



Whole-Genome Metagenomic Analysis of the Gut Microbiome in HIV-1-Infected Individuals on Antiretroviral Therapy

Xiangning Bai^{1,2,3,4*}, Aswathy Narayanan², Piotr Nowak^{2,5,6}, Shilpa Ray^{1,6}, Ujjwal Neogi¹ and Anders Sönnnerborg^{1,2,5}

¹ Division of Clinical Microbiology, Department of Laboratory Medicine, ANA Futura, Karolinska Institutet, Stockholm, Sweden, ² Division of Infectious Diseases, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden, ³ State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, ⁴ Division of Laboratory Medicine, Oslo University Hospital, Oslo, Norway, ⁵ Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden, ⁶ The Laboratory for Molecular Infection Medicine Sweden MIMS, Umeå University, Umeå, Sweden

OPEN ACCESS

Edited by:

George Tsiamis,
University of Patras, Greece

Reviewed by:

Alessandra Borgognone,
IrsiCaixa, Spain
Sonny Lee,
Kansas State University, United States

*Correspondence:

Xiangning Bai
xiangning.bai@ki.se

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 14 February 2021

Accepted: 31 May 2021

Published: 25 June 2021

Citation:

Bai X, Narayanan A, Nowak P,
Ray S, Neogi U and Sönnnerborg A
(2021) Whole-Genome Metagenomic
Analysis of the Gut Microbiome
in HIV-1-Infected Individuals on
Antiretroviral Therapy.
Front. Microbiol. 12:667718.
doi: 10.3389/fmicb.2021.667718

Gut microbiome plays a significant role in HIV-1 immunopathogenesis and HIV-1-associated complications. Previous studies have mostly been based on 16S rRNA gene sequencing, which is limited in taxonomic resolution at the genus level and inferred functionality. Herein, we performed a deep shotgun metagenomics study with the aim to obtain a more precise landscape of gut microbiome dysbiosis in HIV-1 infection. A reduced tendency of alpha diversity and significantly higher beta diversity were found in HIV-1-infected individuals on antiretroviral therapy (ART) compared to HIV-1-negative controls. Several species, such as *Streptococcus anginosus*, *Actinomyces odontolyticus*, and *Rothia mucilaginosa*, were significantly enriched in the HIV-1-ART group. Correlations were observed between the degree of immunodeficiency and gut microbiome in terms of microbiota composition and metabolic pathways. Furthermore, microbial shift in HIV-1-infected individuals was found to be associated with changes in microbial virulome and resistome. From the perspective of methodological evaluations, our study showed that different DNA extraction protocols significantly affect the genomic DNA quantity and quality. Moreover, whole metagenome sequencing depth affects critically the recovery of microbial genes, including virulome and resistome, while less than 5 million reads per sample is sufficient for taxonomy profiling in human fecal metagenomic samples. These findings advance our understanding of human gut microbiome and their potential associations with HIV-1 infection. The methodological assessment assists in future study design to accurately assess human gut microbiome.

Keywords: gut microbiome, shotgun metagenome sequencing, HIV-1 infection, virulome, resistome

INTRODUCTION

It is well-recognized that gut microbiome (GM) is a key player in intestinal and physiological homeostasis, immunity, and energy metabolism. Dysbiosis in the GM may alter intestinal barrier functions, host metabolic and signaling pathways, which are directly or indirectly linked to various diseases and metabolic disorders such as obesity (Turnbaugh et al., 2006), diabetes

(Sonnenburg and Backhed, 2016), autoimmune diseases (Severance et al., 2016), cardiovascular disease (Karlsson et al., 2013), inflammatory bowel disease (Ananthakrishnan et al., 2018), cancer (Schwabe and Jobin, 2013), and neuropsychiatric disorders (Mathee et al., 2020). Accumulating studies in recent years have indicated that GM plays a significant role in human immunodeficiency virus type 1 (HIV-1) immunopathogenesis as well as in HIV-1-associated chronic complications (Vujkovic-Cvijin et al., 2013; Nowak et al., 2015; Dillon et al., 2016). The gastrointestinal tract is severely and rapidly damaged following HIV-1 infection (Hunt et al., 2014), resulting in structural impairment of the epithelial barrier and a disruption of intestinal homeostasis. The resulting translocation of bacterial products and bacteria themselves from the lumen into the systemic circulation is associated with local and systemic inflammation, immune activation, and deregulation (Svard et al., 2015), which are not fully restored with antiretroviral therapy (ART) (Dillon et al., 2016; Ray et al., 2020).

