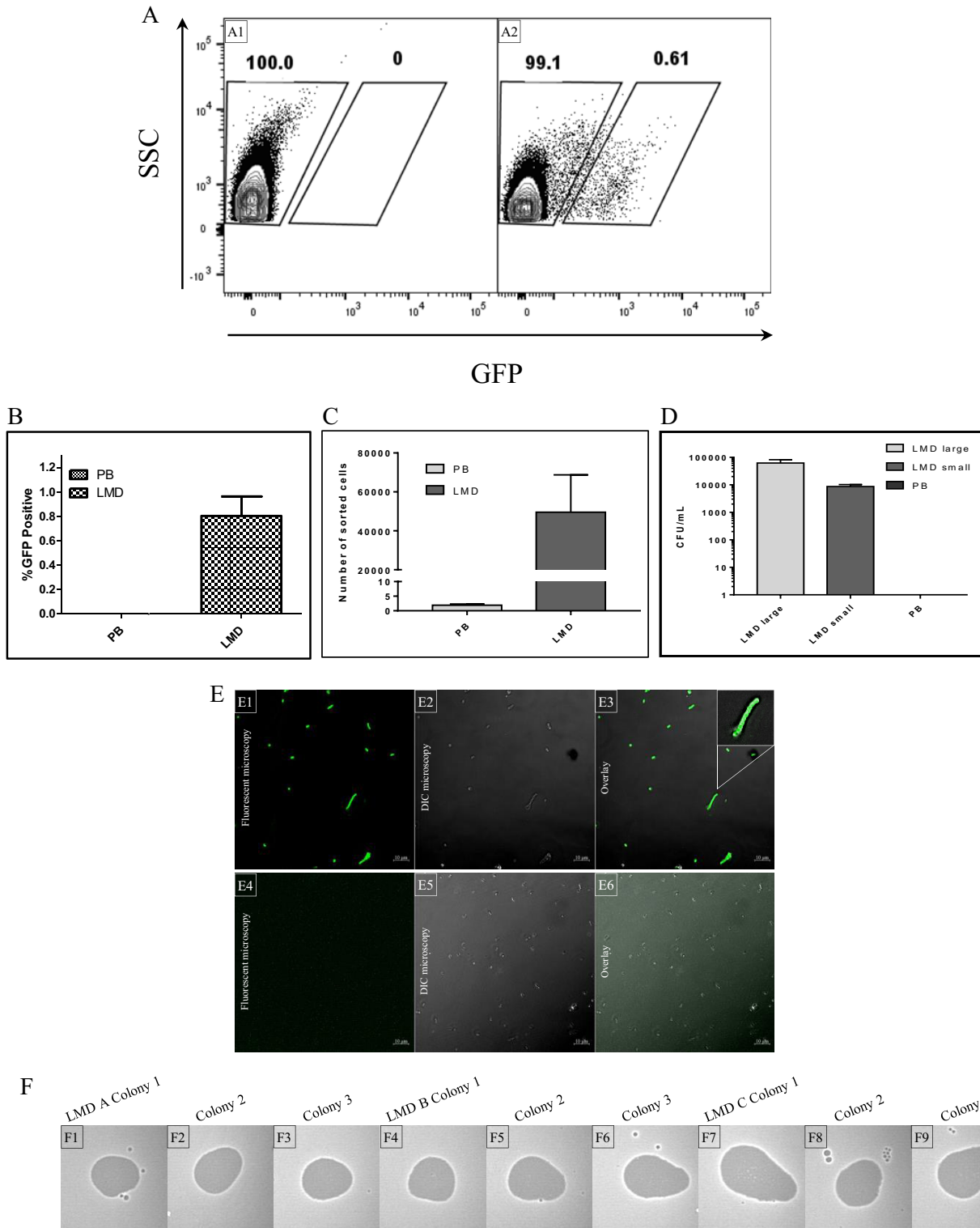


FIG 4 (A) Dot plot analysis of undefined culture MUO12 labeled with protein buffer (PB) (A1) and ^{His}GFP-RBP_{LMD} (A2). SSC, side scatter. (B) Proportion of GFP-positive events expressed as a percentage of total events processed. (C) Number of events sorted using gating strategy targeting GFP-positive cells evident in Fig. 4, panel A2. (D) Plate counts of colony morphologies evident in sorted population following plating on MRS agar. (E) Assessment of ^{His}GFP-RBP_{LMD} binding to *Ln. mesenteroides* (E1 to E3) and *L. lactis* subsp. *lactis* (E4 to E6) FACS coisolates. Colonies were visualized using fluorescence confocal microscopy at a wavelength of 488 nm (E2 and E4) overlaid on differential interference contrast (DIC) microscopy (E1 and E3). (F) Confirmation of recovery of *Ln. mesenteroides* phage LMD_1_2 host by spot assay using original whey sample MUO12, where visible zones of lysis indicate successful infection (F1 to F9).

and the two bacterial isolates in question showed that ^{His}GFP-RBP_{LMD} specifically binds to the *Ln. mesenteroides* isolate only (Fig. 4, panels E1 to E3) and not to the *L. lactis* isolate (Fig. 4, panels E4 to E6).

Having retrieved the presumed *Ln. mesenteroides* host from undefined culture, an attempt was then made to retrieve the infecting viral particle. Nine colonies identified as *Leuconostoc* species (three from each replicate) were tested for infection by phages from the original whey sample, using a double layer spot assay method. All selected colonies exhibited sensitivity to phages present in the whey sample, indicated by a zone of clearing on bacterial lawns (Fig. 4, panels F1 to F9). One of these isolates—termed *Leuconostoc* LMDS5—was selected as a host for phage propagation. A single plaque phage isolate from the whey sample was propagated on LMDS5, and DNA from the resulting lysate was isolated. Upon analysis, the genome sequence of the retrieved phage was found to exhibit 100% nucleotide identity with the previously sequenced ϕ LMD_1_2 from the phageome analysis, confirming isolation of a specific phage and host from whey and undefined starter culture, respectively.

