nature communications

Article

https://doi.org/10.1038/s41467-026-79230-0

Population-level characterisation of the human gut phageome and its association with microbial metabolites

Received: 10 February 2026

Accepted: 15 October 2026

Published online: 17 November 2026

Check for updates

A list of authors and their affiliations appears at the end of the paper

Bacteriophages, viruses infecting bacteria, are abundant in the human gut, but their population-level structure and metabolic associations remain unclear Here we present a foundational characterisation of the gut phageome using deep shotgun metagenomic data in a large population-based cohort (n = 1,069: TwinsUK). We assembled 116,740 non-redundant vOTUs of which 52.9% were predicted to be temperate and 47.0% virulent. Individuals harboured an average of 3,370 vOTUs, mostly at low abundances, exhibiting structured population-level co-abundance patterns with bacterial species including *Faecalibacterium prausnitzii* and *Blautia wexlerae*. Phageome diversity was mostly driven by environmental factors, with modest heritable components. We detected 4 significant vOTU diet associations. Additionally, 6,572 phages were associated with short-chain fatty acids and secondary bile acids, identifying extensive population-level associations between phage-associated genomic variation and microbial metabolites. Over 100 vOTUs encoded auxiliary metabolic genes, while a distinct subset of phages encoded virulence-associated genes, including vacuolating autotransporter toxin (vat) and LPS glucosyltransferases (gtrA, gtrB, gtr) implicated in O-antigen modification and immune invasion, and showed consistent co-abundance patterns with bacterial species previously associated with opportunistic or pathobiont features. These results identify structured population-level associations between phage-associated genomic elements, bacterial communities, and microbial metabolites, motivating future experimental and mechanistic studies.

A growing body of evidence highlights the importance of viruses infecting gut bacteria—bacteriophages (phages)—in modulating microbial ecology and host-associated functions[1]. Indeed, phages are thought to contribute to microbial community dynamics by specifically targeting their hosts and transferring genes that can change bacterial traits and community functional potential[2–4]. Recent work has further highlighted the prevalence of capsid-mediated mobilisation of bacterial DNA in the human gut, particularly Faecalibacterium and Bacteroides[5].

Though recent studies have identified associations between the gut phageome and metabolic diseases[6,7], large-scale investigations into the composition of the gut phageome and its interactions with host factors remain scarce. The precise mechanisms through which phages influence host metabolism, phenotypic heterogeneity, and ecosystem-level functions remain poorly understood[8]. Current studies are constrained by small sample sizes, heterogeneous viral identification pipelines, and the need for deep metagenomic sequencing or viral-like particle isolation[9,10].

e-mail: cristina.menni@kcl.ac.uk

Recent population-scale metagenomic analyses indicate that viral diversity in the human gut is substantially greater than previously appreciated[11]. Mining tens of thousands of gut metagenomes has yielded large catalogues of non-redundant phage genomes, the majority of which show little similarity to existing reference databases, revealing extensive viral "dark matter"[11]. The recently released Unified Human Gastrointestinal Virome (UHGV) catalogue represents the largest such catalogue to date, comprising nearly 874,000 viral genomes with improved host predictions and annotations[12]. Comparative analyses further demonstrate that most viral sequences are unique to individual catalogues and that aggregating datasets across cohorts continues to uncover large numbers of previously unseen viral taxa, indicating that gut viral diversity remains far from saturated[13]. Moreover, the expansion of viral reference catalogues has substantially improved the inference of virus-host associations, viral functional repertoires (virulence factors (VFs) and auxiliary metabolic genes (AMGs)), and ecological network structure, providing new insight into virus-associated patterns within microbial community organisation and functional potential in the human gut. In parallel, advances in computational tools[10,14], are facilitating population-scale investigations into the phageome's relationship to host-microbe interactions.

The aim of this study is to systematically map the human gut phageome in a large, deeply phenotyped, population-based cohort by integrating multi-omics data. Specifically, we investigate (i) phageome heritability; (ii) phage-bacterial correlation structure and network organisation; (iii) associations between phages, microbial metabolites and habitual diet; and (iv) phage-encoded auxiliary metabolic and virulence functions, and their associations with bacterial species. We identify key phages associated with bacterial nodes and metabolites, highlighting the gut phageome as an underappreciated component of the human gut microbiome that shows structured co-abundance relationships with microbial network structure and metabolic variation at the host–microbe interface.

Results

A schematic of the study design is presented in Fig. 1.

Study population

We included 1,069 participants from the TwinsUK cohort, predominantly female ($n = 903$, 84.5%), with a mean age of 58.6 years (SD = 14.8) and BMI of 26.2 kg/m^2 (SD = 5.2) (Table 1).

To characterise the gut phageome, we assembled viral sequences and dereplicated them into 116,740 non-redundant viral operational taxonomic units (vOTUs) using MIUViG-consistent criteria (95% average nucleotide identity and ≥85% alignment fraction). Quality-controlled metagenomic reads (~30 Gb per sample) were mapped to the vOTU catalogue[15], and vOTUs were considered present only if they achieved ≥20% genome coverage and a mean sequencing depth of at least 0.5× following post-mapping quality control[16]. Consistent with the largely unexplored nature of the human gut virome, vOTUs lacked close matches to the UHGV[12], revealing extensive phage diversity not represented in current reference collections (868/116,740 matches, 0.74%). vOTUs were detected across all samples, with participants harbouring on average 3,370 vOTUs (SD = 1,979) (Supplementary Fig. 1). Across the full catalogue, 41,431 vOTUs were present in ≥1% of participants; of these, 8,970 were also present at an average relative abundance > 0.001% and taken forward for downstream analyses. Among these, the majority (92.5)% were assigned to the class *Caudoviricetes*. Phage lifestyle prediction further indicated that 67.3% were temperate, and 32.5% were virulent (Fig. 2).

Phageome heritability

Given that bacteriome diversity and composition have been shown to have a heritable component[17,18], we leveraged the twin nature of our data to estimate the heritability of phageome diversity and composition using structural equation modelling, adjusting for age, sex, and BMI. We decomposed phenotypic variance into additive genetic (A; host genetic influence), common environmental (C; environment shared by both twins, e.g., early life exposures), and unique environmental (E; individual-specific environmental factors) components, fitting ACE, AE, CE, and E models and selecting the best-fitting model based on the Akaike Information Criterion (AIC).

Phageome alpha diversity, as measured by the Shannon index, was best explained by an AE model, indicating a substantial additive genetic contribution (A = 46.8%, 95% confidence interval (CI): 38.3–55.2%; E = 53.2%, 95% CI: 44.8–61.7%). In contrast, taxonomic richness was best fit by a CE model, suggesting a greater influence of shared environmental factors rather than host genetics (C = 34.4%, 95% CI: 25.8–43.1%; E = 65.6%, 95% CI: 56.9–74.2%) (Supplementary Table 1). The divergent heritability estimates may reflect differences in the ecological properties captured by these diversity metrics, with Shannon diversity incorporating both richness and evenness, whereas richness-based metrics primarily reflect taxon presence.

For beta diversity, decomposed into the first two principal coordinates (Supplementary Fig. 2), the first axis (PCo1) of Bray–Curtis dissimilarity was best explained by an AE model, with narrow-sense heritability of 34.8% (95%CI: 24.8–44.8%) and unique environmental effects accounting for the remaining 65.2% (55.2–75.3%). In contrast, the second axis (PCo2) showed negligible heritability, with variance explained by primarily shared (C = 21.3%, 11.9–30.6%) and unique environmental factors (E = 78.7%, 69.4–88.1%) (Supplementary Table 2).

Across 8,970 vOTU (Supplementary Table 3), 2073 (23.1%) were best explained by models including a heritable component (A), accounting on average for 31.3% (95%CI: 21.2–42.6%) of the variance. In contrast, 5,452 vOTUs fitted best to the unique environmental model; 1,445 were best explained by the CE model (C = 14.3%, 4.6–23.9%), likely reflecting individual-specific influences (Supplementary Fig. 3 and Supplementary Table 4).

Together, these results indicate that while overall phageome diversity is predominantly environmentally determined, a substantial subset of individual vOTUs exhibit measurable host genetic influence.

Correlations between phages and bacterial species

To investigate population-level co-abundance relationships between the gut phageome and bacteriome, we next examined concordance in community diversity and composition, followed by species-level phage–bacteria associations.

Alpha diversity. Positive correlations were observed between phage and bacterial alpha diversity (measured by Shannon index and taxonomic richness) (Kendall's $\tau = 0.37$ and 0.19, respectively; $p = \leq 2.0 \times 10^{-16}$) (Supplementary Table 5). These results indicate that more diverse phage communities are generally associated with more diverse bacterial communities, consistent with previous work in a small cohort of 21 twin pairs[19], consistent with ecological interdependencies between phages and their bacterial hosts at the population level.

Beta Diversity. Procrustes analysis of Bray-Curtis distances between the phageome (116,740 vOTUs) and bacteriome (all 2,233 species) revealed strong concordance in community structure (Procrustes correlation = 0.83, $p = 0.001$), with approximately 69% of the variance explained by the Procrustes fit ($m^2 = 0.309$). This indicates that global patterns of β-diversity are broadly shared between phage and bacterial communities while a residual ~31% of the variance may reflect phage-specific compositional variation not fully captured by bacterial structure. Results were consistent when using Jaccard distances ($m^2 = 0.308$, $p = 0.001$).

