Community types of the human gut virome are associated with endoscopic outcome in ulcerative colitis

Daan Jansen  
KU Leuven  
Gwen Falony  
VIB  
Sara Vieira-Silva  
VIB  
Ceren Simsek  
KU Leuven  
Tine Marcelis  
KU Leuven  
Clara Caenepeel  
KU Leuven, University Hospitals Leuven  
Kathleen Machiels  
KU Leuven, University Hospitals Leuven  
Jeroen Raes  
VIB  
Severine Vermeire  
KU Leuven, University Hospitals Leuven  
Jelle Matthijnssens (✉️ jelle.matthijnssens@kuleuven.be)  
KU Leuven

Research Article

**Keywords:** IBD, virome, microbiome, community-typing, bacteriophages, metagenomics

**Posted Date:** October 19th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-2166222/v1

**License:** ©️️ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background

Inflammatory Bowel Diseases (IBD) are a group of chronic inflammatory diseases of the gut. IBD patients have an altered gut microbiota; however, the relationship to disease is unknown. The gut microbiota is a complex ecosystem and bacterial community-typing is an established approach to condense the microbial complexity into enterotypes. A viral counterpart of enterotypes might allow stratification of individuals based on their gut virome. We aim to investigate the existence of such viral community types and assess the impact of therapeutic outcome (and other covariates) on the gut virome in IBD patients.

Methods

Viral particle enrichment followed by deep sequencing (1.52TB) was performed on 432 faecal samples from 181 IBD patients (CD = 126; UC = 55) starting biological therapy. Redundancy analysis and Dirichlet Multinomial Mixtures were applied to determine covariates of the virome composition and to condense the gut virota into viral community types, respectively.

Results

IBD patients were stratified based on unsupervised machine learning into two viral community types. Community type CA showed a low $\alpha$-diversity and a high relative abundance of Caudoviricetes [non-CrAss] phages and was associated to the dysbiotic Bact2-enterotype. Community type CrM showed a high $\alpha$-diversity and a high relative abundance of Caudoviricetes [CrAss] and Malgrandaviricetes phages. The gut virome variation was explained by several factors: patient individuality (75.8%), disease location (1.4%), age (0.5%) and faecal moisture (0.3%), whereas diagnosis did not show a non-redundant effect. Despite our expectations the choice of biological therapy did not show an association with the virome variation. During post-interventional analysis, endoscopic outcome (0.5%) was associated to gut virome variation. Remitting UC, but not CD, patients revealed a high percentage of community type CrM, a high Shannon diversity and a low lysogenic potential. Conversely, non-remitting UC, but not CD, patients revealed a high percentage of community type CA, a low Shannon diversity and a high lysogenic potential. During pre-interventional analysis, we discovered five novel phages associated with treatment success.

Conclusion

The gut virota shows the existence of distinct virome configurations that are associated with endoscopic outcome. Therefore, community typing could be a valuable tool to improve our understanding about IBD subtypes, pathology, and activity.

Background
The gut microbiota is a complex ecosystem that consists of viruses, fungi, bacteria, archaea, and protozoa. It can exert beneficial functions to the human host, such as protection against invading pathogens or production of essential vitamins. At times, this microbial ecosystem gets disrupted, resulting in gut dysbiosis. Gut dysbiosis is associated with several diseases, one of which is inflammatory bowel disease (IBD)\(^1\). IBD is a group of chronic remitting diseases involving inflammation of the gut, and its two main phenotypes include ulcerative colitis (UC) and Crohn’s disease (CD). Although the aetiology of IBD is unknown, the interplay between host genetic susceptibility, a mucosal immune response to the host microbiota and other environmental factors, has been suggested as working hypothesis\(^2\).

In recent years, the bacterial component of the microbiota has been repeatedly associated with the pathology and activity of IBD\(^3-5\). The gut microbiota of active IBD patients is characterized by a high abundance of *Proteobacteria* and a low abundance of *Firmicutes*, combined with a low bacterial α-diversity and cell count\(^2,6\). Another frequently reported alteration is the reduction in anti-inflammatory bacteria, particularly butyrate-producing bacteria (e.g. *Faecalibacterium prausnitzii*)\(^7\). Community analysis can provide a more holistic view of the gut microbiota by collapsing the microbial variation into just a few categories. Enterotyping (or bacterial community typing) is such an analysis that can stratify patients based on their gut microbiota\(^8\). Four enterotypes have been reported, named *Bacteroides*\(^1\) (Bact1), *Bacteroides*\(^2\) (Bact2), *Prevotella* (Prev) and *Ruminococcus* (Rum)\(^6\). One enterotype, Bact2, largely reflects the IBD-specific bacterial alterations, as described above\(^9\). Up to 80% of IBD patients harbor the Bact2-enterotype, which is seen as a dysbiotic enterotype, while less than 15% of healthy individuals possess this enterotype\(^6,10\).