To isolate the host of proposed *L. cremoris* phage BB4_2 from the second whey sample, *orf04*_{BB4_2} was cloned into pHTP9 and an ~57-kDa protein (^{His}GFP-RBP_{BB4_2}) was expressed and purified. Fluorescence microscopy revealed specific labeling of a subpopulation of the starter culture, which had been enriched for lactococci by overnight growth in GM17. Following the binding assay, FACS was performed in triplicate to isolate labeled cells. An average separation of 0.11% ($\pm 0.07\%$) GFP-positive cells from the total community was observed. Plating ^{His}GFP-RBP_{BB4_2} sorted colonies on GM17 yielded two distinct morphologies, large and small, being present at 1.11×10^2 CFU/mL and 2.43×10^2 CFU/mL, respectively. Interestingly, 16S rRNA sequence analysis revealed isolates of both morphologies belong to closely related *L. cremoris*, *L. lactis*, or *Lactococcus laudensis* species.

Several lactococcal isolates were selected from the sorted cells for phage sensitivity assays (32). One isolate (termed SC34) exhibited sensitivity to at least one phage present in the whey sample. SC34 was determined by PCR to harbor a B-type CWPS gene cluster, which correlates with the predicted RBP grouping of phage BB4_2, discussed above. A single plaque-purified phage isolate was propagated, and the RBP-encoding gene was amplified and sequenced to confirm 100% nucleotide identity to that assembled from virome data and used for host retrieval.

Interestingly, although *L. cremoris* and *L. lactis* strains are the most dominant in this mixed starter (accounting for over 95% of reads, based on metagenomic sequencing), CWPS type B strains were not detected during metagenomic read mapping using our method (likely due to insufficient read depth). This indicates that CWPS type B strains were present in very low abundance within the sample. Despite this finding, the presence of strains harboring this gene cluster was in fact detected in the original whey sample by PCR using DNA extracted for metagenome sequencing as the template (32), highlighting the sensitivity of the PhRACS method.

Phageome analysis ii: Intestinal niche. To highlight the versatility of PhRACS, it was applied to an alternative ecological niche, i.e., an infant fecal sample. The activities of bifidobacteria and their infecting phages may be a key ecological parameter in the infant gut microbiota, though little is known about their interactions (36–39). Appropriate development of an infant microbiota is linked to positive health outcomes (40–43), and understanding of key modulators of this dynamic ecological environment merits further research.

In order to gain an appreciation of bacterial diversity in the intestinal niche, 16S rRNA gene-encoding sequence analysis was first conducted on a fecal sample (termed T60III) taken from a 30-day-old, vaginally delivered, full-term, breast-fed baby, in order to determine the bacterial component of the microbiota composition (44). Relative abundances of 16S rRNA gene-based sequences detected by phylum, family, and genus are given in Table 3. Bacterial genera detected in this sample included *Bacteroides*, *Parabacteroides*, *Escherichia-Shigella*, and *Bifidobacterium*. Following this

TABLE 3 Microbial composition (greater than or equal to 0.01% read abundance) in fecal sample T60III by phylum, family, genus, and *Bifidobacterium* species

Analysis	Level	Taxon	Abundance (%)
16S	Phylum	<i>Actinobacteria</i>	18.78
		<i>Bacteroidetes</i>	7.89
		<i>Firmicutes</i>	2.26
		<i>Proteobacteria</i>	71.05
16S	Family	<i>Bifidobacteriaceae</i>	18.72
		<i>Bacteroidaceae</i>	6.36
		<i>Porphyromonadaceae</i>	1.52
		<i>Enterobacteriaceae</i>	71.00
16S	Genus	<i>Bacteroides</i>	6.36
		<i>Parabacteroides</i>	1.52
		<i>Escherichia-Shigella</i>	70.48
		<i>Bifidobacterium</i>	18.72
ITS	<i>Bifidobacterium</i> species	<i>B. longum</i> subsp. <i>longum</i>	60.45
		<i>B. pseudocatenulatum</i>	38.51
		<i>B. bifidum</i>	0.89
		<i>B. adolescentis</i>	0.05
		<i>B. longum</i> subspecies	0.02
		<i>B. parvae</i>	0.02
		<i>B. longum</i> subsp. <i>infantis</i>	0.01
		<i>B. animalis</i> subsp. <i>lactis</i>	0.01
		<i>B. asteroides</i>	0.01
		<i>B. breve</i>	0.01
		<i>B. scardovii</i>	0.01
		Other <i>Bifidobacterium</i> species	0.02

and considering the abundance (Table 3) and relevance of bifidobacteria, internal transcribed spacer (ITS) region sequence analysis was performed to determine the composition of the bifidobacterial population in the sample to species and subspecies level. Species belonging to this genus included *Bifidobacterium longum* subsp. *longum* (species abundance within genus, 60.45%) and *Bifidobacterium pseudocatenulatum* (species abundance within genus, 38.51%), with a diverse range of other bifidobacteria also being detected (Table 3).

To obtain corresponding T60III virome data to enable phage RBP identification, phageome sequencing and analysis of T60III were conducted in parallel. Among 605 (total) phage types whose gene sequences were identified from this analysis was a phage sequence identical to 423phi1, a previously identified inducible prophage of *Bifidobacterium breve* strain 139W4-23 (39, 44). Following this identification, phageome reads from the T60III fecal sample were mapped against genes encoding 423phi1rv1, the RBP of 423phi1 at a 99% nucleotide identity stringency level (39). This analysis yielded 24 mapped reads, representing a small percentage of the overall number of reads and, indeed, of overall phage diversity in the sample. These results indicated the presence of a phage closely related to 423phi1 and (presumably) a permissive host for that phage in the T60III fecal sample (in agreement with the predicted presence of bifidobacteria), and which we endeavoured to retrieve from this complex sample using PhRACS.

Host retrieval ii: Intestinal niche. For host retrieval, the bifidobacterial component of sample T60III was cultivated, and the resulting cells were pooled. A GFP-tagged 423phi1Rv1 protein (termed ^{His}GFP-Rv1_{423phi1}) or PB (as control) was then produced and added to the pooled cells. FACS was then performed in triplicate, and dot plots were generated for one of these replicates in Fig. 5, panel A2. This plot shows separation of 1.08% GFP-positive cells from the total community, compared to 0.008% for the control (Fig. 4, panel A1), enabling sorting of the GFP-positive population. This separation was consistent across all replicates with an average of 1.1% ± 0.14% GFP-positive cells (Fig. 5, panel B).