Cross-sectional studies have shown reduced gut microbiota diversity and compositional shifts from *Bacteroides* to *Prevotella* predominance after HIV-1 infection (Vazquez-Castellanos et al., 2015; Ling et al., 2016; Dillon et al., 2017; Lu et al., 2018), which have been linked to lower CD4 + T-cell count, higher inflammation, and increased immune activation (Vujkovic-Cvijin et al., 2013; Dinh et al., 2015). However, in some studies, such microbiota shifts have been observed neither in animal models (Handley et al., 2012) nor in studies matching for HIV-1 risk groups (Noguera-Julian et al., 2016; Kelley et al., 2017). Importantly, previous microbiome studies in HIV-1 infection have mostly been based on 16S rRNA gene sequencing approach, which has several limitations, for instance, taxonomic resolution at the genus level (Knight et al., 2018). To date, limited HIV-1 microbiome studies have been conducted using whole metagenome sequencing, which can characterize the microbiome at the species and gene levels. The sequencing depth of a few published studies was inadequate (Vazquez-Castellanos et al., 2015, 2018; D'Souza et al., 2019; Rocafort et al., 2019) and underpowered to identify granular differences in both microbial composition and function (Rocafort et al., 2019). It is noteworthy that one previous study indicated that enrichment of bacterial virulence factors and antimicrobial resistance (AMR) genes were observed in HIV-1-infected individuals with progressively lower nadir CD4 + T-cell counts (Guillen et al., 2019). It remains to be seen if such alterations contribute to immune suppression in HIV-1 infection. A reliable and comprehensive pattern of gut microbiome dysbiosis and its impact on HIV-1 disease progression are yet to be fully deciphered.

In the current study, we set up whole metagenomics workflows and characterized the gut microbiome features in terms of microbial composition at the species level, functional genes, metabolic pathways, bacterial virulence factors, and AMR genes, with respect to HIV-1 infection. As previous studies have shown that different DNA extraction methods have an impact on DNA yield and quality, which may influence the structure of bacterial communities (Costea et al., 2017; Angebault et al., 2018; Lim et al., 2018), we thus firstly evaluated the effect of different genomic DNA (gDNA) extraction protocols on the DNA quantity and

quality. Whole metagenome sequencing was then performed on fecal samples from HIV-1-infected individuals on ART (HIV-1-ART) from the Swedish InfCareHIV cohort, as well as HIV-1-negative controls. We evaluated whole metagenome sequencing depth required for recovery of the bacterial species, virulence factors genes, and AMR genes. We assessed proportions of different types of microbes and functionally characterized genes in human gut metagenomic samples. Finally, metagenomics data were analyzed in combination with clinical variables to determine potential associations between gut microbiome and HIV-1 infection.

MATERIALS AND METHODS

Study Participants

This cross-sectional study included HIV-1-ART patients who were recruited from the outpatient clinic at Karolinska University Hospital, Stockholm, Sweden, and HIV-1-negative controls. The inclusion criteria for the HIV-1-ART participants were age >18 years (median: 45 years; range: 38–62 years), and HIV-1 positive for at least 6 months. The exclusion criteria were ongoing HIV-1-related complications or antibiotic treatment during the previous 3 months. The patients received ART for a median of 7.7 years (range: 4.4–20.8 years). The HIV-1-negative controls were healthy individuals (median: 32 years; range: 24–51 years) who did not receive antibiotic treatment during the last 3 months.

This study was approved by the Regional Ethics Committee, Stockholm (2009/1485-31, 2013/1944-31/4, 2014/920-3).

Samples and Data Collection

A sterile tube for fecal sampling without preservation media was used when participants were able to donate feces at the clinic as previously described (Vesterbacka et al., 2017). Samples were frozen and stored at -80°C within 24 h. Stool collection tube with DNA stabilizer (Stratec Biomedical) was used for participants who collected the feces at home. The samples were delivered to the outpatient clinic by the participants or instantly sent by post and stored at -80°C according to the manufacturer's instructions.

Clinical data of the HIV-1-ART participants were collected from the InfCareHIV database, including gender, age, current, and nadir CD4 + T-cell counts, ART regimen at sampling date, total ART duration (years), and transmission mode (Table 1).

Metagenomic DNA Extraction With Different Protocols

First, we did literature review and selected and evaluated two widely used protocols, i.e., QIAamp PowerFecal Pro DNA Isolation kit (QP) (Qiagen, Germany) and the standardized International Human Microbiota Standards (IHMS) Protocol Q recommended by the International Human Microbiome Consortium¹ (Costea et al., 2017), which performed better than

¹<http://www.microbiome-standards.org>

TABLE 1 | Characteristics of study subjects.