Despite the vast number of associations made with bacteria, little is known about the role of the viral component in the pathology and activity of IBD. A growing body of evidence suggests that disease pathology is associated with alterations in the gut virota as well\(^11-14\). These alterations are largely characterized by a high abundance of members of the *Caudovirales* and a low abundance of *Microviridae* members\(^15,16\). Other alterations show a highly lysogenic potential of the gut virota\(^11\). In addition, diversity changes have also been reported, albeit in an inconsistent manner\(^11-14,17,18\). One of the interesting viral groups is the recently identified CrAss-like phages, which are a diverse group that are believed to be the most abundant viruses present in the human gut\(^19\). Recently, Gulyaeva and colleagues described that CrAss-like phages are depleted in the IBD faecal virota\(^20\). Other studies have been suggesting that some lytic phages (e.g. CrAss-like phage crAss001) could exist in symbiosis with their bacterial host, and even drive bacterial diversity through a process called phase variation\(^21\). More specifically, phase variation allows the parallel multiplication of phages and their host by providing a balance between phage resistance and sensitivity. To date, no consistent associations have been made between the gut virota and disease activity. However, we anticipate that unravelling the full complexity of the human gut virome will deepen our understanding of complex human disease. Consequently, a viral counterpart of enterotyping (‘viral community-typing’) might improve this understanding and would allow stratification of individuals based on their gut virota. A first attempt was performed by Song and colleagues using published sequencing data of 2,690 metagenomes, although was unable to describe the viral taxonomic composition of the community types due to a high number of unclassified viruses\(^22\).

In this study, we analyzed faecal samples (n = 432) of a prospective cohort of active IBD patients (n = 181) starting biological therapies and investigated the factors shaping the gut virome composition. Viral community
typing was performed to describe the virome configurations, and to associate with covariates of the virome composition. In doing so, we are able to identify therapy outcome (as measured by endoscopic outcome) as a covariate of the gut virome, thereby highlighting the role of the gut virome in IBD.

**Results**

**The gut virome is dominated by Caudoviricetes and Malgrandaviricetes phages in a multi-therapeutic IBD cohort**

Faecal samples were collected from patients in a prospective multi-therapeutic IBD cohort (n = 181)\(^23\). Patients had either active ulcerative colitis (UC, n = 55), or active Crohn's disease (CD, n = 126) and started biologicals as part of their medical care. Patients were re-evaluated at the pre-defined primary endpoint (post-intervention) and classified as achieving remission or not (Extended Data Fig. 1; Extended Data Methods; Supplementary Table 1). The gut virome was characterized using the NetoVIR protocol to isolate, enrich and sequence viruses in faecal samples (Extended Data Fig. 2a). Computational analyses on an input of 10.2 billion paired end reads (1.52 TB, \(\overline{x} = 23.6\) million reads per sample) was performed with the latest methodologies (Extended Data Fig. 2b). Most of the quality-controlled reads were found to be of viral origin (viral = 63.0%, bacterial = 32.3%, other = 3.0%, dark matter = 1.0%; Extended Data Fig. 3). Most of the identified viruses could reliably be classified at class-level taxonomy (classified = 93%, unclassified = 7%; Extended Data Fig. 4; Supplementary Table 2). The two most abundant viral classes were *Caudoviricetes* (58.1%, dsDNA tailed phages) and *Malgrandaviricetes* (37.7%, ssDNA circular phages), representing most quality-controlled reads (\(\overline{x} = 95.8\)%, range = 2.76%-100% per sample). The former viral class could be broken down into *Caudoviricetes* [non-CrAss] and *Caudoviricetes* [CrAss] phages, encompassing 41.9% and 16.2% of the quality-controlled reads, respectively (Supplementary Table 2).

**The eukaryotic virome is small and is largely composed of plant viruses**

A small percentage of samples contained eukaryotic viruses (32.9%) representing a minority of the quality-controlled viral reads (eukaryotic viruses = 0.2%, phages = 99.8%; Extended Data Fig. 5; Supplementary Table 2). Eukaryotic viruses could be grouped based on the host, which could be known (animal, plant or fungal viruses) or unknown (small circular viruses). Most of the viruses detected in IBD patients belonged to the plant and fungal viral group, with only a few samples containing small circular viruses, or viruses potentially causing gastroenteritis. The two most prevalent viral species were Pepper mild mottle virus (prevalence = 12.9%, genus = *Tobamovirus*) and Pepino mosaic virus (prevalence = 10.8%, genus = *Potexvirus*), likely obtained via the patient’s diet (Supplementary Table 3)\(^24\).

**The gut virota reveals the existence of two virome configurations in IBD patients**

The gut microbiota is complex and variable, further complicating its thorough exploration. One approach is to condense the bacterial complexity into bacterial community types ('enterotypes'). By using Dirichlet Multinomial Mixture (DMM) modelling, bacterial research has consistently stratified large human gut microbiota studies into
four enterotypes (Extended Data Methods)\textsuperscript{9,23,25}. We used the same methodology and applied this for the first time to the gut virota. The gut virota observed in the IBD cohort consisted of no less than 874 genus-like groups (median = 26, range = 3–74 per individual). Applying the DMM algorithm reduced the viral complexity and revealed the existence of two distinct clusters, or virome configurations (n = 363, genus-like group, Bray-Curtis dissimilarity; Fig. 1a; Supplementary Table 4). The groups showed a high probability of cluster assignment and were hereafter referred to as viral community types (median = 99.6%, Extended Data Fig. 6; Supplementary Table 4).

To obtain insights in these viral community types, we investigated the virome compositional variation and found that viral community types were associated with distinct groups of viruses (Fig. 1b). The first viral community type (termed “CA”) revealed a high relative abundance of members of the \emph{Caudoviricetes} [non-CrAss] (n = 363, Mann-Whitney U, r = 0.137, AdjP = 0.0266; Supplementary Table 5), whereas the second (termed “CrM”) revealed a high relative abundance of members of the \emph{Caudoviricetes} [CrAss] and \emph{Malgrandaviricetes} (n = 363, Mann-Whitney U, AdjP < 0.05). Viral community types were characterized by a distinct \( \alpha \)-diversity (Fig. 1c).