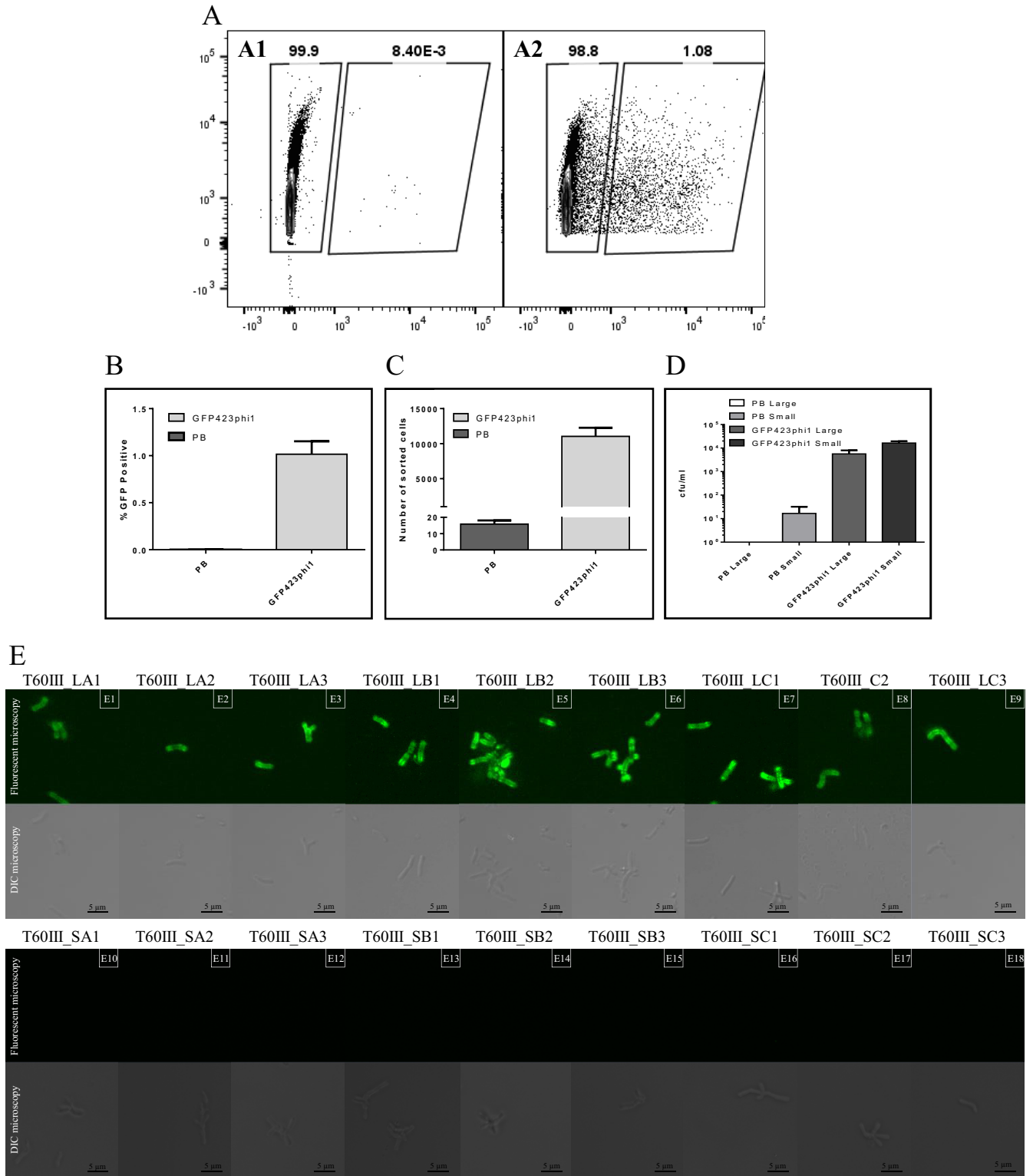


FIG 5 (A) Dot plot analysis of the T60III bifidobacterial community labeled with PB (A1) and HisGFP-RV1_{423phi1} (A2). (B) Proportion of GFP-positive events expressed as a percentage of total events processed. (C) Number of events sorted using gating strategy targeting GFP-positive cells evident in Fig. 4, panel A. (D) Plate counts (on MRS agar) of populations exhibiting distinct colony morphologies present in sorted population. (E) Confirmation of recovery of potential host by GFP-RBP labeling of *B. pseudocatenulatum* cultures derived from isolated colonies (E1 to E9). Unlabeled coisolates (*B. longum* subsp. *longum*) are shown in E10 to E18.

Normalizing processed events across samples to 1,500,000 resulted in an average of $11,034 \pm 1,229$ events sorted following incubation with $^{\text{His}}\text{GFP-Rv1}_{423\text{phi1}}$, while incubation with PB yielded 15.87 ± 2.18 events (Fig. 5, panel C). Plating of $^{\text{His}}\text{GFP-Rv1}_{423\text{phi1}}$ -sorted cells on RCM agar yielded two colony morphologies, large and small, present at an average of 5.67×10^3 CFU/mL (T60III_L-) and 1.63×10^4 CFU/mL (T60III_S-), respectively, across three replicates (Fig. 5, panel D). In contrast, plating of sorted events from the PB samples yielded an average of 8.33×10^0 colonies.

Eighteen sorted colonies (three representatives of each colony morphology type from each of three replicates) were selected and tested for binding by $^{\text{His}}\text{GFP-Rv1}_{423\text{phi1}}$. Subsequent labeling of isolated colonies revealed that T60III_L-morphotype colonies were labeled (Fig. 5, panels E1 to E9), while all tested T60III_S-morphotype colonies were not labeled by $^{\text{His}}\text{GFP-Rv1}_{423\text{phi1}}$ (Fig. 5, panels E10 to E18). To confirm that isolation of bifidobacterial hosts was achieved, genome sequences of all 18 isolates were determined. *In silico* analysis revealed that all T60III-L isolates were clonal and members of the *B. pseudocatenuatum* species, while all T60III-S isolates were clonal and members of *B. longum* subsp. *longum*. The dominance of the nonlabeled strain (*B. longum* subsp. *longum*) may be due to a physical interaction between labeled and unlabeled cells, since bifidobacteria have previously been shown to auto- and coaggregate in a strain-dependent manner (45–48).