Sample ID	Group	Age at sampling date (years)	Gender	CD4 + T-cell count (cells/mm ³)		ART at sampling date	Total ART duration (years)	Transmission mode
				Current	Nadir			
Q-C1	Case	43	Male	440	270	Abacavir/Lamivudine/Dolutegravir	4.4	Heterosexual
Q-C2	Case	48	Female	400	150	Abacavir/Lamivudine/Dolutegravir	8.3	Heterosexual
Q-C3	Case	38	Male	770	344	Tenofovir (TAF)/Emtricitabine/Rilpivirine	6.4	MSM
Q-C4	Case	42	Female	810	530	TDF/Emtricitabine/Rilpivirine	5.1	Heterosexual
Q-C5	Case	43	Female	740	430	TDF/Emtricitabine/Dolutegravir	5.7	Heterosexual
Q-C6	Case	54	Female	690	280	TDF/Emtricitabine/Efavirenz	6.8	Heterosexual
Q-C7	Case	56	Female	800	570	TDF/Emtricitabine/Rilpivirine	5.0	Heterosexual
Q-C8	Case	45	Male	680	300	Abacavir/Lamivudine/Dolutegravir	8.1	Heterosexual
Q-C9	Case	62	Male	330	140	TAF/Emtricitabine/Dolutegravir	8.8	Drug use
Q-C10	Case	38	Female	780	300	Abacavir/Lamivudine/Dolutegravir	7.7	Heterosexual
Q-D2	Case	44	Female	1020	273	Dolutegravir/Abacavir/Lamivudine	20.8	Heterosexual
Q-D3	Case	55	Female	390	170	TDF/Emtricitabine/Dolutegravir	20.7	Blood product
Q-D4	Case	50	Female	970	380	TDF/Emtricitabine/Dolutegravir	10.9	Heterosexual
Q-H1	Control	29	Male	-	-	-	-	-
Q-H2	Control	35	Male	-	-	-	-	-
Q-H3	Control	51	Male	-	-	-	-	-
Q-H4	Control	32	Male	-	-	-	-	-
Q-H5	Control	40	Male	-	-	-	-	-
Q-H6	Control	24	Male	-	-	-	-	-
Q-H7	Control	30	Male	-	-	-	-	-
Q-H8	Control	27	Male	-	-	-	-	-
Q-H9	Control	42	Male	-	-	-	-	-
Q-H10	Control	33	Male	-	-	-	-	-
Q-H11	Control	30	Male	-	-	-	-	-

other methods in terms of both DNA quality and microbial diversity in human fecal microbiota studies (Angebault et al., 2018; Lim et al., 2018; Fiedorova et al., 2019), while never having been compared before. A subset of fecal samples was randomly selected from cases and controls and then subjected to the two protocols according to instructions with minor modifications. The quantity and quality of gDNA extracted by the two protocols were measured using NanoDropTM One (Thermo Scientific, United States) and Agilent 2200 TapeStation (Agilent Technologies, United States) including gDNA concentration, purity, and integrity. The protocol yielding higher gDNA quantity and better quality was then used to extract gDNA of the remaining samples. The gDNA samples were stored at -80°C until library preparation and sequencing.