Viral community type CrM revealed a higher viral richness (n = 363, Mann-Whitney U, \( r = 0.542, \) AdjP < 2.2e-16) and Shannon diversity (n = 363, Mann-Whitney U, \( r = 0.384, \) AdjP < 4.76e-10) compared to viral community type CA. We hypothesized that this diversity variation might be the result of a different phage lifestyle (lytic versus lysogenic), as determined by the presence or absence of lysogeny-specific proteins such as integrases (Supplementary Table 6)\textsuperscript{26}. We found that viral community type CrM, but not CA, was associated with a richness expansion of lytic compared to lysogenic phages (n = 121, Mann-Whitney U, \( r = 0.308, \) AdjP = 6.86e-06; Fig. 1d). In addition, only phages belonging to the \emph{Caudoviricetes} class were capable of lysogeny (\emph{Caudoviricetes} [non-CrAss] = 69.4\%, \emph{Caudoviricetes} [CrAss] = 10.8\%, \emph{Malgrandaviricetes} = 0\%; Fig. 1d). Overall, both community types showed a high relative abundance of lytic compared to lysogenic phages (Mann-Whitney U, AdjP < 0.05; Supplementary Table 5), but only in viral community CrM we observed a high relative abundance of lytic \emph{Caudoviricetes} phages (CrAss + non-CrAss, n = 121, Mann-Whitney U, \( r = 0.307, \) AdjP = 0.00360; Supplementary Table 5), suggesting that \emph{Caudoviricetes} prophages were induced in patients with viral community type CA.

Furthermore, viral community types were associated with IBD subtype (proportion test, \( X^2 = 5.20 \) AdjP = 0.0472; Supplementary Table 5) but not disease location (proportion test, \( X^2 = 1.04 \) AdjP = 0.791; Supplementary Table 5), as characterized by a higher prevalence of viral community type CA in CD compared to UC patients (CD = 69.8\%, UC = 57.9\%, Supplementary Table 5).

**Viral community type CA is mostly associated with the dysbiotic Bact2-enterotype**

Next, we implemented \emph{in silico} phage host prediction to determine the bacterial host. Most bacterial hosts were predicted on phylum-level (83.6\%) and only a few hosts could be predicted on genus-level (27.7\%; Extended Data Fig. 7). The majority of the host phyla were \emph{Bacteroidetes}, \emph{Firmicutes} and \emph{Proteobacteria}. Most of the host genera were \emph{Bacteroides} and \emph{Prevotella}, which were preferentially infected by \emph{Caudoviricetes} [non-CrAss] and \emph{Caudoviricetes} [CrAss] phages, respectively (\( \geq 1\% \) reads; Fig. 1e). Viral community type CrM revealed a high relative abundance of \emph{Prevotella}-infecting phages compared to viral community type CA (n = 363, Mann-Whitney U, \( r = 0.235, \) AdjP = 1.51e-05; Supplementary Table 5), but no significant differences were found in relative
abundance of *Bacteroides*-infecting phages between community types (*n* = 363, Mann-Whitney U, *r* = 0.00530, AdjP = 1.00; Supplementary Table 5).

The gut microbiota of IBD patients has been repeatedly associated with the dysbiotic Bact2-enterotype\(^9,23\). Extended Data Fig. 8 shows a statistical correlation between the bacterial enterotypes (Bact1, Bact2, Prev and Rum) and the viral community types. Viral community type CA was mostly associated with the Bact2-enterotype. This dysbiotic enterotype showed a higher prevalence in viral community type CA compared to viral community type CrM in UC (CA = 57.1%, CrM = 25.6%, *n* = 88, proportion test, \(X^2 = 7.55, \text{AdjP} = 0.012\); Supplementary Table 7) and in CD patients (CA = 79.6%, CrM = 38.3%, *n* = 207, proportion test, \(X^2 = 3.13, \text{AdjP} = 8.96e-08\)). Conversely, viral community type CrM was mostly associated with the Bact1-enterotype. This enterotype showed a higher prevalence in viral community CrM compared to viral community type CA in UC (CA = 16.3%, CrM = 48.7%, *n* = 88, proportion test, \(X^2 = 9.24, \text{AdjP} = 9.44e-03\)) and CD patients (CA = 12.2%, CrM = 35.0%, *n* = 88, proportion test, \(X^2 = 1.30, \text{AdjP} = 4.20e-04\)).

The gut virome composition is individual and is associated with disease location, patient's age, and moisture content of the faecal samples

The factors shaping the gut virome composition were determined by distance-based redundancy analysis (dbRDA) in the IBD cohort (Supplementary Table 9). Patient individuality was identified as the largest explanatory variable, thereby reinforcing the notion that the virome is highly individual-specific (*n* = 363, multivariate dbRDA, genus-like group, \(R^2 = 75.8\%\), AdjP = 0.001; Fig. 2a). Disease location was identified as the second largest explanatory variable of virome variation (*n* = 363, multivariate dbRDA, genus-like group, \(R^2 = 1.40\%\), AdjP = 0.001). Next, patient age and moisture content of the stool also showed a limited contribution to the virome variation (*n* = 363, multivariate dbRDA, genus-like group, \(R^2 = 0.5\%\), moisture \(R^2 = 0.3\%\), AdjP < 0.05). Remarkably, disease location (Montreal classification) stratifying patients according to ileal (L1CD), colonic (L2CD/UC) and ileocolonic (L3CD) phenotypes, revealed a higher explanatory power than simple diagnosis (UC/CD), and was observed in bacterial research as well (*n* = 363, univariate dbRDA, genus-like group, location \(R^2 = 1.34\%\), diagnosis \(R^2 = 0.47\%\), AdjP < 0.05)\(^23\). Despite our expectations the choice of biological therapy did not show an association (univariate or multivariate dbRDA) with the virome variation.