https://doi.org/10.1038/s41467-026-79230-0

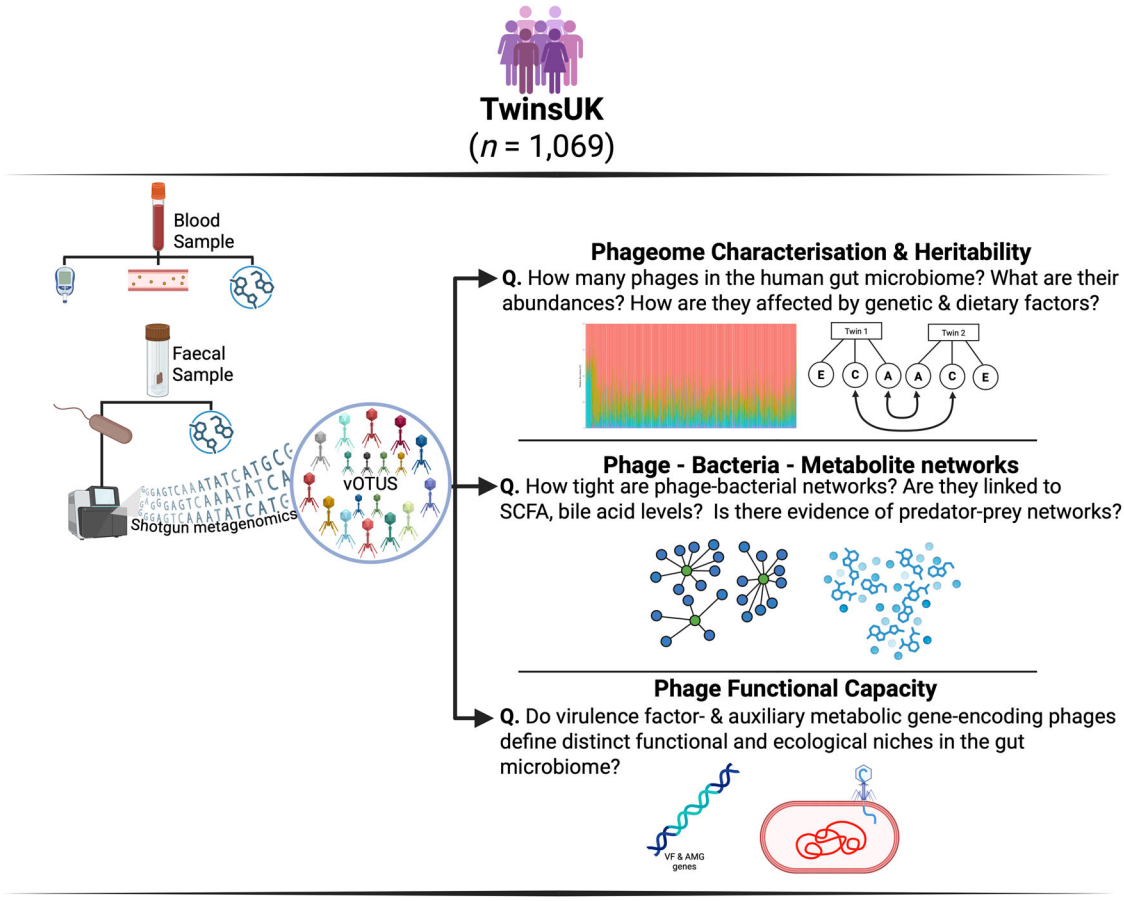


Fig. 1 | Study design and analytical framework for gut phageome characterisation. Schematic overview of the sampling strategy, data generation, and analytical questions. Stool and blood samples were collected from TwinsUK participants for metagenomic and metabolomic profiling. vOTUs were assembled from shotgun metagenomes and characterised to assess phage diversity, abundance, and lifestyle, followed by analyses of phageome heritability using the twin design. Correlation-based network analyses were used to visualise phage–bacterial co-abundance patterns and to examine associations between phages, bacterial taxa, microbial metabolites (including short-chain fatty acids and bile acids), and habitual dietary intake. Functional annotation of sequences facilitated investigation of whether virulence-factor- and auxiliary metabolic gene-encoding vOTUs occupy distinct niches within the gut microbiome. SCFA short-chain fatty acids, VF virulence factors, AMG auxiliary metabolic genes. Created with BioRender. *Menni, C. (2026).*

Table. 1 | Cohort Demographics and Clinical Characteristics

	n	**%**
n	1,069	
MZ twin pairs	257	48.1
DZ twin pairs	147	27.5
Singletons	261	24.4
Female participants	903	84.5
	mean	***sd***
Age, years	58.6	14.8
BMI, kg/m^2	26.2	5.2
Phageome Shannon diversity	7.0	0.9
Phageome observed taxonomic richness	3,369.5	1,979.3
Bacteriome Shannon diversity	3.7	0.5
Bacteriome observed taxonomic richness	246.5	57.5

MZ, Monozygotic twins; DZ, Dizygotic twins; BMI, body mass index.

Species-level bacterial associations. To resolve phage–bacteria relationships at higher taxonomic resolution, we examined correlations between 8,970 vOTUs and 274 bacterial species using Spearman correlation. We identified 1,126,841 (45.9%) significant associations (FDR < 0.01). All 8,970 vOTUs were significantly associated with at least one bacterial species. Of these, 856,098 (76.0%) were positive and 270,743 (24.0%) were negative. As expected, positive associations were modestly stronger than negative (median rho 0.145 vs. 0.126, respectively), consistent with shared population-level structure, prophage carriage, or other sources of covariance. Negative associations may reflect indirect ecological or compositional processes. Results were robust to alternative model approaches with consistent findings when using linear mixed models accounting for family as a random effect for all but one phage (8,969 vOTUs, FDR < 0.01).

To assess whether observed phage-bacterial species associations were confounded by phage-host co-abundance, we performed conditional analyses adjusting for the predicted host genera inferred using iPHoP[16]. Of the 8,970 vOTUs, 6,013 (67%) had a predicted host genus that was both annotated in our metagenome data and detected in ≥ 10% of samples, yielding 662,429 testable phage-species pairs. After conditional analyses, 525,291 associations remained significant (79.3%). Importantly, 6,012 vOTUs continued to show significant associations with at least one bacterial species (FDR < 0.01), suggesting that many observed correlations are not fully accounted for by predicted host assignments, although shared metagenomic origin and other sources of dependence remain important considerations. Consistent with this, a complementary analysis directly testing vOTU-iPHoP-predicted host genus associations showed that

https://doi.org/10.1038/s41467-026-79230-0

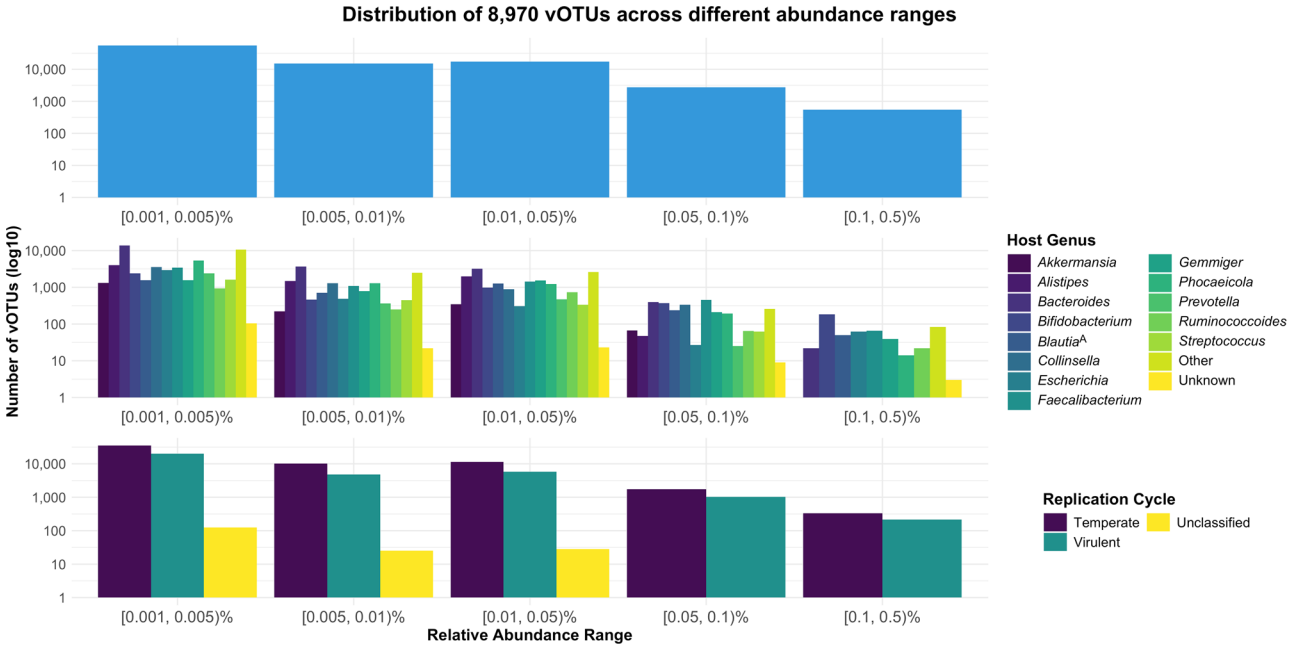


Fig. 2 | Gut phageome composition by relative abundance and predicted phage characteristics. Distribution of 8,970 vOTUs with a prevalence of > 1% and mean relative abundance of > 0.001% identified in 1069 individuals, grouped by relative abundance ranges and coloured by the predicted host genus, or the predicted phage replication cycle. Abundance ranges (x axis) are shown using interval notation ("[" indicates inclusion, and ")" exclusion). Relative abundance represents the proportion of each vOTU to the gut phageome (all detected vOTUs).

only 72.8% of vOTUs were significantly correlated with their predicted hosts (FDR < 0.01).

Co-abundance Network of phages and bacteria. Given the extensive and structured phage–bacteria associations observed, we constructed bipartite correlation networks to identify higher-order association structure. One comprised positive correlations (Spearman's Rho > 0.4), including 2,445 unique vOTUs (67.3% temperate, 32.5% virulent), and a second comprised negative correlations (Spearman's Rho < −0.3), including 557 vOTUs (69.3% temperate, 30.7% virulent) (Fig. 3).

(i) Positive correlation network. Bipartite network analysis revealed a densely interconnected phageome–bacteriome landscape, with several bacterial taxa acting as central hubs (Fig. 3A). *Faecalibacterium prausnitzii* exhibited the greatest connectivity, co-abundant with 172 vOTUs, followed by *Anaerobutyricum hallii* (167 vOTUs) and *Blautia wexlerae* (161 vOTUs).

(ii) Negative correlation network. The network of negative associations similarly highlighted *F. prausnitzii* as one of the most connected bacterial species (111 vOTUs). The most connected taxa included *Alistipes communis* (125 vOTUS), followed by *B. wexlerae* (119 vOTUs) and *Alstipes putredinis* (119 vOTUs). The Fruchterman–Reingold layout revealed distinct clusters centred on these bacterial hubs (Fig. 3B), suggesting that subsets of phages co-vary with key bacterial taxa, although these patterns may reflect indirect biological or genomic relationships rather than direct ecological interactions.

Associations between phages and bacterial metabolites

Short-chain fatty acids (SCFAs) and secondary bile acids (BAs) are key microbial-derived metabolites that serve as important readouts of bacterial metabolic activity. Given the close correspondence between bacterial composition, microbial metabolism, and viral abundance, we examined associations between 8,970 vOTUs and serum and faecal levels of acetate, butyrate, and propionate, as well as 23 secondary BAs (11 serum, 12 stool), adjusting for age, sex, BMI, family structure, and batch effects.