The above demonstrates enrichment of the $^{\text{His}}\text{GFP-Rv1}_{423\text{phi1}}$ binding population from 1.1% to 34.8% (Fig. 5, panel D) and the ability to isolate pure culture of this $^{\text{His}}\text{GFP-Rv1}_{423\text{phi1}}$ -interacting population. Current laboratory conditions are not permissive of plaque formation using bifidobacteria (39), limiting additional analyses of this bifidobacterial phage-host interaction.

Concluding remarks. Here, we established a method enabling isolation of phage-host combinations from mixed, complex bacterial communities using FACS, where hosts are labeled using probes derived from the cell-surface recognition protein of its infecting phage.

To infect its host, a bacteriophage must specifically bind to a molecule that is exposed at the host cell surface, representing the first stage of infection (12). As such, PhRACS relies on prediction of a phage-encoded RBP by phageome analysis, followed by construction of a GFP-RBP fusion protein which allows for specific labeling of its putative host within a complex community. These labeled hosts are then isolated by fluorescence-mediated cell sorting, as previously seen in methods such as viral tagging (11). A major advantage of PhRACS over similar methods is that it allows isolation of a host-phage combination from a given environmental sample using just phageome sequence data as a starting point. In contrast, other methods (such as viral tagging) allow only isolation of phages infecting a known host.

To demonstrate its utility, proof of concept was first established, and PhRACS was then applied to two complex samples, i.e., a whey sample derived from a fermentation utilizing an undefined mesophilic culture (MUO12), and an infant fecal sample (T60III). Following identification of a putative *Leuconostoc* phage RBP using phageome analysis of MUO12, construction of a GFP-fusion protein and PhRACS application resulted in an enrichment of the target population from 0.6% to 87%, enabling straightforward host isolation by routine methods.

Using a fluorescently tagged RBP of a previously described bifidophage, we successfully isolated a fecal sample-derived bifidobacterial strain, to which this RBP specifically binds. Due to the apparent recalcitrance of *Bifidobacterium* phages to form *in vitro* plaques, further study of the interaction is currently difficult. Nonetheless, this method represents an ideal tool for further exploration of these interactions with the capacity to prove both binding activity of a putative *Bifidobacterium* (or, indeed, any) phage and isolation/identification of potential hosts.

At present, several limitations of the current method can be identified. First, the RBP-encoding gene of the target phage must be predicted *in silico* to produce a suitable “hook” with which to retrieve the phage host. This may be an issue in cases where phages do not encode an easily recognizable RBP-encoding gene. To overcome this (at

least for phages binding to carbohydrate receptors), identified carbohydrate binding domains (CBMs), responsible for phage host adhesion (49) and which have been shown to be host specific (50, 51), could be employed.

Second, retrieval of two distinct bacterial isolates using a single phage RBP hook occurred during whey and infant fecal sample processing. While the reason for this remains unclear, we postulate that this is due to close association of coisolated strains in their natural environment, either in biofilms or through general aggregation. Another cause may be the use of an 85- μm nozzle during cell sorting (see Materials and Methods), which was employed to decrease cell stress and improve purity. Future studies may benefit from the use of a smaller nozzle, or indeed single cell sorting, as has been employed in other studies (52). In addition, physical or chemical pretreatment of samples may be necessary to disrupt biofilms prior to performing binding assays. Third, in both examples presented above (dairy and intestinal niches), target bacterial populations were enriched by cultivation prior to GFP-RBP binding to increase the likelihood of host identification. Further studies are required to investigate whether this cultivation is necessary, which may provide further evidence of the robustness of PhRACS.

Beyond isolation of hosts for phages, this protocol has further applications that could be investigated. The advent of metagenomics has resulted in a large-scale exploration of diverse and microbe-rich communities, including soil (53), marine (54, 55), and wastewater-related sites (56, 57) and the gut (43). Considering that prophages are widely distributed on bacterial genomes sequenced from these environments and that the prophage must have adsorbed to the host surface through a cell-surface receptor, a recombinant phage RBP may act as a hook by which bacterial strains of interest may be isolated. Theoretically, any prophage-harboring bacterial strain may be isolated from a complex community using this technique. In the event that a host is not isolated, enrichment of the RBP-bound bacterial host may be sufficient to enable determination of its genome sequence, by comparing metagenome data from the original sample to that of a sample having undergone PhRACS. Further applications of PhRACS (or variations thereof) will undoubtedly be identified in future attempts to bridge the current gap between metagenomics and more traditional culture-based microbiology.

MATERIALS AND METHODS

Bacterial and phage culture conditions and storage. Phages, bacterial cultures, and plasmids employed in this study are listed in Table 1. Standard bacterial and phage culture conditions (see supplementary method 1 in Text S1 in the supplemental material) were employed.

Molecular cloning. Primers used in this study (Table 2) were synthesized by Eurofins Genomics (Germany). PCR amplifications were performed using Phusion High Fidelity DNA polymerase (New England Biolabs, UK). Restriction enzymes were purchased from Roche (Germany), and T4 DNA ligase was obtained from Promega (USA). All restriction digestions and ligations were performed using standard molecular cloning techniques (58) employing restriction enzymes (where appropriate) specified in Table 2 and as described in supplementary method 2 in Text S1. Recombinant protein production is described in supplementary method 3 in Text S1.

Binding assays. Qualitative assessment of the binding of GFP-RBP fusion proteins to bacterial cell surfaces was performed by a method adapted from Dieterle et al. (19) and as described in supplementary method 4 in Text S1. Fluorescence microscopy was performed using a Zeiss LSM 5 Exciter (Carl Zeiss, Germany) utilizing an argon laser and detector and filter sets for monitoring GFP (excitation, 488 nm; emission, 505 to 550 nm).

Flow cytometry cell preparation. For assessment using GFP^{UAL}_{TUC2009}, 150 μL of early exponential phase (optical density at 600 nm [OD₆₀₀] \approx 0.2) of *L. lactis* UC509.9::pNZ44 cells was harvested, resuspended in 150 μL of SM buffer, and mixed with similar phase *L. lactis* 3107 wild type. Plate counts of both strains were undertaken at the point of harvesting prior to binding assays.