To further evaluate the factors shaping the gut virome composition in active IBD patients, we focused on baseline (pre-intervention) samples (Fig. 2b; Supplementary Table 10). We identified body-mass index (BMI), haemoglobin concentration, age, and smoking behaviour as unique contributors to gut virome variation in baseline samples (*n* = 151, multivariate dbRDA, genus-like group, \(R^2 = 1.94\%\), AdjP = 0.050, full model). We found that the patient's BMI correlated with hemoglobin concentration (*n* = 151, \(\rho = 0.230, \text{AdjP} = 0.00460\); Supplementary Table 11) and reasoned that a low BMI and a low hemoglobin concentration (anaemia) probably expressed frailty of active IBD patients. In addition, we hypothesized that the gut virome might have a predictive capacity to determine therapeutic outcome (remission or non-remission at primary endpoint). Unfortunately, none of the disease activity indices could be associated to virome composition at baseline; however, viral community types could be associated to endoscopic outcome (*n* = 151, AdjP < 0.050; Supplementary Table 10). Here, remitting patients harboured a 225% increased probability of hosting viral community type CrM, while simultaneously presenting a poor predictive power (univariate logistic regression,
AUC = 60.0, AdjP = 0.0348; Supplementary Table 10), suggesting that community types had slight pharmacodynamic capabilities able to track patient outcome.

The gut virome composition and community types are associated with endoscopic outcome in UC patients

To determine the effect of therapy outcome on the gut virome composition we investigated the association of all covariates on primary endpoint (post-intervention) samples in IBD patients (Fig. 2b; Supplementary Table 12). Disease location and patient’s age were once more identified as unique explanatory variable (n = 166, univariate dbRDA, genus-like group, Location R^2 = 1.01%, Age R^2 = 0.370%, AdjP < 0.050). Interestingly, therapeutic outcome appeared as a new covariate of the gut virome composition, as measured by the golden standard of endoscopic evaluation (n = 166, multivariate dbRDA, genus-like group, R^2 = 0.460%, AdjP = 0.0320). On the other hand, other disease activity indices (clinical and biomarker evaluation) did not have a significant contribution to the virome variation. Moreover, no association was found between remission rates and disease location (n = 138, proportion test, X^2 = 7.47, r = 0.233, AdjP = 0.0584; Supplementary Table 13), most likely because of limited samples sizes.

By focusing on post-intervention samples, this study premiered endoscopic outcome as a significant covariate of the gut virome composition in IBD patients. We hypothesized that virome covariates might be associated to viral community types in post-intervention samples, and therefore modelled the association between viral community types and the three significant covariates of virome variation (n = 166, logistic regression; Fig. 2c; Supplementary Table 14). The disease location and patient’s age could not be associated to viral community types (n = 166, AdjP > 0.05), while the endoscopic outcome could be associated to viral community types (n = 166, AdjP = 0.0280). Remitting patients harboured a 265% increased probability of hosting viral community type CrM (endoscopic remission relative risk (RR) = 2.65) and a 62% decreased probability of hosting viral community type CA (endoscopic non-remission RR = 0.38). Stratification of endoscopic outcome for each IBD subtype demonstrated that UC, but not CD patients, had a higher prevalence of viral community type CrM in remitting patients (Fig. 2d). Viral community type CrM was found in 54.8% of the remitting and 25.0% of the non-remitting UC patients (n = 51, proportion test, X^2 = 4.41, r = 0.254, AdjP = 0.0357; Supplementary Table 15). Conversely, viral community type CrM was found in 31.7% of the remitting and 19.6% of the non-remitting CD patients (n = 115, proportion test, X^2 = 2.21, r = 0.136, AdjP = 0.345; Supplementary Table 15). Next, we evaluated the impact of treatment success (endoscopic outcome) on the gut virome, as determined by compositional changes between baseline and primary endpoint (paired statistics, n = 2x103, Supplementary Table 9). There were no changes in the gut virome composition over time associated with endoscopic remission, across all patients (paired dbRDA, n = 2x53, R^2 = 0.869%, AdjP = 0.688), or following the stratification of the patients into corresponding UC and CD subtypes (Fig. 2e; Supplementary Table 14).