Library Preparation and Shotgun Metagenome Sequencing

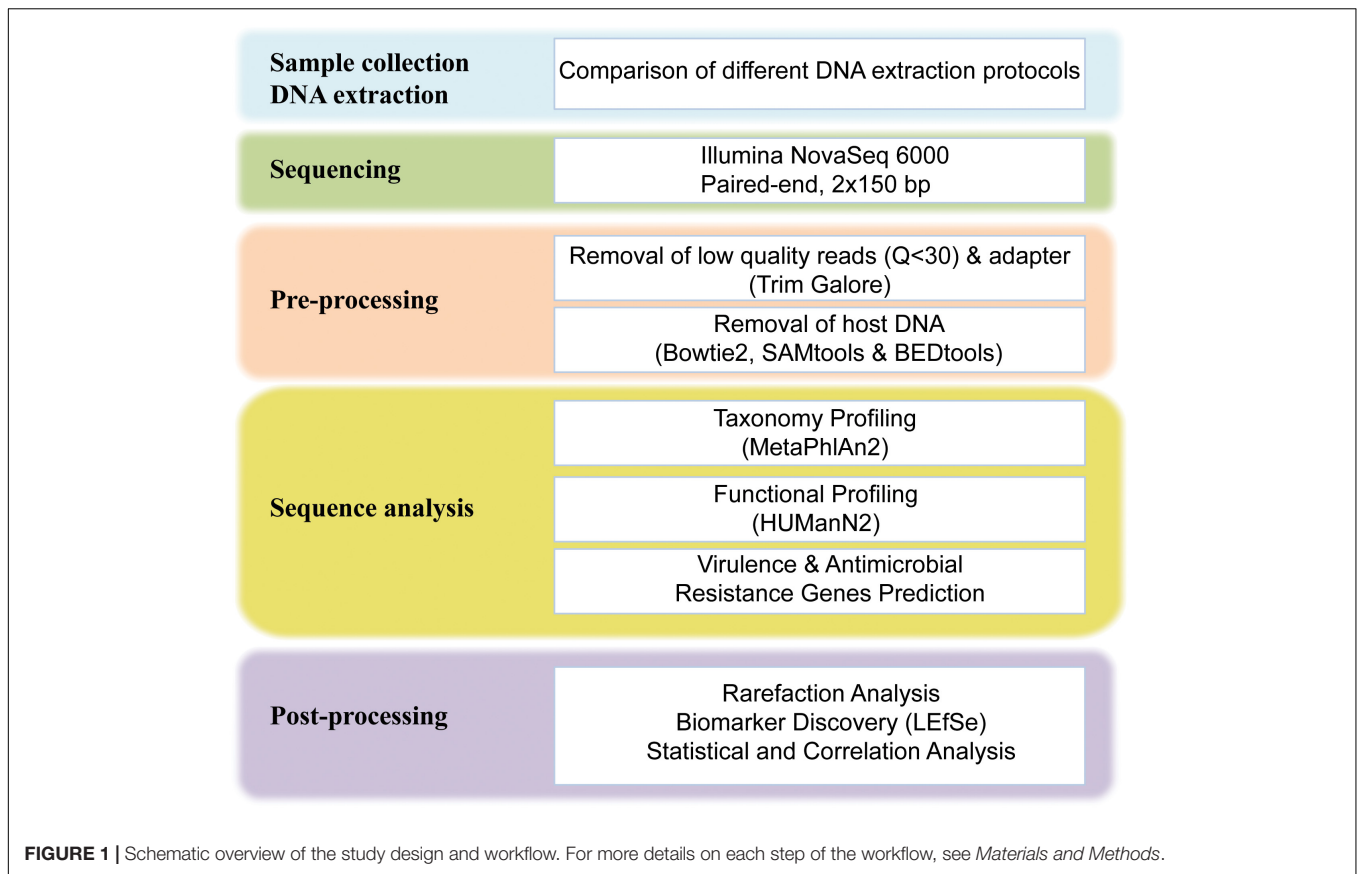
Sequencing libraries were prepared with the Nextera DNA Flex kit (Illumina, Inc.) following the manufacturer's instructions. One paired-end (PE) library with insert size of approximately 320 bp was constructed for each sample. Libraries were normalized with Qubit assay; the pooled library was then sequenced on one lane on NovaSeq6000 platform (NovaSeq

Control Software 1.6.0/RTA v3.4.4) with a 2×150 setup using "NovaSeqXp" workflow in "S1" mode flowcell. The Bcl to FastQ conversion was performed using bcl2fastq_v2.20.0.422 from the CASAVA software suite. The quality scale used is Sanger/phred33/Illumina 1.8 +.

Data Preprocessing

A total of 146 GB raw sequencing data were generated for 24 fecal samples, approximately 6.1 GB per sample was obtained. The raw sequencing data were preprocessed using our in-house bioinformatics pipelines (**Figure 1**). Briefly, the adapter and low-quality reads (a quality score of less than Q30) were removed using Trim galore (v0.6.4).² After the quality trimming, Bowtie2 (v2.3.5.1) (Langmead and Salzberg, 2012) was used in combination with SAMtools (v1.19) (Li et al., 2009) and BEDtools (v2.29.2) (Quinlan and Hall, 2010) to identify and remove human DNA sequences. The human unmapped reads were then used for downstream analysis. After filtering, an average of 41 million 150-bp PE reads per sample was obtained, the adapter and low-quality reads accounted for 1.32% of total reads, and the average rate of human DNA contamination in all samples was 0.58% (**Supplementary Table S1**).