For assessment using GFP^{UL}_{TP901-17}, 150 μL of exponential-phase *L. lactis* 3107-pNZ44 was harvested and resuspended in 150 μL of SM buffer. A serial dilution was made in SM buffer to a theoretical CFU/mL of 1. To do this, 150 μL of undiluted resuspended *L. lactis* UC509.9 wild type and 150 μL of each serial dilution were mixed to generate eight samples with a sequential 10-fold reduction in target concentration and with a constant nontarget cell concentration. Binding assays were then performed as described above.

In the case of GFP^{RBP}_{LMDV}, bulk starter culture was inoculated into GM17 and MRS broth separately and incubated at 30°C until it reached an OD₆₀₀ \approx 0.2 to 0.3. Cells were harvested, resuspended in SM buffer, and mixed in equal volumes to 300 μL prior to binding assays. In the case of GFP^{RBP}_{BB4_27}, bulk starter was cultured in GM17 as described above.

In the case of ^{GFP}Rv1_{423phi1}, the presumed bifidobacterial component (having been enriched as described) of the T60III fecal sample was cultured in RCM as described above until it reached an OD₆₀₀ ≈ 0.2 before binding assays.

Flow cytometry. GFP-RBP fusion protein-labeled bacteria were analyzed and sorted on a BD ARIA Fusion cell sorter using 488-nm laser for GFP excitation and 530/30 band pass detector filter. To minimize electronic noise and detect relatively small bacterial cells, adjustments were made as follows: side scatter and forward scatter signals were acquired with amplifiers set to logarithmic scale, and threshold level was set for side scatter and adjusted to minimize noise as much as possible. Samples were sorted with 85- μ m nozzle, fluidics pressure at 47 lb/in² using drop delay of 31.04 μ s, and DIVA software sorting precision protocol 16-16-0. One hundred thousand events were recorded per sample. Data analysis was performed using FlowJo (59).

Metavirome extraction, sequencing, and analysis. Whey phageome sequencing was performed as follows. NaCl was added to 1 ml of whey sample (to a final concentration of 1 M) which was incubated at 4°C for at least 1 h. The pH was adjusted to 4.6 to clarify, and the solution was centrifuged for 15 min at 28,000 \times g. Polyethylene glycol 8000 (PEG8000) (Sigma-Aldrich) was added to a final concentration of 10% (wt/vol), and supernatant was incubated on ice for 1 h. PEG-precipitated phages were harvested by centrifugation at 15,000 \times g for 15 min and resuspended in 1 ml of SM buffer. DNase treatment and viral DNA extraction were as previously described (5). DNA preparation, sequencing, assembly, and read mapping and retrieval were as previously described (5). Gene products of LMD_1_2 were interrogated for similarity to RBPs using HHPred (26) and BLASTP (60), while putative host species for phage contigs were predicted using HostPhinder (9). Phageome sequencing of infant fecal samples was performed as previously described (5). Quality filtering and classification of reads were performed as described in supplementary method 5 in Text S1.

Fecal sample collection and processing. A fecal sample was collected from a 30-day-old, full-term, breast-fed infant. A fecal sample was collected in a sterile container, immediately frozen at -20°C prior to processing (see supplementary method 5 in Text S1). The study was approved by the Regional Ethical Committee of Asturias Public Health Service (reference 12/16), and informed written consent was obtained from the parents.

Genome sequencing. Genome sequences of 18 *Bifidobacterium* isolates from the T60III sample were determined as previously described (61) using a Illumina NextSeq 550 platform with NextSeq 500/550 v2.5 High Output kit v2.5 300 cycles, and assembled and annotated according to supplementary method 6 in Text S1.

16S rRNA and ITS sequence-based microbiota determination. Bacterial DNA was used as the template for partial 16S rRNA gene PCR amplification, followed by Illumina MiSeq sequencing as previously described (62). Presumed bifidobacterial DNA (extracted from the enriched bifidobacterial portion of the T60III fecal sample; see supplementary method 7 in Text S1) was used as the template for partial ITS PCR amplification, followed by Illumina MiSeq sequencing as previously described (37, 63).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.1 MB.

FIG S1, EPS file, 2.3 MB.

ACKNOWLEDGMENTS

This research was supported by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) to D.V.S. (in conjunction with Science Foundation Ireland [SFI], grant 15/JPI-HDHL/3280) and to M.V. (in conjunction with MIUR, Italy). D.V.S. is a member of APC Microbiome Ireland funded by SFI through the Irish Government's National Development Plan (grants SFI/12/RC/2273a and SFI/12/RC/2273b). J.M. is supported by an SFI-funded Starting Investigator Research Grant (SIRG) (reference number 15/SIRG/3430). S.A. was the recipient of a postdoctoral Juan de la Cierva Contract (Ministry of Science, Innovation and Universities, reference IJCI-2017-32156).

This publication has emanated from research supported in part by a grant from Science Foundation Ireland under grants (15/JPI-HDHL/3280, SFI/12/RC/2273a, SFI/12/RC/2273b, and 15/SIRG/3430). For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

E.C., B.M., K.W., P.S., T.C., and S.A. performed the experiments. E.C., P.S., S.H., G.A.L., M.V., K.N., I.O., I.M., M.G., J.M., and D.V.S. designed experiments and/or interpreted the results. E.C., P.S., B.M., J.M., and D.V.S. wrote the manuscript.

Ines Martinez and Fabio dal Bello are employees of Sacco Srl.