Intestinal inflammation in UC patients is associated with a low viral diversity and a high lysogenic potential of Caudoviricetes

To investigate the effect of the endoscopic outcome on the major viral classes in more detail we focused on post-interventional samples of the IBD patients. The major viral classes of UC, but not CD patients, were characterized by a distinct α-diversity between remitting and non-remitting patients (Fig. 3a,c; Supplementary Table 16–18). Non-remitting UC patients revealed a low Shannon diversity of Caudoviricetes [non-CrAss] (n = 51,
Mann-Whitney U, r = 0.372, AdjP = 0.0143; Supplementary Table 16) and *Malgrandaviricetes* (*n* = 51, Mann-Whitney U, r = 0.329, AdjP = 0.0378; Supplementary Table 16) phages compared to remitting UC patients. Other metrics (Pielou’s evenness and richness) were not seen to be affected by the endoscopic outcome. Conversely, none of the metrics (Shannon diversity, Pielou’s evenness and richness) were affected by the endoscopic outcome in CD patients (*n* = 94, Mann-Whitney U, AdjP < 0.05; Supplementary Table 16–18). Next, we hypothesized that therapeutic outcome might be associated to a changing phage lifestyle (Fig. 3b,c). A high relative abundance of lysogenic phages was observed in non-remitting compared to remitting UC patients (*n* = 51, Mann-Whitney U, r = 0.335, AdjP = 0.0344), while no differences in phage lifestyle were found between non-remitting and remitting CD patients (*n* = 94, Mann-Whitney U, r = 0.219, AdjP = 0.0677; Supplementary Table 19). Moreover, the relative abundance of lysogenic phages in (non)-remitting UC patients was not associated to viral community types (*n* = 20, Mann-Whitney U, r = 0.127, AdjP = 1.00; Supplementary Table 19). Therefore, the expanded lysogenic potential of phages was thought to be associated with endoscopic outcome rather than viral community types. The high lysogenic potential in non-remitting UC patients appeared to be associated with the entire class of *Caudoviricetes* phages ([CrAss + non-CrAss], *n* = 51, Mann-Whitney U, r = 0.313, AdjP = 0.0498; Supplementary Table 20). To evaluate the finding by Gulyaeva and colleagues that CrAss-like phages in the human gut were depleted in IBD patients, we also assessed the association between the prevalence of *Caudoviricetes* [CrAss] and endoscopic outcome\(^{20}\). UC patients showed a prevalence of 77.4% in remission compared to 60.0% in non-remission (*n* = 51, proportion test, \(\chi^2 = 1.04\), AdjP = 0.617; Supplementary Table 21). CD patients showed a prevalence of 71.9% in remission compared to 62.1% in non-remission (*n* = 94, proportion test, \(\chi^2 = 1.76\), AdjP = 0.419). In addition, no associations were found between the relative abundance of *Caudoviricetes* [CrAss] and endoscopic outcome.

**Individual gut phages are associated with treatment success in IBD patients undergoing biological therapy**

To investigate the capacity of individual phages to determine therapeutic success, we focussed on pre-interventional samples of IBD patients. We argue that these phages should be (1) highly shared between patients and (2) differentiate between endoscopic treatment success (remission versus non-remission). Firstly, the 20 most prevalent viruses were determined for each IBD subtype and at least one of them was found in 97.7% of UC and 87.6% of CD patients (Extended Data Fig. 9, Supplementary Table 22). Secondly, phages with a capability to differentiate between endoscopic treatment success were identified for each IBD subtype (LDA score (log10) > 2, AdjP < 0.05; Fig. 4a, Supplementary Table 23). Accordingly, five individual novel phages (both highly prevalent and differentially present) were found to meet these criteria (*n*\(_{UC}\) = 44, *n*\(_{CD}\) = 99; Fig. 4b; Supplementary Table 24–25). Specifically, UC patients harboured two prevalent phages associated with endoscopic remission (CrAssella-R and Cripes-R) and one prevalent phage associated with endoscopic non-remission (Croccus-NR). On the Contrary, CD patients harboured two prevalent phages associated with endoscopic remission (CrAssella-R and Croides-R) and one prevalent phage associated with endoscopic non-remission (Croides-NR). CrAssella-R is a *Caudoviricetes* [CrAss] phage that was associated with treatment success and the only phages detected in both of the IBD subtypes. We argued that next-generation sequencing methods, in combination with stringent bioinformatic criteria might underestimate the presence and quantity of phages in IBD patients. Therefore, qPCR was used to quantify and validate the association with endoscopic outcome of the five above mentioned phages (Supplementary Table 26). Consequently, two phages in UC (Cripes-R and Croccus-NR) revealed a higher qPCR positivity rate than NGS positivity rate (*n* = 44, qPCR =
79.5%/79.5%, NGS = 16.9%/18.2%, respectively; Fig. 4c; Supplementary Table 27). Three phages in CD (CrAssella-R, Croides-R and Croides-NR) revealed a higher qPCR positivity rate as well (n = 88, qPCR = 21.4%/52.4%/36.4%, NGS = 16.2%/21.2%/10.1%; Supplementary Table 27). Only one phage (CrAssella-R) revealed a slightly lower qPCR positivity rate (n = 44, qPCR = 22.7%, NGS = 29.5%). Next, the true predictive capacity (IBD predictive value [IPV]) of individual phages was determined based on the viral copy number (Supplementary Table 27). Remitting IBD patients were predicted by a positive IPV (n = 82, IPV > 0, sensitivity = 0.67; Fig. 4d; Supplementary Table 27) and non-remitting IBD patients were predicted by a negative IPV (n = 82, IPV < 0, specificity = 0.68). No predictions were made for IBD patients with a neutral IPV (n = 46, IPV = 0).

Next, we modelled the association between prevalence of remission and IPV and found a predictive power of 72.8% (n = 82, logistic regression, AUC = 72.8%, AdjP = 0.000394; Fig. 4e; Supplementary Table 28). At last, we assessed a combination of IPV and viral community types and found that the latter variable does not significantly contribute to the model. In general, individual viruses revealed a capacity to determine therapeutic success in IBD patients, in which CrAss-like phages play an important role.