We identified 6,572 unique vOTUs that were significantly associated with at least one SCFA or secondary BA (FDR < 0.01; Supplementary Table 6). These associations were predominantly observed with faecal metabolite abundances, whereas associations with serum metabolites were comparatively limited, as expected, given that there is no clear mechanism for gut phages to interact with serum metabolites.

To visualise the structure and polarity of these associations, we constructed a correlation network linking bacterial metabolites and the associated vOTUs (Fig. 4). The network highlights densely connected hubs around faecal acetate levels, highlighting co-abundance of phages with microbial fermentation products within the gut. In addition, interconnected hubs around faecal secondary BAs, including ursocholate, isoursodeoxycholate, 7-ketodeoxycholate and hyocholate, emerge as prominent hubs.

The concentration of viral associations around distinct metabolic hubs suggests that microbial metabolic variation may represent an important axis of population-level co-abundance among phages, bacterial taxa, and metabolites. In contrast, correlations with serum metabolites were sparse and weaker. Notably, isoursodeoxycholate, one of the prominent bile acid hubs identified, has previously been linked to poorer cardiometabolic health and increased appetite, underscoring its potential clinical relevance[20].

As SCFAs and secondary BAs are known to be strongly influenced by bacterial composition[20,21], and as phage associations may indirectly reflect these patterns, we additionally performed a parallel analysis of bacterial species and metabolite associations. Across 274 bacterial species and 29 metabolites, we identified 1,574 significant species–metabolite associations (FDR < 0.01), involving 246 species and 26 metabolites (Supplementary table 7). Associations were most prevalent for faecal secondary bile acids, particularly dehydrolithocholate, isoursodeoxycholate, ursocholate, and 7-ketodeoxycholate, consistent with the established role of gut bacteria in secondary bile acid metabolism.

Of the 6,572 metabolite-associated vOTUs, 3,877 (59%) were predicted to infect a bacterial genus containing at least one species that

https://doi.org/10.1038/s41467-026-79230-0

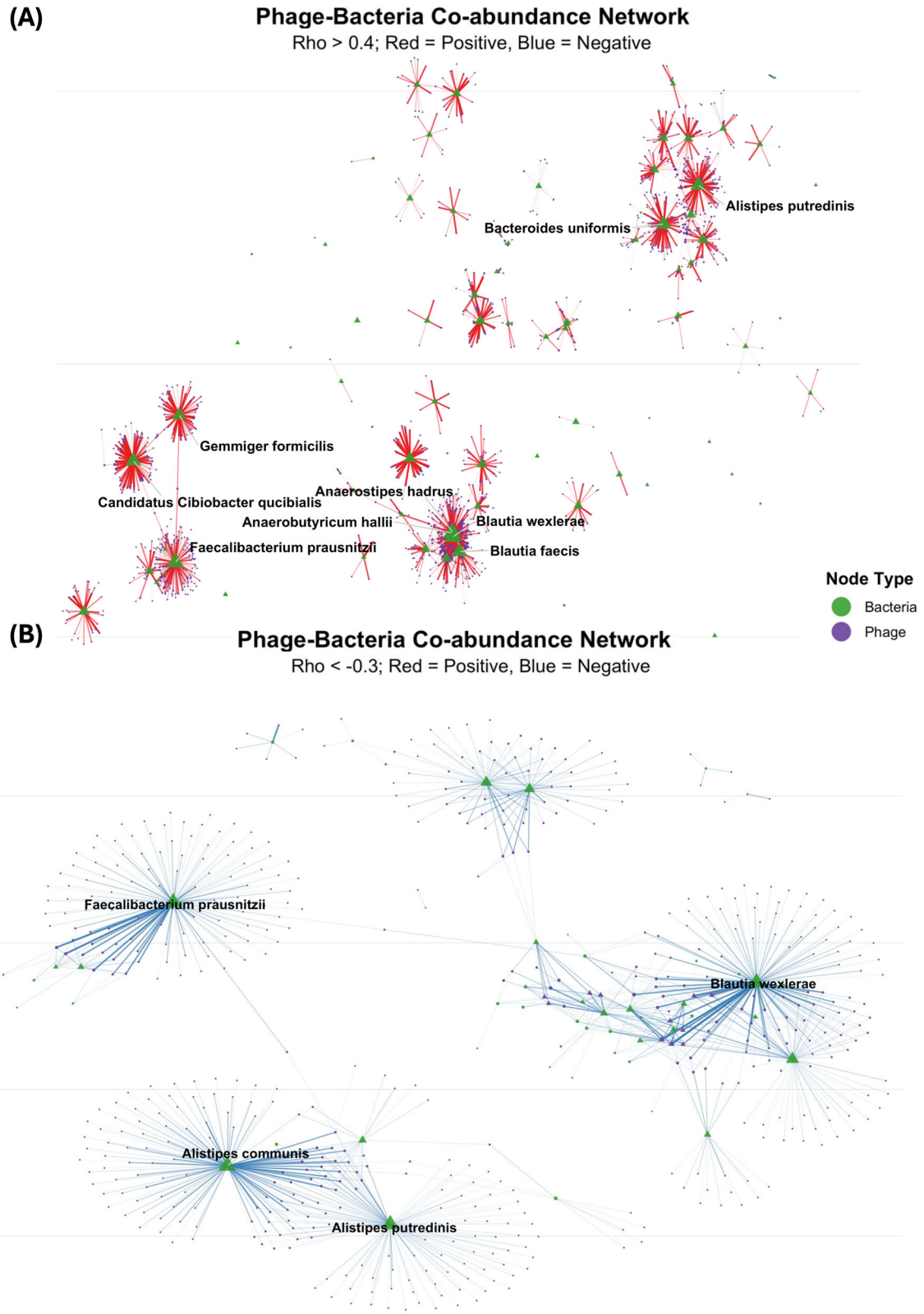


Fig. 3 | Phage-Bacteria correlation networks. Bipartite correlation networks illustrate **A** positive monotonic relationships (Spearman's Rho > 0.4) and **B** moderate negative relationships (Spearman's Rho < −0.3) between vOTUs and bacterial taxa. Nodes represent vOTUs (purple) and bacterial species (green), with node size proportional to degree (number of connections). Bacterial nodes are labelled if they are conne cted to more than 100 vOTUs. Edges indicate correlation direction and strength: red for positive and blue for negative, with edge thickness and transparency scaled to the absolute correlation coefficient. Networks were generated using the Fruchterman–Reingold layout.

https://doi.org/10.1038/s41467-026-79230-0

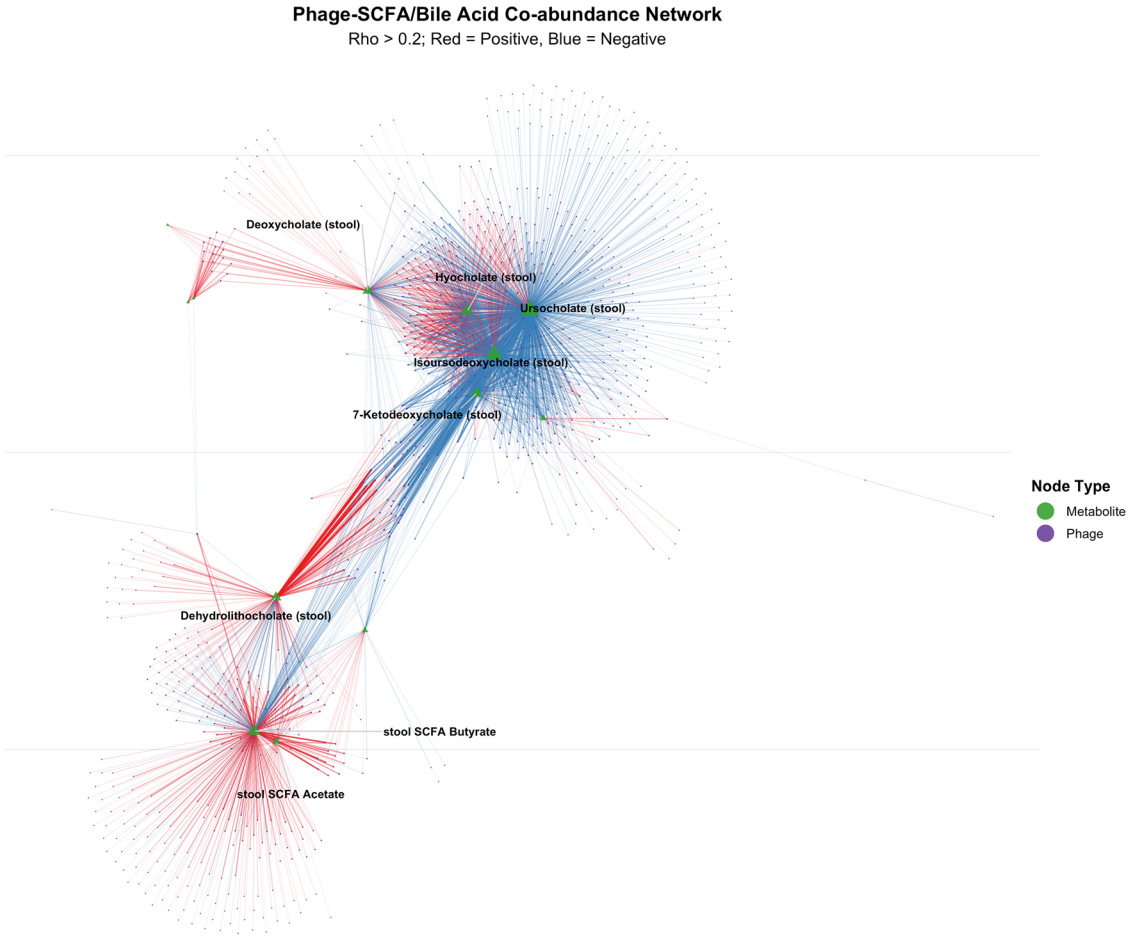


Fig. 4 | Phage-Metabolite co-abundance network. A correlation network (Fruchterman-Reingold layout) illustrating strong monotonic relationships (Spearman's Rho > 0.2) between 950 vOTUs and 18 stool or serum SCFAs and BAs previously identified as significantly associated using linear mixed-effects models (FDR < 0.01). A total of 260,130 models were performed, comprising 53,820 tests for vOTUs – SCFA associations (8,970 vOTUs × 6 SCFA measures: 3 serum and 3 stool) and 206,310 tests for viral contig – bile acid associations (8,970 vOTUs × 23 bile acids). Nodes represent individual vOTUs (purple) and metabolites (green), with node size proportional to degree (number of connections). Edges indicate correlation strength and direction, with red lines representing positive correlations and blue lines representing negative correlations; edge thickness reflects the absolute correlation coefficient. Metabolite nodes are labelled if they are connected to more than 100 vOTUs.

was also significantly associated with the same metabolite, suggesting that a proportion of phage–metabolite associations may partially reflect underlying bacterial composition. However, given that host prediction operates at genus level, this figure represents an upper bound on true host-mediated confounding.