REFERENCES

- Lillehaug D. 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. *J Appl Microbiol* 83:85–90. <https://doi.org/10.1046/j.1365-2672.1997.00193.x>.
- Terzaghi BE, Sandine W. 1975. Improved medium for lactic streptococci and their bacteriophages. *Appl Microbiol* 29:807–813. <https://doi.org/10.1128/am.29.6.807-813.1975>.
- Adams MH, Lark G. 1950. Mutation to heat resistance in coliphage T5. *J Immunol* 64:335–347.
- Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, Ryan FJ, Draper LA, Plevy SE, Ross RP, Hill C. 2019. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microbe* 26:764–778.e5. <https://doi.org/10.1016/j.chom.2019.10.009>.
- Milani C, Casey E, Lugli GA, Moore R, Kaczorowska J, Feehily C, Mangifesta M, Mancabelli L, Duranti S, Turrioni F, Bottacini F, Mahony J, Cotter PD, McAuliffe FM, van Sinderen D, Ventura M. 2018. Tracing mother-infant transmission of bacteriophages by means of a novel analytical tool for shotgun metagenomic datasets: METAnnotatorX. *Microbiome* 6:145. <https://doi.org/10.1186/s40168-018-0527-z>.
- Tikhe CV, Husseneder C. 2017. Metavirome sequencing of the termite gut reveals the presence of an unexplored bacteriophage community. *Front Microbiol* 8:2548. <https://doi.org/10.3389/fmicb.2017.02548>.
- Edwards RA, Vega AA, Norman HM, Ohaeri M, Levi K, Dinsdale EA, Cinek O, Aziz RK, McNair K, Barr JJ, Bibby K, Brouns SJJ, Cazares A, de Jonge PA, Desnues C, Díaz Muñoz SL, Fineran PC, Kurilshikov A, Lavigne R, Mazankova K, McCarthy DT, Nobrega FL, Reyes Muñoz A, Tapia G, Trefault N, Tyakht AV, Vinuesa P, Wagemans J, Zhernakova A, Aarestrup FM, Ahmadov G, Alassaf A, Anton J, Asangba A, Billings EK, Cantu VA, Carlton JM, Cazares D, Cho G-S, Condeff T, Cortés P, Cranfield M, Cuevas DA, De la Iglesia R, Decewicz P, Doane MP, Dominy NJ, Dziewit L, Elwasila BM, Eren AM, Franz C, et al. 2019. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat Microbiol* 4:1727–1736. <https://doi.org/10.1038/s41564-019-0494-6>.
- Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiol Rev* 40:258–272. <https://doi.org/10.1093/femsre/fuv048>.
- Villaruel J, Kleinheinz KA, Jurtz VI, Zschach H, Lund O, Nielsen M, Larsen MV. 2016. HostPhinder: a phage host prediction tool. *Viruses* 8:116. <https://doi.org/10.3390/v8050116>.
- Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J, Amann R, Sullivan MB. 2013. Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* 15:2306–2318. <https://doi.org/10.1111/1462-2920.12100>.
- Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, Sullivan MB. 2014. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* 513:242–245. <https://doi.org/10.1038/nature13459>.
- Sturino JM, Klaenhammer TR. 2006. Engineered bacteriophage–defence systems in bioprocessing. *Nat Rev Microbiol* 4:395–404. <https://doi.org/10.1038/nrmicro1393>.
- Ainsworth S, Sadvovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, Van Sinderen D. 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. *mBio* 5:e00880-14. <https://doi.org/10.1128/mBio.00880-14>.
- Campanacci V, Veessler D, Lichère J, Blangy S, Sciarra G, Moineau S, Van Sinderen D, Bron P, Cambillau C. 2010. Solution and electron microscopy characterization of lactococcal phage baseplates expressed in *Escherichia coli*. *J Struct Biol* 172:75–84. <https://doi.org/10.1016/j.jsb.2010.02.007>.
- Spinelli S, Veessler D, Bebeacua C, Cambillau C. 2014. Structures and host-adhesion mechanisms of lactococcal siphophages. *Front Microbiol* 5:3. <https://doi.org/10.3389/fmicb.2014.00003>.
- Collins B, Bebeacua C, Mahony J, Blangy S, Douillard FP, Veessler D, Cambillau C, van Sinderen D. 2013. Structure and functional analysis of the host recognition device of lactococcal phage Tuc2009. *J Virol* 87:8429–8440. <https://doi.org/10.1128/JVI.00907-13>.
- Hayes S, Duhoo Y, Neve H, Murphy J, Noben JP, Franz C, Cambillau C, Mahony J, Nauta A, van Sinderen D. 2018. Identification of dual receptor binding protein systems in lactococcal 936 group phages. *Viruses* 10:668. <https://doi.org/10.3390/v10120668>.
- Bielmann R, Habann M, Eugster MR, Lurz R, Calendar R, Klumpp J, Loessner MJ. 2015. Receptor binding proteins of *Listeria monocytogenes* bacteriophages A118 and P35 recognize serovar-specific teichoic acids. *Virology* 477:110–118. <https://doi.org/10.1016/j.virol.2014.12.035>.