**Discussion And Conclusion**

To date, this research represents one of the most comprehensive gut virome studies (using viral particle enrichment followed by deep sequencing) of IBD patients undergoing biological treatment. Here, we implemented the concept of viral community typing and found that patients could be stratified into two (CA/CrM) virome constellations (Fig. 1a). We hypothesize that, patients harbouring viral community type CA exhibited a dysbalanced virome, characterised by a high abundance of *Caudoviricetes* [non-CrAss] and low viral diversity, which has been observed before (Fig. 1b,c)\(^{12}\). The association with the dysbiotic Bact2-enterotype strengthens the belief that this constellation reflects the viral counterpart of gut dysbiosis (Extended Data Fig. 8). Conversely, patients harbouring viral community CrM might exhibit a more eubiotic-like virome characterised by a high abundance of members of the *Caudoviricetes* [CrAss] and *Malgrandaviricetes*, a high viral diversity and a shift away from the dysbiotic Bact2-enterotype.

Furthermore, remitting UC patients are associated to viral community type CrM, whereas non-remitting UC patients are associated to viral community type CA, thereby reinforcing the notion that viral community types might reflect a potential dysbalanced or eubiotic-like state of the gut virome, and may therefore be clinically relevant (Fig. 2b-d). However, these findings could not be reproduced for CD patients suggesting that these patients retain a dysbalanced state even in the case of remission or quiescent disease. Thus far, twin studies revealed a larger genetic impact in CD pathology and in contrast, a larger non-genetic or environmental (e.g., virome) impact in UC pathology\(^{27}\). These findings are corroborated in this study, further revealing a substantial role of the gut microbial ecosystem in UC pathology (Fig. 3). Consequently, the thorough analysis of the gut inflammation in IBD revealed a decrease in viral diversity and an expansion of lysogenic phages. However, changes in viral diversity and phage lifestyle were not observed in CD pathology, thereby once more confirming the different role of the virome regarding the IBD subtype.

We argue that a persistent inflamed state of the intestine will shape the gut virome by acting on different mechanisms. Firstly, an increase in gut motility might lower the microbial (and viral) diversity due to an increase defecation frequency\(^{28}\). Secondly, phage-mediated lysis might explain the observed expansion of lysogenic gut phages\(^{29}\). Phage-mediated lysis describes a positive-feedback loop between phage induction and intestinal
inflammation. Briefly, inflammation stimulates enterocytes to produce stressors (e.g. reactive oxygen species) activating a stress response in the host bacteria (‘SOS response’). The stress response will trigger prophages to initiate the lytic lifecycle, leading to the lysis of the bacterial host cell. An increased bacterial lysis will be accompanied by an increase of pathogen-associated molecular patterns (e.g., lipopolysaccharide, bacterial DNA) that are stimulating pattern recognition receptors on enterocytes. These cells will produce more stressors and further promote prophage induction, thus, starting a positive feedback-loop and increase the lysogenic potential under intestinal inflammation. Taken together, these mechanisms might explain the observed decrease in viral diversity and lysogenic expansion under inflammatory conditions (Fig. 3).

Following the viral exploration of the gut inflammation, we also discovered five novel phages that were associated with treatment success, as confirmed by qPCR results (Fig. 4). One phage, a novel CrAss-like phage (CrAssella-R), revealed an association with treatment success for both IBD subtypes. Shkoporov and colleagues described that long-term persistence of crAss0001 with its bacterial host could drive diversity (hallmark of eubiosis) by a process called phase variation\(^\text{21}\). We argue that biological treatment induces remission and that mechanism such as phase variation could provide a path to maintain remission by stimulating eubiosis. Phages with predictive abilities might be those capable of driving diversity in combination with the respective host.

In conclusion, in this study we have shown that, viral community types exist and allow the stratification of IBD patients based on a distinct viral composition in the gut, and could be used to better understand IBD subtypes, pathology and disease activity in the future.

**Methods**

Faecal samples were collected from a prospective IBD cohort (126 CD and 55 UC) based on sample availability and pairing (\(n = 432\)). All IBD patients had active disease at baseline defined by endoscopy and were started on one of four approved biological therapies (infliximab, adalimumab, ustekinumab or vedolizumab). Every patient provided a baseline and a follow-up (primary endpoint) sample. The NetoVIR protocol was used to prepare faecal samples for viral metagenomics, as described before (Extended Data Fig. 2A)\(^\text{30}\). Further bioinformatic processing, viral community typing, diversity analysis and other details regarding the protocols were elaborately described in the attachment (Extended Data Methods).

**Declarations**

**Ethics approval and consent to participate**

The study was approved by the ethical commission of UZ Leuven (KU Leuven, reference number: S53684). Participants provided signed informed consent to participate with the study. The design of the study was in accordance with the Declaration of Helsinki and Belgian privacy law.

**Availability of data and materials**

Metadata can be found in Supplementary Table 1. The raw sequence data were deposited to the NCBI Sequence Read Archive under the BioProject accession number PRJNA804384. Sequences (predictive markers) were
deposited to GenBank under the following accession numbers: ON493177-ON493181. The ViPER (Virome Paired-End Reads pipeline) script was used to process raw paired-end reads and is publicly available at https://github.com/Matthijnssenslab/ViPER. All the data required to reproduce virome analyses will be made available at https://github.com/Matthijnssenslab/IBDVirome.