To assess whether species–metabolite associations were independent of viral pressure on the predicted host genus, we repeated the species–metabolite association analyses, additionally adjusting for the summed abundance of phages predicted to infect each species' genus. Of the 854 species–metabolite associations that were significant in the base model and had a mappable phage covariate, 666 (78%) remained significant after phage adjustment (FDR < 0.01), suggesting that the majority of species–metabolite associations are not explained by co-varying phage abundance.

Population-level co-abundance patterns between vOTUs, bacterial taxa and microbial metabolites

To explore potential ecological and functional associations involving phage-bacteria relationships, we selected vOTUs previously identified as strongly correlated with key bacterial hubs and mapped their associations with microbial metabolites (Fig. 5). Specifically, we selected a subset of top-ranking vOTUs with the strongest correlations to microbial metabolites that were also present in the phage-bacteria networks.

The resulting heatmap revealed distinctive patterns of associations: vOTUs in the positive network were frequently associated with lower levels of specific stool BAs, and many were negatively associated with stool butyrate levels. Conversely, phages in the negative network had weaker associations with SCFAs and were positively associated with specific secondary BAs such as ursocholate and isoursodeoxycholate, metabolites previously implicated in metabolic dysregulation[20].

Notably, *B. wexlerae* appears across both positive and negative phage networks and several vOTUs associated with *B. wexlerae* were associated with several metabolites, underscoring its prominent position within the phageome–bacteriome–metabolome axis.

These patterns illustrate structured population-level co-abundance relationships between phage-associated genomic variation, bacterial taxa, and microbial metabolites. These observations are strictly descriptive in nature and do not imply direct mechanistic or causal interactions.

Phageome & dietary intake

Given the observed associations between vOTUs and SCFAs and secondary BAs, we also hypothesise possible links between vOTUs

https://doi.org/10.1038/s41467-026-79230-0

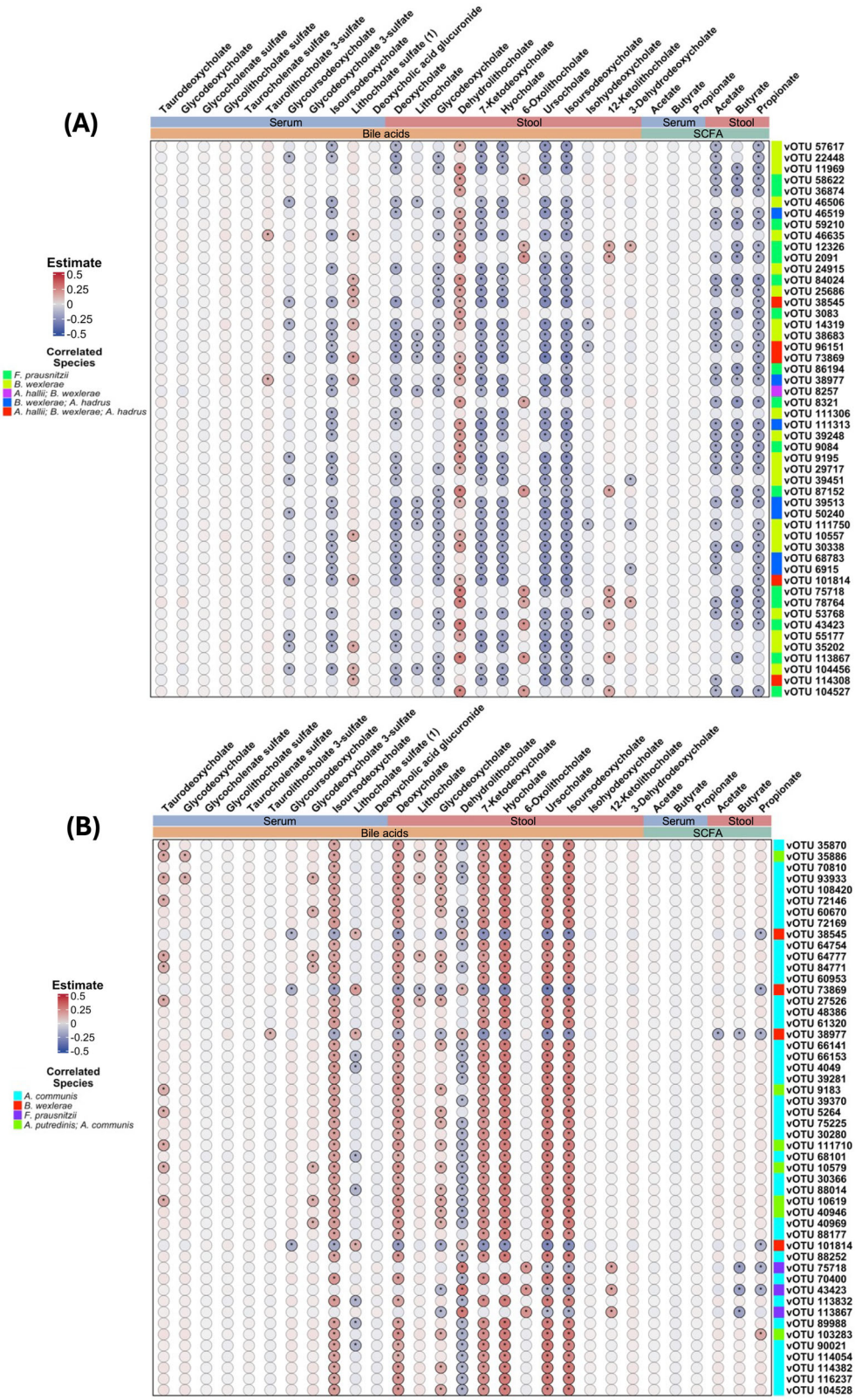


Fig. 5 | Phage – Bacteria – Metabolite Association Patterns. Heatmap of the 50 strongest vOTU-metabolite associations. vOTUs were pre-filtered to include only those with significant correlations to key bacterial nodes **A** *Faecalibacterium prausnitzii, Anaerobutyricum hallii* and *Blautia wexlerae* (Spearman's Rho > 0.4, FDR < 0.01) **B** *Faecalibacterium prausnitzii*, *Alistipes communis, Blautia wexlerae*, and *Alistipes putredinis* (Spearman's Rho < −0.3, FDR < 0.01). From these, vOTU – microbial metabolite associations with FDR < 0.001 were retained and ranked by absolute coefficient values from linear mixed-effects models. Correlated species legend is ordered by the frequency of correlated vOTUs. Non-significant associations are displayed semi-transparently; significant associations are marked with asterisks.

and phage diversity with habitual dietary intake (nutrient intakes and healthy eating index (HEI)). After adjusting for age, sex, family relatedness, batch effects, energy intake, and multiple testing (FDR < 0.01). We identified 4 significant associations and two dietary measures, namely galactose intake (3 positively associated vOTUs) and dietary fibre (1 positively associated vOTU). The strongest positive association was detected for galactose intake (β [95%CI] 0.22 [0.15, 0.28], $p = 3.6 \times 10^{-11}$) (Supplementary Fig. 4, Supplementary Table 8). Collectively, these findings suggest that while certain nutrients are linked to phageome variation, the broad associations between vOTUs and SCFAs or secondary BAs are unlikely to be explained solely by habitual dietary intake.

Functional potential of phages: Auxiliary metabolic genes & Virulence factors

AMGs were identified in a subset of vOTUs. A total of 24 vOTUs (prevalence 1.8–85.7%) encoded at least one AMG and were significantly associated with at least one bacterial taxon (FDR < 0.01) (Fig. 6A). Collectively, AMG-encoding vOTUs were correlated with 265 of the 274 bacterial species, with a median of 9 AMG-containing vOTU associated per species.

The most prevalent AMGs were dominated by functions involved in nucleotide metabolism and nutrient acquisition. The most abundant individual annotations included ribonucleotide reductase ($n = 13$ vOTUs), cellulase/endo-β−1,4-mannosidase ($n = 6$), glycohydrolase family 16 (β−1,3-glucanase) ($n = 5$), and the phosphate transport system substrate-binding protein PstS ($n = 3$). The bacterial species most frequently associated with AMG-encoding vOTUs included *Clostridium innocuum*, *Erysipelatoclostridium ramosum*, *Blautia producta*, and *Clostridium symbiosum*.

In contrast, VFs were detected in a small but distinct subset of vOTUs. Across the full catalogue of 116,740 vOTUs, 139 (0.1%) harboured at least one VF. Within the 8,970 filtered vOTU set, 71 vOTUs (prevalence 1.2–45.2%) were significantly associated with at least one bacterial taxon (FDR < 0.01), collectively associating with 259 of the 274 bacterial species tested. In contrast to AMG-encoding vOTUs, VF-encoding vOTUs showed more restricted host linkage, with a median of 67 VF-containing vOTUs associated per species. VF burden per vOTU was low, with a median of 1 unique VF per vOTU (IQR: 1–2). A subset of 38 bacterial species were associated with all 71 VF-encoding vOTUs at FDR < 0.01, with exclusively positive correlations observed for the most strongly associated species, including *Erysipelatoclostridium ramosum* (mean rho = 0.29), *Clostridium innocuum* (mean rho = 0.25), and *Eggerthella lenta* (mean rho = 0.24).

The most prevalent VFs with resolved gene-level annotation included the vacuolating autotransporter toxin (*vat*, $n = 6$ vOTUs) and LPS glucosyltransferases (*gtrA*, *gtrB*, *gtr;* $n = 2$ vOTUs each), the latter implicated in O-antigen modification and immune evasion. The majority of remaining VF hits matched genomic coordinates within the *Escherichia coli* O157:H7 EDL933 reference genome.

Collectively, the annotated VF repertoire suggests that a subset of viral genomic elements contains sequences annotated as bacterial virulence factors. Because these analyses were performed on bulk metagenomic assemblies, these annotations should be interpreted cautiously and require experimental validation to confirm phage localisation and function.