- Dieterle ME, Bowman C, Batthyany C, Lanzarotti E, Turjanski A, Hatfull G, Piuri M. 2014. Exposing the secrets of two well-known *Lactobacillus casei* phages, J-1 and PL-1, by genomic and structural analysis. *Appl Environ Microbiol* 80:7107–7121. <https://doi.org/10.1128/AEM.02771-14>.
- Szymczak P, Rau MH, Monteiro JM, Pinho MG, Filipe SR, Vogensen FK, Zeidan AA, Janzen T. 2019. A comparative genomics approach for identifying host-range determinants in *Streptococcus thermophilus* bacteriophages. *Sci Rep* 9:7991. <https://doi.org/10.1038/s41598-019-44481-z>.
- Sumrall ET, Rohrig C, Hupfeld M, Selvakumar L, Du J, Dunne M, Schmelcher M, Shen Y, Loessner MJ. 2020. Glycotyping and specific separation of *Listeria monocytogenes* with a novel bacteriophage protein toolkit. *Appl Environ Microbiol* 86:e00612-20. <https://doi.org/10.1128/AEM.00612-20>.
- Hautefort I, Proença MJ, Hinton JCD. 2003. Single-copy green fluorescent protein gene fusions allow accurate measurement of *Salmonella* gene expression in vitro and during infection of mammalian cells. *Appl Environ Microbiol* 69:7480–7491. <https://doi.org/10.1128/AEM.69.12.7480-7491.2003>.
- Legrand P, Collins B, Blangy S, Murphy J, Spinelli S, Gutierrez C, Richet N, Kellenberger C, Desmyter A, Mahony J, van Sinderen D, Cambillau C. 2016. The atomic structure of the phage Tuc2009 baseplate tripod suggests that host recognition involves two different carbohydrate binding modules. *mBio* 7:e01781-15. <https://doi.org/10.1128/mBio.01781-15>.
- Frantzen CA, Kleppen HP, Holo H. 2018. *Lactococcus lactis* diversity in undefined mixed dairy starter cultures as revealed by comparative genome analyses and targeted amplicon sequencing of epsD. *Appl Environ Microbiol* 84:e02199-17. <https://doi.org/10.1128/AEM.02199-17>.
- Pujato SA, Guglielmotti DM, Martínez-García M, Quiberoni A, Mojica FJM. 2017. *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides* bacteriophages: genomics and cross-specific host ranges. *Int J Food Microbiol* 257:128–137. <https://doi.org/10.1016/j.jfoodmicro.2017.06.009>.
- Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248. <https://doi.org/10.1093/nar/gki408>.
- Mahony J, Randazzo W, Neve H, Settanni L, van Sinderen D. 2015. Lactococcal 949 group phages recognize a carbohydrate receptor on the host cell surface. *Appl Environ Microbiol* 81:3299–3305. <https://doi.org/10.1128/AEM.00143-15>.
- Farenc C, Spinelli S, Vinogradov E, Tremblay D, Blangy S, Sadvovskaya I, Moineau S, Cambillau C. 2014. Molecular insights on the recognition of a *Lactococcus lactis* cell wall pellicle by the phage 1358 receptor binding protein. *J Virol* 88:7005–7015. <https://doi.org/10.1128/JVI.00739-14>.
- McDonnell B, Mahony J, Neve H, Hanemaaijer L, Noben JP, Kouwen T, van Sinderen D. 2016. Identification and analysis of a novel group of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*. *Appl Environ Microbiol* 82:5153–5165. <https://doi.org/10.1128/AEM.00835-16>.
- Li X, Koç C, Kühner P, Stierhof Y-D, Krismer B, Enright MC, Penadés JR, Wolz C, Stehle T, Cambillau C, Peschel A, Xia G. 2016. An essential role for the baseplate protein Gp45 in phage adsorption to *Staphylococcus aureus*. *Sci Rep* 6:26455. <https://doi.org/10.1038/srep26455>.
- Kropinski AM, Waddell T, Meng J, Franklin K, Ackermann H-W, Ahmed R, Mazzocco A, Yates J, Lingohr EJ, Johnson RP. 2013. The host-range, genomics and proteomics of *Escherichia coli* O157: H7 bacteriophage rV5. *Virol J* 10:76. <https://doi.org/10.1186/1743-422X-10-76>.
- Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sørensen SJ, Hammer K, Cambillau C, Vogensen FK, van Sinderen D. 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. *Appl Environ Microbiol* 79:4385–4392. <https://doi.org/10.1128/AEM.00653-13>.
- Murphy J, Bottacini F, Mahony J, Kelleher P, Neve H, Zomer A, Nauta A, Van Sinderen D. 2016. Comparative genomics and functional analysis of the 936 group of lactococcal *Siphoviridae* phages. *Sci Rep* 6:21345. <https://doi.org/10.1038/srep21345>.
- Gopal PK, Davey GP, Crow VL, Pillidge CJ, Coolbear T, Holland R. 1996. Extracellular material from *Leuconostoc mesenteroides* subsp. *dextranicum* mediates the aggregation of lactococcal cells: implications for the isolation of single strains. *J Appl Bacteriol* 81:48–56. <https://doi.org/10.1111/j.1365-2672.1996.tb03281.x>.
- Hemme D, Foucaud-Scheunemann C. 2004. *Leuconostoc*, characteristics, use in dairy technology and prospects in functional foods. *Int Dairy J* 14:467–494. <https://doi.org/10.1016/j.idairyj.2003.10.005>.