Competing interest

SV has received grants from AbbVie, J&J, Pfizer, Galapagos, Takeda. SV has received consulting and/or speaking fees from AbbVie, AbolerIS Pharma, AgomAb, Alimentiv, Arena Pharmaceuticals, AstraZeneca, Avaxia, BMS, Boehringer Ingelheim, Celgene, CVasThera, Dr Falk Pharma, Ferring, Galapagos, Genentech-Roche, Gilead, GSK, Hospira, Imidomics, Janssen, J&J, Lilly, Materia Prima, MiroBio, Morphic, MrMHealth, Mundipharma, MSD, Pfizer, Prodigest, Progenity, Prometheus, Robarts Clinical Trials, Second Genome, Shire, Surrozen, Takeda, Theravance, Tillots Pharma AG, Zealand Pharma. Other authors report no conflict of interest. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to JM (jelle.matthijnssens@kuleuven.be).

Funding

This research was supported by the ‘Fonds Wetenschappelijk Onderzoek’ (Research foundation Flanders) (Daan Jansen: 1S78021N; Kathleen Machiels: 12M9118N).

Author Contributions

The study was conceived by JM, SV and JR. Experiments were designed by DJ, SV, JR and JM. Sampling was set up by CC, KM and SV. Experiments were performed by DJ, TM and CS. Bioinformatic and statistical analysis of the sequences reads was performed by DJ, SS and GF. DJ, CS, GF, SS and JM drafted the manuscript. All authors revised the article and approved the final version for publication.

Acknowledgements

We would like to thank all participating patients for their contribution. We thank Nooshin Ardeshir Davani for technical support regarding faecal sample logistics and Prof. Dr. Jurgen Vercauteren for the support in statistical analysis. We thank Prof. Dr. Evelien Adriaenssens for assisting in the nomenclature of the phages with predictive potential and the VIB nucleomics Core for sequencing production (www.nucleomics.be). We thank Winston Chiu for extensive proofreading of the manuscript.

Author information

SV has received grants from AbbVie, J&J, Pfizer, Galapagos, Takeda. SV has received consulting and/or speaking fees from AbbVie, AbolerIS Pharma, AgomAb, Alimentiv, Arena Pharmaceuticals, AstraZeneca, Avaxia, BMS, Boehringer Ingelheim, Celgene, CVasThera, Dr Falk Pharma, Ferring, Galapagos, Genentech-Roche, Gilead, GSK, Hospira, Imidomics, Janssen, J&J, Lilly, Materia Prima, MiroBio, Morphic, MrMHealth, Mundipharma, MSD, Pfizer, Prodigest, Progenity, Prometheus, Robarts Clinical Trials, Second Genome, Shire, Surrozen, Takeda, Theravance, Tillots Pharma AG, Zealand Pharma. Other authors report no conflict of interest. Readers are
welcome to comment on the online version of the paper. Correspondence should be addressed to JM (jelle.matthijnssens@kuleuven.be).

References


Figures
Figure 1

**Viral community types to stratify individuals based on their virome composition.** A, Principal coordinates analysis (PCoA) of inter-individual differences of the gut virome composition (genus-like group, Bray-Curtis dissimilarity) of the IBD cohort (coloured by viral community types, n=363) with shape representing categorical richness (median richness=31, open circles<31 (low richness), closed circles³31 (high richness), R²=0.0193, P=0.001, Supplementary Table 8). B, Boxplot showing relative abundance of major phage classes (³1% of reads) stratified according to viral community type (n=363, Mann-Whitney U, AdjP<0.05). C, Boxplot showing alpha-diversity metrics (Observed richness and Shannon diversity) stratified according to viral community type (n=363, Mann-Whitney U, AdjP<0.05). D (left), Violin plot showing comparisons of observed richness stratified according to phage lifestyle (lytic versus lysogenic) within each viral community type (comparison within viral community type CrM, n=121, Mann-Whitney U, AdjP<0.05) and between viral community types (n=363, Mann-
Whitney U, AdjP<0.05). **D (right),** Barplot showing the distribution of phage lifestyles (lytic versus lysogenic) for major phage classes (≥1% of reads). **E (left),** Donut plot visualizing host prediction of phages on contig and read level (≥1% of reads). **E (right),** Boxplot visualizing host prediction of phages on read level (relative abundance) stratified according to viral community type (comparison *Prevotella*-infecting phages, *n*=363, Mann-Whitney U, AdjP<0.05). Adjustment for multiple testing (AdjP) was performed using the Benjamini-Hochberg method. Significant associations (AdjP<0.05) were visualized by an astrix (*). Abbreviations: Relative abundance (RA), Community type *Caudoviricetes* [non-CrAss] (CA) and community type CrAss-*Malgrandaviricetes* (CrM).