The correlation patterns between bacterial species and viral VF-encoding phages revealed a clear molecular signature distinguishing beneficial commensals from likely pathobionts (Fig. 6B). Among the 38 species consistently associated with the 71 VF-encoding phages (FDR < 0.01), 23 exhibited positive associations across all such phages. This cluster was enriched in species with opportunistic pathogenic potential, including established or emerging pathogens, such as *Clostridium innocuum*, *Eggerthella lenta*, *Ruminococcus gnavus*, *Enterocloster bolteae*, *Erysipelatoclostridium ramosum*, and members of the *Streptococcus anginosus* group (*S. anginosus*, *S. constellatus*), which are linked to bacteraemia, abscesses and other infections.

In contrast, the remaining 15 species displayed universally negative associations with VF-encoding phages. These were predominantly beneficial commensals and short-chain fatty acid producers with anti-inflammatory and barrier-protective roles, including Roseburia hominis, Lachnospira eligens, Coprococcus eutactus, Barnesiella intestinihominis, Alistipes spp., Bacteroides uniformis, and other Ruminococcaceae members.

This inverse relationship indicates that, at the population level, vOTUs containing putative virulence-associated genes were more frequently co-abundant with bacterial species that have been previously linked to opportunistic pathogenic characteristics.

Discussion

In this large and comprehensive human cohort study integrating phage, bacterial, metabolite, and dietary data, we provide an integrated characterisation of the human gut phageome and its population-level organisation within the human gut microbiome. Although individual phages were typically present at low relative abundances (< 0.3%), they showed structured and reproducible co-abundance networks with bacterial species and microbial metabolites. We further observed that the gut phageome composition in this cohort is moderately heritable but mainly explained by environmental factors. Importantly, negative correlations with VF-carrying vOTUs predominantly involved bacterial species previously characterised as beneficial SCFA-producing commensals, whereas positive correlations were observed with species previously associated with opportunistic pathogenic potential. Our findings provide novel insights into the population-level organisation of the gut phageome and its statistical alignment with functional potential within the gut microbiome.

Because vOTUs were reconstructed from bulk metagenomic data rather than virus-like particle–enriched preparations, the phageome characterised here likely reflects a composite signal of free phages and host-integrated prophages, capturing phage-associated genomic variation at the bacterial strain level.

Correlation-based network analyses revealed extensive co-variation between vOTUs and bacterial species, with several bacterial species, including *F. prausnitzii*, and *B. wexlerae*, emerging as highly connected nodes in both positive and negative association networks. These taxa are well-recognised as contributors to gut metabolic function, including SCFA production and cross-feeding processes[22,23]. The presence of these species as central hubs in viral-bacterial co-abundance networks is consistent with their high connectivity and strain-level diversity within this cohort. The observation that these taxa are present in both positive and negative networks may reflect heterogeneity or may arise from the coexistence of multiple phages with different genomic relationships to their bacterial hosts. However, distinguishing direct biological relationships from indirect or confounding associations remains challenging, and cannot be resolved using cross-sectional correlation-based analysis alone[4,24]. Thus, these observed associations should be interpreted as strictly descriptive rather than mechanistic.

Consistent with this view, we observed widespread associations between vOTUs and both stool- and serum-derived microbial metabolites, including SCFAs and secondary BAs, such as ursocholate and isoursodeoxycholate. These metabolites influence gut barrier integrity, immune tone, and systemic insulin sensitivity[25,26], consistent with phage-bacteria co-abundance patterns being associated with variation in microbial metabolic profiles across the cohort. However, these associations do not establish that phages directly influence metabolite production and may partly reflect shared bacterial composition or other features of the underlying metagenomic data.

Functional annotation identified a small subset of vOTUs encoding AMG or virulence-associated functions. Consistent with the

https://doi.org/10.1038/s41467-026-79230-0

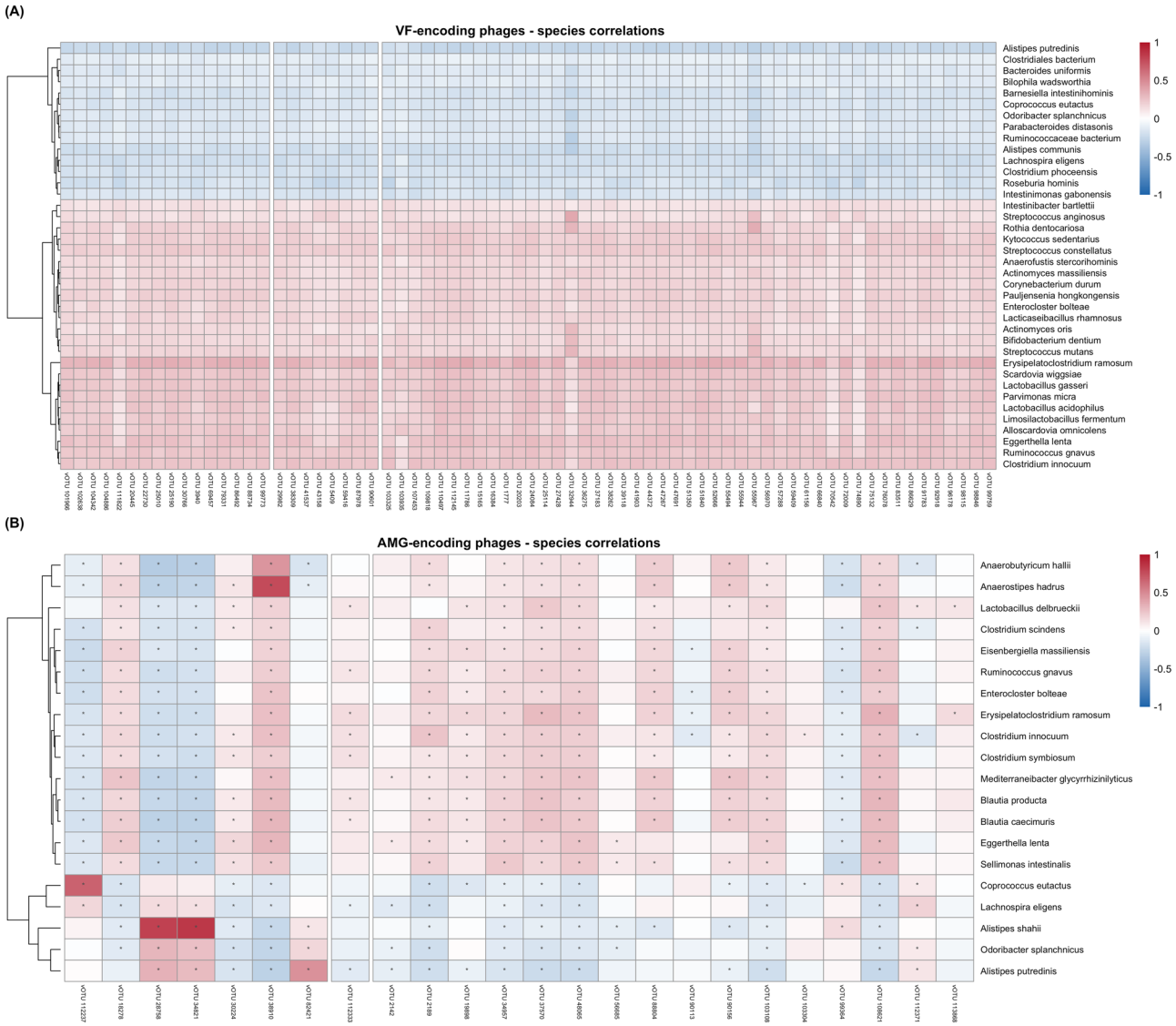


Fig. 6 | Associations between virulence factor- and auxiliary metabolic gene-encoding vOTUs and bacterial species. A Heatmap showing Spearman correlation coefficients between 71 virulence factor-encoding vOTUs and 38 bacterial species significantly associated with all VF-encoding vOTUs (FDR < 0.01). Colours indicate the strength and direction of correlations, with blue representing negative and red representing positive associations. vOTUs (columns) are ordered and grouped from left to right according to their estimated heritability, progressing from highly heritable vOTUs ($H^2 > 0.2$, $n = 16$), through moderately heritable vOTUs ($H^2 = 0.1–0.2$, $n = 8$), to non-heritable vOTUs ($H^2 < 0.1$ not represented; Non-heritable: $n = 47$). **B** Heatmap of Spearman correlations between 24 auxiliary metabolic gene-encoding vOTUs and the 20 bacterial species most frequently associated across vOTUs. Asterisks denote statistically significant correlations after multiple-testing correction (FDR < 0.01). Colours indicate the strength and direction of correlations, with blue representing negative and red representing positive associations. vOTUs (columns) are ordered and grouped from left to right according to their estimated heritability, progressing from highly heritable vOTUs ($H^2 > 0.2$, $n = 0$), through moderately heritable vOTUs ($H^2 = 0.1–0.2$, $n = 7$), to low heritable vOTUs ($H^2 < 0.1$, $n = 1$), and non-heritable contigs ($n = 16$). Abbreviations: HEI healthy eating index, vOTU operational taxonomic unit.

structured phage-bacteria networks, functional annotation identified a small but distinct subset of vOTUs encoding VFs. A core set of 38 bacterial species was consistently associated with all 71 VF-encoding vOTUs (FDR < 0.01). Among these, 23 species showed positive associations across all VF-carrying phages, forming a cluster enriched in opportunistic pathogens and pathobionts. This included *C. innocuum*, *E. lenta*, *R. gnavus*, *E. bolteae*, *E. ramosum*, and *S. anginosus* group members linked to bacteraemia, abscesses and invasive infections, as well as oral and cariogenic taxa (*Streptococcus mutans*, *Scardovia wiggsiae*, *Bifidobacterium dentium*, *Parvimonas micra*, *Actinomyces* spp., *Rothia dentocariosa*).

In contrast, the remaining 15 species displayed universally negative associations. These were predominantly beneficial commensals and short-chain fatty acid producers with anti-inflammatory and barrier-protective roles, including *R. hominis*, *L. eligens*, *C. eutactus*, *B. intestinihominis*, *A.* spp., *B. uniformis*, and other *R.* members.

These findings indicate that VF-encoding vOTUs were more commonly co-abundant with bacterial species previously associated with opportunistic pathogenic potential than alongside health-associated commensals. Whether this reflects horizontal gene transfer, shared ecological niches, prophage carriage, or other aspects of bacterial genome structure cannot be determined.