36. Duranti S, Lugli GA, Mancabelli L, Armanini F, Turrioni F, James K, Ferretti P, Gorfer V, Ferrario C, Milani C, Mangifesta M, Anzalone R, Zolfo M, Viappiani A, Pasolli E, Bariletti I, Canto R, Clementi R, Cologna M, Crifo T, Cusumano G, Fedi S, Gottardi S, Innamorati C, Masè C, Postai D, Savoì D, Soffiati M, Tateo S, Pedrotti A, Segata N, van Sinderen D, Ventura M. 2017. Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* 5:66. <https://doi.org/10.1186/s40168-017-0282-6>.
37. Lugli GA, Milani C, Mancabelli L, van Sinderen D, Ventura M. 2016. MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS Microbiol Lett* 363:fnw049. <https://doi.org/10.1093/femsle/fnw049>.
38. Ventura M, Turrioni F, Lima-Mendez G, Foroni E, Zomer A, Duranti S, Giubellini V, Bottacini F, Horvath P, Barrangou R, Sela DA, Mills DA, van Sinderen D. 2009. Comparative analyses of prophage-like elements present in bifidobacterial genomes. *Appl Environ Microbiol* 75:6929–6936. <https://doi.org/10.1128/AEM.01112-09>.
39. Mavrich TN, Casey E, Oliveira L, Bottacini F, James K, Franz C, Lugli GA, Neve H, Ventura M, Hatfull GF, Mahony J, van Sinderen D. 2018. Characterization and induction of prophages in human gut-associated *Bifidobacterium* hosts. *Sci Rep* 8:12772. <https://doi.org/10.1038/s41598-018-31181-3>.
40. Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. 2014. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin Exp Allergy* 44:842–850. <https://doi.org/10.1111/cea.12253>.
41. Penders J, Thijs C, van den Brandt PA, Kummeling I, Snijders B, Stelma F, Adams H, van Ree R, Stobberingh EE. 2007. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. *Gut* 56:661–667. <https://doi.org/10.1136/gut.2006.100164>.
42. Shen X, Wang M, Zhang X, He M, Li M, Cheng G, Wan C, He F. 2019. Dynamic construction of gut microbiota may influence allergic diseases of infants in Southwest China. *BMC Microbiol* 19:123. <https://doi.org/10.1186/s12866-019-1489-4>.
43. Milani C, Duranti S, Bottacini F, Casey E, Turrioni F, Mahony J, Belzer C, Delgado Palacio S, Arboleya Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M, Margolles A, van Sinderen D, Ventura M. 2017. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev* 81:e00036-17. <https://doi.org/10.1128/MMBR.00036-17>.
44. Leblond-Bourget N, Philippe H, Mangin I, Decaris B. 1996. 16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific *Bifidobacterium* phylogeny. *Int J Syst Bacteriol* 46:102–111. <https://doi.org/10.1099/00207713-46-1-102>.
45. Del Re B, Sgorbati B, Miglioli M, Palenzona D. 2000. Adhesion, autoaggregation and hydrophobicity of 13 strains of *Bifidobacterium longum*. *Lett Appl Microbiol* 31:438–442. <https://doi.org/10.1046/j.1365-2672.2000.00845.x>.
46. Vlkova E, Rada V, Smehilova M, Killer J. 2008. Auto-aggregation and co-aggregation ability in bifidobacteria and clostridia. *Folia Microbiol (Praha)* 53:263–269. <https://doi.org/10.1007/s12223-008-0040-z>.
47. Andriantsoanirina V, Teolis AC, Xin LX, Butel MJ, Aires J. 2014. *Bifidobacterium longum* and *Bifidobacterium breve* isolates from preterm and full term neonates: comparison of cell surface properties. *Anaerobe* 28:212–215. <https://doi.org/10.1016/j.anaerobe.2014.07.002>.
48. Rahman MM, Kim W-S, Kumura H, Shimazaki K. 2008. Autoaggregation and surface hydrophobicity of bifidobacteria. *World J Microbiol Biotechnol* 24:1593–1598. <https://doi.org/10.1007/s11274-007-9650-x>.
49. Goulet A, Cambillau C. 2021. Structure and topology prediction of phage adhesion devices using AlphaFold2: the case of two *Oenococcus oeni* phages. *Microorganisms* 9:2151. <https://doi.org/10.3390/microorganisms9102151>.
50. Hayes S, Mahony J, Vincentelli R, Ramond L, Nauta A, Sinderen D, Cambillau C. 2019. Ubiquitous carbohydrate binding modules decorate 936 lactococcal siphophage virions. *Viruses* 11:631. <https://doi.org/10.3390/v11070631>.
51. Hayes S, Vincentelli R, Mahony J, Nauta A, Ramond L, Lugli GA, Ventura M, van Sinderen D, Cambillau C. 2018. Functional carbohydrate binding modules identified in evolved dits from siphophages infecting various Gram-positive bacteria. *Mol Microbiol* 110:777–795. <https://doi.org/10.1111/mmi.14124>.
52. Zhang Y, Delbrück AI, Off CL, Benke S, Mathys A. 2020. Flow cytometry combined with single cell sorting to study heterogeneous germination of *Bacillus* spores under high pressure. *Front Microbiol* 10:3118. <https://doi.org/10.3389/fmicb.2019.03118>.
53. Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* 73:7059–7066. <https://doi.org/10.1128/AEM.00358-07>.
54. Meier DV, Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, Amann R, Meyerdierks A. 2017. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *ISME J* 11:1545–1558. <https://doi.org/10.1038/ismej.2017.37>.
55. Vigil-Stenman T, Ininbergs K, Bergman B, Ekman M. 2017. High abundance and expression of transposases in bacteria from the Baltic Sea. *ISME J* 11:2611–2623. <https://doi.org/10.1038/ismej.2017.114>.
56. Zhang L, Calvo-Bado L, Murray AK, Amos GCA, Hawkey PM, Wellington EM, Gaze WH. 2019. Novel clinically relevant antibiotic resistance genes associated with sewage sludge and industrial waste streams revealed by functional metagenomic screening. *Environ Int* 132:105120. <https://doi.org/10.1016/j.envint.2019.105120>.
57. Guo J, Li J, Chen H, Bond PL, Yuan Z. 2017. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Res* 123:468–478. <https://doi.org/10.1016/j.watres.2017.07.002>.
58. Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
59. FlowJo LLC. FlowJo v10.0.7r2 (2014). <https://www.flowjo.com>.
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
61. Lugli GA, Duranti S, Albert K, Mancabelli L, Napoli S, Viappiani A, Anzalone R, Longhi G, Milani C, Turrioni F, Alessandri G, Sela DA, van Sinderen D, Ventura M. 2019. Unveiling genomic diversity among members of the species *Bifidobacterium pseudolongum*, a widely distributed gut commensal of the animal kingdom. *Appl Environ Microbiol* 85:e03065-18. <https://doi.org/10.1128/AEM.03065-18>.
62. Milani C, Hevia A, Foroni E, Duranti S, Turrioni F, Lugli GA, Sanchez B, Martin R, Gueimonde M, van Sinderen D, Margolles A, Ventura M. 2013. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* 8:e68739. <https://doi.org/10.1371/journal.pone.0068739>.
63. Milani C, Lugli GA, Turrioni F, Mancabelli L, Duranti S, Viappiani A, Mangifesta M, Segata N, van Sinderen D, Ventura M. 2014. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS Microbiol Ecol* 90:493–503. <https://doi.org/10.1111/1574-6941.12410>.
64. Kuipers OP, de Ruyter PGG, Kleerebezem M, de Vos WM. 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. *J Biotechnol* 64:15–21. [https://doi.org/10.1016/S0168-1656\(98\)00100-X](https://doi.org/10.1016/S0168-1656(98)00100-X).
65. Braun V, Jr, Hertwig S, Neve H, Geis A, Teuber M. 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. *Microbiology* 135:2551–2560. <https://doi.org/10.1099/00221287-135-9-2551>.
66. Ainsworth S, Zomer A, de Jager V, Bottacini F, van Hijum SAFT, Mahony J, van Sinderen D. 2013. Complete genome of *Lactococcus lactis* subsp. *cremoris* UC509.9, host for a model lactococcal P335 bacteriophage. *Genome Announc* 1:e00119-12. <https://doi.org/10.1128/genomeA.00119-12>.
67. McGrath S, Fitzgerald GF, van Sinderen D. 2001. Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. *Appl Environ Microbiol* 67:608–616. <https://doi.org/10.1128/AEM.67.2.608-616.2001>.