**Figure 2**

**Virome covariates of the IBD cohort, and their association to viral community types. A,** Principal coordinate analysis of inter-individual differences of the gut virome composition (genus-like group, Bray-Curtis dissimilarity) of the IBD cohort (coloured by viral community types, *n*=363). Insert top: Barplot represents significant covariates of virome composition of the IBD cohort, as identified in a multivariate model. The cumulative model (patient ID, age, location and moisture content) explains 78.0% of the virome variation. **B,**
Metadata variables correlating significantly to virome variation in (left) pre-intervention and (right) post-intervention samples (dbRDA, genus-like group, Bray-Curtis dissimilarity). Effect sizes of correlating metadata are calculated independently (univariate coloured in black) or in a cumulative model (multivariate coloured in black). C, D, E, Modeling the association between the prevalence of viral community type CrM and significant non-redundant covariates (location, therapeutic outcome as measured by endoscopy and age) of post-intervention samples (logistic regression model, n=166). C, Relative risk ratio of IBD patients hosting community type CrM associated with significant non-redundant covariates of the virome composition in post-intervention samples (endoscopic outcome [remission], n=166, RR=2.65, AdjP<0.05). D, Barplot showing viral community type (CA/CrM) prevalence in post-intervention samples stratified according to endoscopic outcome (R, NR, U) for UC (left) and CD (right) patients (comparison UC endoscopic outcome, n=51, proportion test, AdjP<0.05). E, Alluvial diagram showing transitions of viral community types after successful intervention for UC (left) and CD (right) patients (logistic regression, n=79x2, AdjP>0.05). A total of 25.0% of UC and 9.09% of CD patients shift towards viral community type CrM, whilst only 20.0% of UC and 9.09% of CD patients shift towards viral community type CA. Adjustment for multiple testing (AdjP) was performed using the Benjamini-Hochberg method. Significant associations (AdjP<0.05) were visualized by an astrix (*). Abbreviations: Ulcerative colitis (UC), Crohn's disease (CD), Remission (R), Non-remission (NR), Unknown (U), Community type Caudoviricetes[non-CrAss] (CA) and community type CrAss-Malgrandaviricetes (CrM).
Figure 3

Virome characteristics of post-intervention samples stratified according to endoscopic outcome. A, Boxplot showing alpha-diversity metrics (Observed richness and Shannon diversity) and evenness (Pielou’s evenness) for major phages classes (> 1% reads) stratified according to endoscopic outcome (NR/R) for UC (left) and CD (right) patients (comparison Malgrandaviricetes and Caudoviricetes [non-CrAss] diversity in UC, Mann-Whitney U test, AdjP<0.05). B, Boxplot showing relative abundance of lysogenic (top) and lysogenic Caudoviricetes phages stratified according to endoscopic outcome (NR/R) for UC (left) and CD (right) patients (comparison lysogenic (Caudoviricetes [CrAss + non-CrAss]) phages in UC, Mann-Whitney U test, AdjP<0.05). C, Graphical summary of virome characteristics detected in post-intervention samples. Adjustment for multiple testing (AdjP) was
performed using the Benjamini-Hochberg method. Significant associations (AdjP<0.05) were visualized by an asterix (*). Abbreviations: Remission (R), Non-remission (NR), Community type Caudoviricetes [non-CrAss] (CA) and community type CrAss-Malgrandaviricetes (CrM).

Figure 4

**Individual gut phages are associated with endoscopic outcome in IBD patients undergoing biological therapy.** A. Pre-interventional differential abundance analysis of gut viruses to determine endoscopic outcome (non-remission/remission) by linear discriminant analysis (LDA) and effect size (Lefse) (LDA score (log10) $³ 2$, $n=143$, Mann-Whitney U test, AdjP<0.05). The NODE names of prevalent viruses to determine endoscopic outcome within UC (left, $n=44$) or CD (right, $n=99$) patients are coloured in white. B, Barplot representing the
prevalence of five phages associated to determine endoscopic outcome for UC (left, three phages) and CD (right, three phages) patients. NODE_1_B26 is a prevalent phage associated with endoscopic outcome in both IBD subtypes (bold). NGS positivity rate (prevalence in baseline samples) is shown for each phage in the white boxes (\(n=143\)). C, Scatterplot representing the number of viral copies in baseline samples for each of these five phages (\(n=128\)). qPCR positivity rate (prevalence in baseline samples) is shown for each phage in the white boxes (limit of detection \(>100\) viral copies, represented by red dashed line) for IBD subtypes. D, The association of individual phages with endoscopic outcome is measured by the concept of IPV based on viral copy number. IPV is calculated by the natural logarithm of the remission phages over the non-remission phages for IBD subtypes (bottom formula, limit of quantification \(>500\) viral copies). D, Mosaic plot displaying the distribution of (excluding samples with no predictions) the categorical variables endoscopic remission and qPCR prediction (based on IPV, sensitivity=0.67, specificity=0.68). E, Modelling the association between prevalence of remission and IPV (logistic regression, \(n=88\), RR=1.32, AdjP<0.05). ROC curve representing the capability of individual phages to determine endoscopic outcome (\(n=88\), AUC=72.8%). Adjustment for multiple testing (AdjP) was performed using the Benjamini-Hochberg method. Abbreviations: IBD predictive value (IPV), Ulcerative colitis (UC), Crohn's disease (CD), Inflammatory bowel disease (IBD), Next-generation sequencing (NGS) and Receiver operating characteristic (ROC) curve.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Figure1Metadataendoscopicremission.pdf
- Figure2Wetlabdrylab.pdf
- Figure3Readsandcontigs.pdf
- Figure4Characteristicgutvirome.pdf
- Figure5Eukaryoticvirusesgenus.pdf
- Figure6BICstabilityclusternumber.pdf
- Figure7Hostpredictionadditions.pdf
- Figure8Enterotypeprofiles.pdf
- Figure9Baselineheatmaptop20UCCD.pdf
- SupplementaryTableDJ.xlsx