Extending these observations, a subset of vOTUs contained putative AMG annotations involved in essential bacterial functions, including nucleotide synthesis, central carbon metabolism, and polysaccharide degradation.

Heatmap-based visualisation of phage-bacterial-metabolite correlations further suggests that these relationships align with distinct

microbial metabolic profiles, with associations varying according to the underlying bacterial network context. These patterns illustrate population-level co-abundance relationships between vOTUs, bacterial taxa, and microbial metabolites. Although they are consistent with variation in bacterial functional capacity, they do not distinguish direct biological interactions from indirect associations arising through shared ecological structure or metagenomic dependence. For example, members of the *Clostridium* and *Eubacterium* genera possess enzymes required to deconjugate and convert primary BAs into secondary derivatives[27]. However, the relative importance of phages in driving microbial community composition remains inconclusive and an area of active debate from both ecological and evolutionary perspectives[28]. Given that several phages associated with these bacterial hubs also co-varied with specific BAs and SCFAs, it is plausible that phage–associated genomic variation co-varies with bacterial populations contributing to variation in host-associated metabolite profiles.

Previous research has shown that the human gut virome is established early in life and shaped by both genetics and environmental exposures, including mode of delivery (vaginal vs Caesarean) and breastfeeding status[29]. In this study, we provide the first quantitative estimate of phageome heritability. We show that while host genetics contributes to specific features of the phageome, such as particular axes of beta diversity and individual vOTUs, the overall composition is largely environmentally determined. This observation aligns with previous reports on bacteriome heritability, which demonstrated modest overall genetic effects alongside heritable signals in specific taxa[17].

We also identified four significant associations between phages and dietary components, namely, fibre and galactose intakes. Studies have shown that nutrient availability can be associated with changes in phage adsorption and intracellular replication dynamics, as phage receptors are often linked to bacterial nutrient transport systems and metabolic state[8]. These observations suggest that diet may contribute to population-level variation in phage-bacteria co-abundance patterns, although mechanistic inferences cannot be drawn.

Galactose intake has been associated with modulation of gut microbial communities and has been proposed to influence phage dynamics in experimental settings[30]. As a selective substrate, galactose may favour specific phage-bacteria configurations, potentially associated with differences in microbial community structure[31]. While our findings are consistent with such models, they should be interpreted as hypothesis-generating, and further experimental work will be required to determine whether galactose directly influences phage activity or bacterial susceptibility in vivo.

A key strength of our study lies in the integration of high-resolution multi-omic data from a large, well-characterised population-based cohort, combined with a comprehensive analytical framework.

However, several limitations should be considered. First, the cross-sectional design limits the ability to draw causal conclusions or assess temporal dynamics. Second, viral taxonomic resolution remains limited by incomplete reference databases, constraining precise functional attribution. Third, differences in taxonomic resolution across analytical approaches may affect comparability, as phage–taxa associations were assessed at the species level using shotgun metagenomic data, whereas adjustment for predicted bacterial hosts relied on iPHoP assignments restricted to the genus level. Fourth, although stringent viral identification and quality-control procedures were applied, some virulence factor annotations may reflect integrated prophages, neighbouring bacterial DNA, or other mobile genetic elements that cannot be completely distinguished using bulk metagenomic assemblies. Fifth, sample-level vOTU detection was based on empirically selected genome-coverage and mean-depth thresholds designed to limit low-support detections while preserving sensitivity for non-spurious, low-abundance viral populations. This threshold favours sensitivity over specificity, and absolute estimates of vOTU richness, prevalence, and diversity should accordingly be interpreted as indicative rather than exact and may be sensitive to detection threshold choice. Some low-support viral signals may nevertheless remain relative to more stringent mapping criteria; accordingly, our findings should be interpreted as identifying putative virulence-associated viral genomic elements rather than definitive phage-encoded virulence genes. Sixth, TwinsUK is predominantly composed of older females of European ancestry, limiting generalisability. Moreover, dietary intake was assessed via FFQs, which are subject to recall and reporting bias[32]. Seventh, our analyses are based on deep shotgun metagenomic sequencing rather than dedicated virome sequencing (e.g., VLP-enrichment), resulting in enrichment for host-associated prophages. This limits the ability to distinguish integrated from extracellular phages; integrated prophages are prevalent components of bacterial genomes and therefore represent an important component of phage-associated genomic variation captured by bulk metagenomic sequencing. Nevertheless, lifestyle prediction suggested that the recovered catalogue comprised both predicted temperate and virulent viral elements. Finally, our analyses focus on population-level statistical associations rather than direct ecological interactions and do not capture within-individual or within-twin-pair dynamics. Nevertheless, population-level co-abundance analyses are valuable for identifying reproducible patterns of phage-associated genomic variation across a large cohort and for generating hypotheses for future mechanistic studies[33]. Future longitudinal and experimental studies incorporating dedicated virome enrichment will be essential to disentangle causal relationships and validate the ecological and functional roles of gut phages suggested here.

In summary, this study provides a population-scale characterisation of the human gut phageome and its organisation within the gut microbiome. By integrating viral genome reconstruction with bacterial networks, metabolomics, and host factors, we describe structured patterns of phage–bacteria–metabolite co-abundance associations across a large human cohort. These findings provide a basis for investigating how phage-associated genomic variation relates to microbial community structure and metabolic function in the human gut and motivate future longitudinal, mechanistic and experimental studies.

Methods

Study population: participants were drawn from the TwinsUK Registry, a national cohort of adult twins recruited from the general population without selection based on disease status or traits[34]. Volunteers completed health questionnaires and attended clinical visits, during which biological samples were collected[34].

We included data from 1,069 twins with concurrent shotgun gut metagenomic data, and metabolomic profiling (serum and stool) measured within a one-year period. No statistical approach was used to predetermine sample size.

A subset of 805 individuals also completed a 131-item modified food frequency questionnaire from the European Prospective Investigation into Cancer and Nutrition[35], within five years of faecal sample collection. Nutrient levels were derived using FETA[36]. Nutrients were adjusted for energy intake using the residual method, as previously described[37]. The HEI (2010) was constructed to capture overall diet quality[38]. Underweight individuals (BMI < 18.5 kg/m^2) were excluded.

Shotgun metagenomic sequencing

Deep shotgun metagenomic sequencing in stool samples from TwinsUK was performed by GenomeScan[39]. This enabled simultaneous profiling of both bacterial and viral communities as detailed below.

Faecal Sample collection and processing

Stool samples were collected at home using pre-labelled kits (2 × 25 ml tubes or 1 × 25 ml tube and 1 × 10 ml Zymo buffer). Samples were either brought to the clinic or posted using Royal Mail safe boxes. In the laboratory, samples were homogenised, aliquoted into four bijou tubes, and stored at −80 °C within two hours of receipt to minimise degradation.

DNA extraction, library preparation and DNA sequencing. DNA was extracted from 0.5 g of stool using a combination of mechanical and enzymatic lysis. Homogenised samples in a lysis buffer were treated with Proteinase K and subjected to bead-beating with garnet beads, followed by incubation at 65 °C and 95 °C. DNA was extracted using a Chemagic MSM I (Perkin Elmer) system, followed by inhibitor removal using the Zymo OneStep Inhibitor Removal Kit. DNA was quantified using Quant-iT on an Eppendorf AF2200 plate reader.

Libraries were prepared manually using the Nextera XT protocol (Illumina), with DNA normalised to 0.2 ng/µl. Libraries were tagged, PCR-amplified, cleaned with AMPure XP beads, and fragment sizes assessed using a LabChip GX Touch. Sequencing was performed on an Illumina HiSeq 2500 using SBS Kit v4 (2 × 125 bp), generating a mean of 54 million paired-end reads per sample.

Metagenome quality control, preprocessing, and taxonomic profiling. Raw reads were processed using the YAMP pipeline (v. 0.9.5.3)[40]. Identical reads were removed, and reads were filtered for adaptors, known artefacts, and phiX174 sequences. Low-quality bases (Phred < 10) were trimmed and reads < 60 bp discarded. Singleton reads were retained, and human reads were filtered using the GRCh37 human genome reference. Samples with <10 million reads after QC were discarded ($n = 4$).

Bacterial taxonomic profiling. Bacterial taxonomic profiling was conducted using MetaPhlAn4 (v4.0.0)[41] as integrated within the YAMP pipeline[40].

Alpha diversity was estimated using the vegan package (version 2.6-10)[42] and quantified using (i) the Shannon diversity index (Shannon entropy)[43] to assess community evenness and richness, and (ii) observed taxonomic richness as the count of unique species. Bacterial species were modelled as relative abundances.

Viral community profiling. The viral community profiling pipeline is illustrated in the schematic (Supplementary Fig 5) and at github.com/PanayiotisLouca/twinsuk-viromics-pipeline[44].

To profile the gut phageome, we performed de novo assembly of quality-controlled and host-depleted metagenomic reads from all samples with MEGAHIT[45]. Viral contigs were identified using geNomad[46], VIBRANT[47] and VirSorter2[48]. Viral contigs identified by the three viral discovery tools were combined and dereplicated using MMseqs2 (95% sequence identity and 85% coverage thresholds)[16] to generate a non-redundant set of vOTUs[49]. The non-redundant vOTUs were subsequently processed with CheckV[50] to assess completeness and identify host-derived regions. For vOTUs identified as integrated proviruses, flanking host regions were removed, and the resulting viral sequences were retained for the curated viral reference library. Only vOTUs >5 kb with a clear viral signature and classified as complete, high-quality, or medium-quality were retained for downstream analyses. Consequently, subsequent read mapping and abundance estimation were performed against the curated viral sequences without the flanking host regions.

Abundance tables were generated by mapping quality-controlled reads to the curated vOTU library with Bowtie2[51], which showed higher sensitivity compared with Minimap2[52] in benchmarking tests using our dataset. To minimise poorly supported viral signals, vOTU detection thresholds were empirically determined for the study cohort by evaluating the number of detected vOTUs across a range of genome-coverage and mean mapping depth cutoffs (Supplementary Fig. 6). The number of detected vOTUs per sample progressively stabilised ≥20% genome coverage and ≥0.5 × sequencing depth; these thresholds were therefore selected to balance sensitivity against the risk of spurious low-support detections. Viral abundances were estimated by normalising read counts to vOTU genome length, scaling to one million per sample, and converting to relative abundances for community-level analyses.

The replication cycle of vOTUs was determined using BACPHLIP[53]. VFs were predicted using ABRicate (https://github.com/tseemann/abricate). Hits were considered valid if they met minimum sequence identity (≥ 90%) and alignment coverage (≥ 90%) thresholds. The taxonomy of the vOTUs was predicted using geNomad, and Host prediction was performed using iPHoP[16]. Only predicted genera with a confidence score ≥ 90% were retained. For functional annotation, we used DRAM-v[54] to identify AMGs, applying stringent viral-context filters, including exclusion of genes adjacent to viral hallmark genes and validation against multiple domain databases (KEGG, Pfam, VOGDB). Only AMGs with an auxiliary score of 1 were retained.

Short chain fatty acid and secondary bile acid profiling. SCFAs and secondary bile acids were measured in stool and fasting serum samples by Metabolon Inc. (Durham, NC, USA). Samples were stored at −80 °C until processing.

Acetate, propionate, and butyrate were quantified from a subset of 953 participants following methanol-based protein precipitation[21]. Internal standards were used to monitor extraction efficiency and analytical performance. The resulting extracts were analysed by LC-MS/MS, and metabolites were identified and quantified based on their chromatographic retention and mass spectrometric signals.

For bile acid profiling, samples from all 1,069 included participants underwent methanol precipitation and LC-MS analysis under complementary chromatographic and ionisation conditions[20]. Metabolites were identified using an in-house reference library based on retention time/index, m/z, and MS/MS spectra, with internal standards and quality-control samples used to monitor analytical performance. Peak areas were normalised for inter-day and batch variation, and missing values were imputed using the minimum observed value after median scaling.

Following quality control and data processing, 11 serum and 12 stool secondary bile acids were retained for analysis. These metabolites were detected in at least 80% of samples.

Statistics and reproducibility. All statistical analyses were performed using R (version 4.4.1)[55] with resources from King's College London Computational Research, Engineering, and Technology Environment (CREATE)[56]. *P*-values were corrected for multiple testing using the Benjamini and Hochberg method. Dietary data and all continuous covariates were z-scored.

Taxonomic Data QC and normalisation approaches. To reduce sparsity and ensure robust statistical comparisons, vOTUs present in at least 1% and bacterial species present in at least 10% of samples were retained. A total of 41,431 vOTUs and 274 bacterial species passed prevalence filtering. To further refine the dataset, a minimum mean relative abundance threshold of > 0.001% was applied, resulting in a final set of 8,970 vOTUs and 274 bacterial species for downstream analysis.

Gut microbial features were centred log-ratio (CLR) transformed to account for compositional effects, with a pseudocount of 1×10^{-6} added to handle zero values prior to transformation, followed by inverse-rank normalisation. Alpha diversity metrics, metabolite levels, and dietary data were also inverse-rank normalised. Continuous covariate data (e.g., age) were z-scored.

Heritability analysis. Heritability of the phageome was estimated for both alpha diversity (measured by the Shannon diversity index and observed taxonomic richness) and beta diversity (first and second principal components of Bray-Curtis dissimilarities). We included 261 monozygotic and 148 dizygotic twin pairs. Structural equation modelling was implemented with the mets package (version 1.3.5)[57] to decompose phenotypic variance into additive genetic (A), shared environmental (C), and unique environmental (E) components. ACE, AE, CE and E models were fitted using the twinlm function, adjusting for age, sex and BMI as covariates. Narrow-sense heritability (h^2) was defined as the proportion of phenotypic variation attributable to additive genetic factors (A), calculated as $h^2 = \frac{(A)}{(A+C+E)}$ for ACE models. Model fit was assessed using the AIC, with the lowest AIC indicating the best-fitting model, and *p*-values were derived from likelihood ratio tests.

Phage-bacteria correlation network analysis. To identify potential phage-bacteria co-abundance associations, Spearman's correlation coefficients were calculated between vOTUs and bacterial species. Correlations exceeding predefined thresholds were visualised in a bipartite network using the igraph[58] and ggraph[59] packages with a Fruchterman-Reingold layout. Edges were weighted by absolute correlation values. To account for potential confounding by bacterial hosts, we repeated the linear mixed models for phage ~ species analysis, further adjusting for predicted host genera (inferred using iPHoP as described above). Results were adjusted for multiple testing using FDR < 0.01.

vOTU associations with metabolites, and dietary intake. We investigated associations between vOTUs and microbial metabolites, specifically SCFAs and secondary BAs, and dietary intakes. Analyses were performed using linear mixed-effects models implemented in glmmTMB[60]. Models were adjusted for age, sex, and BMI (fixed effects), while family relatedness and batch effects were modelled as random intercepts.

Ethics

TwinsUK was carried out under TwinsUK BioBank ethics, approved by North West—Liverpool Central Research Ethics Committee (REC reference 19/NW/0187), IRAS ID 258513. This approval supersedes earlier approvals granted to TwinsUK by the St Thomas' Hospital Research Ethics Committee, later London—Westminster Research Ethics Committee (REC reference EC04/015, approved on 24 May 2019), which have now been subsumed within the TwinsUK BioBank. All twins provided written informed consent.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the findings described in this manuscript have been deposited to the TwinsUK BioResource data management team and are available by application to the Twin Research Executive Access committee (TREC) at King's College London. Researchers can apply for access to individual level data via the TwinsUK Data Access portal at https://twinsuk.ac.uk/researchers/access-data-and-samples/request-access/. TwinsUK data users are bound by data sharing agreement set out in the data access application form, which includes responsibilities with respect to third party data sharing and maintaining participant privacy. Further responsibilities include a responsibility to acknowledge data sharing. Source Data are provided with this paper for all Figures with the exception of Fig. 1, where no source data was used. The raw metagenomic sequence data used in this study have been deposited under the NCBI BioProject (Accession: PRJNA1357304).

Code availability

Code relating to the vOTU workflow is available at github.com/PanayiotisLouca/twinsuk-viromics-pipeline[44]. Statistical analysis code is publicly available at github: github.com/PanayiotisLouca/twinsuk-phageome-analysis[61].

References

1. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 https://doi.org/10.1038/s41579-020-0433-9 (2021).
2. Wei, Y. Bacteriophages: a double-edged sword in the gastrointestinal tract. *Front. Microbiomes* **3** https://doi.org/10.3389/frmbi.2024.1450523 (2024).
3. Kirk, D. et al. Bacteriophages, gut bacteria, and microbial pathways interplay in cardiometabolic health. *Cell Rep.* **43** https://doi.org/10.1016/j.celrep.2024.113728 (2024).
4. Mirzaei, M. K. & Maurice, C. F. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* **15**, 397–408 https://doi.org/10.1038/nrmicro.2017.30 (2017).
5. Borodovich, T. et al. Large-scale capsid-mediated mobilisation of bacterial genomic DNA in the gut microbiome. *Nat. Commun.* **17**, 2046 https://doi.org/10.1038/s41467-026-68726-4 (2026).
6. de Jonge, P. A. et al. Gut virome profiling identifies a widespread bacteriophage family associated with metabolic syndrome. *Nat. Commun.* **13**, 3594 https://doi.org/10.1038/s41467-022-31390-5 (2022).
7. Zhang, Y. & Wang, R. The human gut phageome: composition, development, and alterations in disease. *Front. Microbiol.* **ume 14**, 2023 https://doi.org/10.3389/fmicb.2023.1213625 (2023).
8. Liang, X. et al. Bacteriophage-driven microbial phenotypic heterogeneity: ecological and biogeochemical importance. *npj Biofilms Microbiomes* **11**, 82 https://doi.org/10.1038/s41522-025-00727-5 (2025).
9. Shkoporov, A. N. & Hill, C. Bacteriophages of the human gut: the "known unknown" of the microbiome. *Cell Host Microbe* **25**, 195–209 https://doi.org/10.1016/j.chom.2019.01.017 (2019).
10. Yu, X. et al. Gut phageome: challenges in research and impact on human microbiota. *Front. Microbiol.* **ume 15**, 2024 https://doi.org/10.3389/fmicb.2024.1379382 (2024).
11. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109.e1099 https://doi.org/10.1016/j.cell.2021.01.029 (2021).
12. Camargo, A. P. et al. A genomic atlas of the human gut virome elucidates genetic factors shaping host interactions. *bioRxiv*, 2025.2011.2001.686033 https://doi.org/10.1101/2025.11.01.686033(2025).
13. Galperina, A. et al. The aggregated gut viral catalogue (AVrC): a unified resource for exploring the viral diversity of the human gut. *PLoS Comput. Biol.* **21**, e1012268 https://doi.org/10.1371/journal.pcbi.1012268 (2025).
14. Ru, J., Khan Mirzaei, M., Xue, J., Peng, X. & Deng, L. ViroProfiler: a containerized bioinformatics pipeline for viral metagenomic data analysis. *Gut Microbes* **15**, 2192522 https://doi.org/10.1080/19490976.2023.2192522 (2023).
15. Louca, P. & Menni, C. TwinsUK non-redundant viral reference catalogue. *Zenodo* https://doi.org/10.5281/zenodo.21111469 (2026).
16. Roux, S. et al. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* **21**, e3002083 https://doi.org/10.1371/journal.pbio.3002083 (2023).
17. Goodrich, J. K. et al. Human genetics shape the gut microbiome. *Cell* **159**, 789–799 https://doi.org/10.1016/j.cell.2014.09.053 (2014).
18. Xu, F. et al. The interplay between host genetics and the gut microbiome reveals common and distinct microbiome features for

complex human diseases. *Microbiome* **8**, 145 https://doi.org/10.1186/s40168-020-00923-9 (2020).
19. Moreno-Gallego, J. L. et al. Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host Microbe* **25**, 261–272.e265 https://doi.org/10.1016/j.chom.2019.01.019 (2019).
20. Louca, P. et al. The secondary bile acid isoursodeoxycholate correlates with post-prandial lipemia, inflammation, and appetite and changes post-bariatric surgery. *Cell Rep. Med* **4**, 100993 https://doi.org/10.1016/j.xcrm.2023.100993 (2023).
21. Nogal, A. et al. Genetic and gut microbiome determinants of SCFA circulating and fecal levels, postprandial responses and links to chronic and acute inflammation. *Gut Microbes* **15**, 2240050 https://doi.org/10.1080/19490976.2023.2240050 (2023).
22. Hosomi, K. et al. Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nat. Commun.* **13**, 4477 https://doi.org/10.1038/s41467-022-32015-7 (2022).
23. Culp, E. J. & Goodman, A. L. Cross-feeding in the gut microbiome: ecology and mechanisms. *Cell Host Microbe* **31**, 485–499 https://doi.org/10.1016/j.chom.2023.03.016 (2023).
24. Shkoporov, A. N., Turkington, C. J. & Hill, C. Mutualistic interplay between bacteriophages and bacteria in the human gut. *Nat. Rev. Microbiol.* **20**, 737–749 https://doi.org/10.1038/s41579-022-00755-4 (2022).
25. Chen, M. L., Takeda, K. & Sundrud, M. S. Emerging roles of bile acids in mucosal immunity and inflammation. *Mucosal Immunol.* **12**, 851–861 https://doi.org/10.1038/s41385-019-0162-4 (2019).
26. Luo, S. et al. Helicobacter pylori infection alters gut virome by expanding temperate phages linked to increased risk of colorectal cancer. *Gut* **73**, 1592–1595 https://doi.org/10.1136/gutjnl-2023-330362 (2024).
27. Winston, J. A. & Theriot, C. M. Impact of microbial derived secondary bile acids on colonization resistance against Clostridium difficile in the gastrointestinal tract. *Anaerobe* **41**, 44–50 https://doi.org/10.1016/j.anaerobe.2016.05.003 (2016).
28. Castledine, M. & Buckling, A. Critically evaluating the relative importance of phage in shaping microbial community composition. *Trends Microbiol.* **32**, 957–969 https://doi.org/10.1016/j.tim.2024.02.014 (2024).
29. Lathakumari, R. H. et al. The gut virome and human health: From diversity to personalized medicine. *Eng. Microbiol.* **5**, 100191. https://doi.org/10.1016/j.engmic.2025.100191 (2025).
30. Glowacki Robert, W. P. & Martens Eric, C. If you eat it or secrete it, they will grow: the expanding list of nutrients utilized by human gut bacteria. *J. Bacteriol.* **203**, 10.1128/jb.00481-00420 https://doi.org/10.1128/jb.00481-20 (2021).
31. Azcarate-Peril, M. A. et al. Impact of short-chain galactooligosaccharides on the gut microbiome of lactose-intolerant individuals. *Proc. Natl. Acad. Sci. USA* **114**, E367–E375 https://doi.org/10.1073/pnas.1606722113 (2017).
32. Miller, T. M., Abdel-Maksoud, M. F., Crane, L. A., Marcus, A. C. & Byers, T. E. Effects of social approval bias on self-reported fruit and vegetable consumption: a randomized controlled trial. *Nutr. J.* **7**, 18 https://doi.org/10.1186/1475-2891-7-18 (2008).
33. Wirbel, J. et al. Long-read metagenomics reveals phage dynamics in the human gut microbiome. *Nature* https://doi.org/10.1038/s41586-025-09786-2 (2025).
34. Verdi, S. et al. TwinsUK: the UK adult twin registry update. *Twin Res Hum. Genet* **22**, 523–529 https://doi.org/10.1017/thg.2019.65 (2019).
35. Bingham, S. A. et al. Nutritional methods in the European prospective investigation of cancer in Norfolk. *Public Health Nutr.* **4**, 847–858 https://doi.org/10.1079/PHN2000102 (2001).
36. Mulligan, A. A. et al. A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open* **4**, e004503 https://doi.org/10.1136/bmjopen-2013-004503 (2014).
37. Willett, W. & Stampfer, M. J. Total energy intake: implications for epidemiologic analyses. *Am. J. Epidemiol.* **124**, 17–27 https://doi.org/10.1093/oxfordjournals.aje.a114366 (1986).
38. Bowyer, R. C. E. et al. Use of dietary indices to control for diet in human gut microbiota studies. *Microbiome* **6**, 77 https://doi.org/10.1186/s40168-018-0455-y (2018).
39. Pope, R. et al. Faecal metabolites as a readout of habitual diet capture dietary interactions with the gut microbiome. *Nat. Commun.* **16**, 10051 https://doi.org/10.1038/s41467-025-66046-7 (2025).
40. Visconti, A., Martin, T. C. & Falchi, M. YAMP: a containerized workflow enabling reproducibility in metagenomics research. *GigaScience* **7** https://doi.org/10.1093/gigascience/giy072 (2018).
41. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 https://doi.org/10.1038/s41587-023-01688-w (2023).
42. vegan: Community Ecology Package v. 2.6-7 (2024).
43. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication*. (The University of Illinois Press, 1949).
44. Louca, P. *TwinsUK Viromics Profiling Pipeline (V1.0.0)*, https://github.com/PanayiotisLouca/twinsuk-viromics-pipeline (2026).
45. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 https://doi.org/10.1093/bioinformatics/btv033 (2015).
46. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 https://doi.org/10.1038/s41587-023-01953-y (2024).
47. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 https://doi.org/10.1186/s40168-020-00867-0 (2020).
48. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 https://doi.org/10.1186/s40168-020-00990-y (2021).
49. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 https://doi.org/10.1038/nbt.3988 (2017).
50. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 https://doi.org/10.1038/s41587-020-00774-7 (2021).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 https://doi.org/10.1038/nmeth.1923 (2012).
52. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 https://doi.org/10.1093/bioinformatics/bty191 (2018).
53. Hockenberry, A. J. & Wilke, C. O. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ.* **9**, e11396 https://doi.org/10.7717/peerj.11396 (2021).
54. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 https://doi.org/10.1093/nar/gkaa621 (2020).
55. R: A language and environment for statistical computing. v. 4.3.1 (R Foundation for Statistical Computing, 2020).
56. King's College London. *King's Computational Research, Engineering and Technology Environment (CREATE)* (King's College London, 2025).
57. Holst, K. K., Scheike, T. H. & Hjelmborg, J. B. The liability threshold model for censored twin data. *Comput. Stat. Data Anal.* **93**, 324–335 https://doi.org/10.1016/j.csda.2015.01.014 (2016).

58. igraph for R: R interface of the igraph library for graph theory and network analysis v. (v2.1.4) (2025).
59. ggraph: an implementation of grammar of graphics for graphs and networks. v. 2.2.1.9000 (2025).
60. McGillycuddy, M., Popovic, G., Bolker, B. M. & Warton, D. I. Parsimoniously fitting large multivariate random effects in glmmTMB. *J. Stat. Softw.* **112**, 19, https://doi.org/10.18637/jss.v112.i01 (2025).
61. Louca, P. *TwinsUK Phageome Analysis*, https://github.com/PanayiotisLouca/twinsuk-phageome-analysis (Github, 2026).

Acknowledgements

We express our sincere thanks to all participants and staff of the TwinsUK cohort.

Author contributions

Conceived and designed the experiments: M.K.M., L.D., A.M.V. and C.M. Data curation: A.M.V., M.K.M., P.L. and C.M. Analysed the data: P.L. and C.M. Contributed reagents/materials/analysis tools: E.K., A.K., Y.L., X.P., R.P., A.V., F.A., D.K., R.C., N.S., M.F., J.T.B. and T.D.S. Wrote the manuscript: P.L., C.M., M.K.M., A.M.V., A.K. and L.A.E. Revised the manuscript: all.

Funding

This research was funded in whole, or in part, by the UKRI (MR/Y010175/1, MR/T004142/1) and by the Chronic Disease Research Foundation. This work was also supported by the Wellcome Trust (WT212904/Z/18/Z). For the purpose of open access, the authors have applied a CC-BY public copyright to any Author Accepted Manuscript version arising from this submission. TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Wellcome Leap Dynamic Resilience Programme (co-funded by Temasek Trust), Zoe Ltd, the National Institute for Health and Care Research (NIHR) Clinical Research Network (CRN) and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. CM is funded by the Chronic Disease Research Foundation (CDRF), by the Italian Ministry of Education and Research: Dipartimenti di Eccellenza Programme 2023 to 2027 and by the Italian Ministry of Health – Bando Ricerca Corrente. AMV is supported by the National Institute for Health and Care Research Nottingham Biomedical Research Centre. LD is funded by the Helmholtz Association (INF W2/W3-161), German Research Foundation (D.F.G. Emmy Noether programme, Project No. 273124240; SPP 2330, Project No. 464797012), Marie Sklodowska-Curie Actions Innovation Training Networks grant agreement no. 955974(VIROINF) and the European Research Council Starting grant (ERC StG 803077).

Competing interests

TDS is a co-found of ZOE Ltd. AMV, and NS are consultants to ZOE Ltd. TDS and NS have received options from ZOE Ltd. All other authors declare no competing financial interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-79230-0.

Correspondence and requests for materials should be addressed to Cristina Menni.

Peer review information *Nature Communications* thanks other anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

Panayiotis Louca[1,12], Mohammadali Khan Mirzaei[2,3,4,12], Afroditi Kouraki[5,6], Erfan Khamespanah[2,3], Yu Lin[1], Xue Peng[2,3], Robert Pope[1], Alessia Visconti[1,7], Francesco Asnicar[8], Daniel Kirk[1], Ricardo Costeira[1], Nicola Segata[8], Mario Falchi[1], Jordana T. Bell[1], Tim D. Spector[1], Lindsey A. Edwards[9], Li Deng[2,3,4,12], Ana M. Valdes[5,6,12] & Cristina Menni[1,10,11,12] ✉

[1]Department of Twin Research, Ageing & Genomic Epidemiology, King's College London, London, UK. [2]Institute of Virology, Helmholtz Centre Munich - German Research Centre for Environmental Health, Neuherberg, Germany. [3]Technical University of Munich, TUM School of Life Sciences, Chair of Prevention of Microbial Diseases, Central Institute of Infection Prevention (ZIP), Freising, Germany. [4]Bavaria Competent Phage Therapy Centre, Freising, Germany. [5]NIHR Nottingham Biomedical Research Centre, Queens Medical Centre, Nottingham, UK. [6]Injury, Recovery and Inflammation Sciences, School of Medicine, University of Nottingham, Nottingham, UK. [7]Centre for Biostatistics, Epidemiology, and Public Health, Department of Clinical and Biological Sciences, University of Turin, Turin, Italy. [8]Department CIBIO, University of Trento, Trento, Italy. [9]Centre for Host Microbiome Interactions, Faculty of Dentistry, Oral & craniofacial sciences, Guy's Hospital, King's College London, London, UK. [10]Department of Pathophysiology and Transplantation, Università Degli Studi di Milano, Milan, Italy. [11]Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy. [12]These authors contributed equally: Panayiotis Louca, Mohammadali Khan Mirzaei, Li Deng, Ana M. Valdes, Cristina Menni.
✉e-mail: cristina.menni@kcl.ac.uk