²https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/



Taxonomical Profiling

The taxonomic assignment and abundance estimation were performed with MetaPhlAn 2.0 (Truong et al., 2015) using default parameters. MetaPhlAn 2.0 relies on ~1 million unique clade-specific marker genes identified from ~17,000 reference genomes (~13,500 bacterial and archaeal, ~3,500 viral, and ~110 eukaryotic genomes) (Truong et al., 2015). This computational tool provides the relative abundances of each microbial clade with species-level resolution. The filtered reads were mapped to all reference genome sequences to determine the presence and abundance of different taxonomic groups present in human fecal samples. Given that this study focused on bacterial communities, downstream analysis was performed on bacterial taxa.

Microbial Gene Richness Assessment

Filtered reads were mapped against the integrated gene catalog (IGC) (Li et al., 2014) using the bwa software (Li and Durbin, 2009). Gene richness was measured as the total number of different genes present in the sample regardless of their abundance and length. A minimum of one filtered mapped read was set to consider the presence of a gene, as described previously (Guillen et al., 2019).

Functional Genes and Pathway Analysis

Metagenomics functional analysis was performed using the HMP Unified Metabolic Analysis Network 2 (HUMAnN2),

which identifies the species profile from metagenomic shotgun sequencing data and aligns reads to their pan-genomes, performs translated search on unclassified reads, and quantifies gene families and pathways (Franzosa et al., 2018). By default, gene families were annotated using a comprehensive protein database UniRef90 (Suzek et al., 2015) and metabolic pathways using MetaCyc database (Caspi et al., 2016). The UniRef90 gene family abundance from HUMAnN2 was then regrouped to Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) orthology (KO). To gain a comprehensive profile of functional pathways, the filtered reads were further mapped to the KEGG database, which contains more compounds while having less reactions and pathways than does the MetaCyc database (Altman et al., 2013).

Characterization of Gut Microbial Virulome and Resistome

To characterize the presence of bacterial virulence factors genes and AMR genes, we mapped filtered reads from each sample against the VFDB (Liu et al., 2019) and the CARD (Alcock et al., 2020), respectively, using Bowtie2 (v2.3.5.1). The mapped read segments were estimated using SAMtools idxstats. The copy number of each gene was estimated by dividing the total reads mapping to a gene divided by the gene's length. The relative abundance of genes was calculated using R function `make_relative` in R package `funrar` (v1.4.1).

Rarefaction Analysis of Microbiome, Virulome, and Resistome

Rarefaction curves were generated using R function `rarecurve` (R package `vegan` v2.5.6) to assess the saturation of samples at different sequencing depth for recovery of bacterial species, bacterial virulence factors genes, and AMR genes.

Statistical Analysis

Species richness and Shannon diversity were calculated using R function `estimate_richness`. Beta diversity was measured by Bray–Curtis, weighted UniFrac, and unweighted UniFrac distances using R package `Phyloseq` (v1.30.0) (McMurdie and Holmes, 2013). Samples were clustered according to their species composition using non-metric multidimensional scaling (NMDS) approach based on Bray–Curtis, weighted UniFrac, and unweighted UniFrac distances in `Phyloseq` (v1.30.0).

Due to the small sample size in this study, different methods were used to determine and verify the results obtained from the other method. Differences in abundance of bacterial species, pathways, bacterial virulence genes, and AMR genes between two groups were determined using Wilcoxon rank sum test with a significance level of <0.05 . The linear discriminant analysis (LDA) effect size (LEfSe) algorithm (Segata et al., 2011) was used to identify specific bacterial taxa and metabolic pathways as taxonomic and functional biomarkers and to verify the Wilcoxon rank sum test. Kruskal–Wallis test was used to process the dataset with LEfSe alpha values set at 0.05. The threshold used to consider a discriminative feature for the logarithmic LDA score was set at >2 . The biomarker discovery was performed at bacterial order, family, genus, and species levels and all functional levels. All the correlation analyses were performed using the Spearman's rank correlation coefficient (library "psych"; function "corr.test"). Spearman's correlation (ρ) with $p < 0.05$ was considered statistically significant. Statistical correction for multiple testing was not applied because of the low sample size and the exploratory focus of this study.

RESULTS

Study Subjects

The study subjects were categorized into case (HIV-1-ART individuals, $n = 13$) and control (HIV-1-negative controls, $n = 11$) groups. Cases were classified into groups based on median value (740 cells/mm^3) of current CD4 + T-cell counts: $\geq 740 \text{ cells/mm}^3$ ($n = 7$) and $< 740 \text{ cells/mm}^3$ ($n = 6$).

Influence of Extraction Protocols on Genomic DNA Quantity and Quality

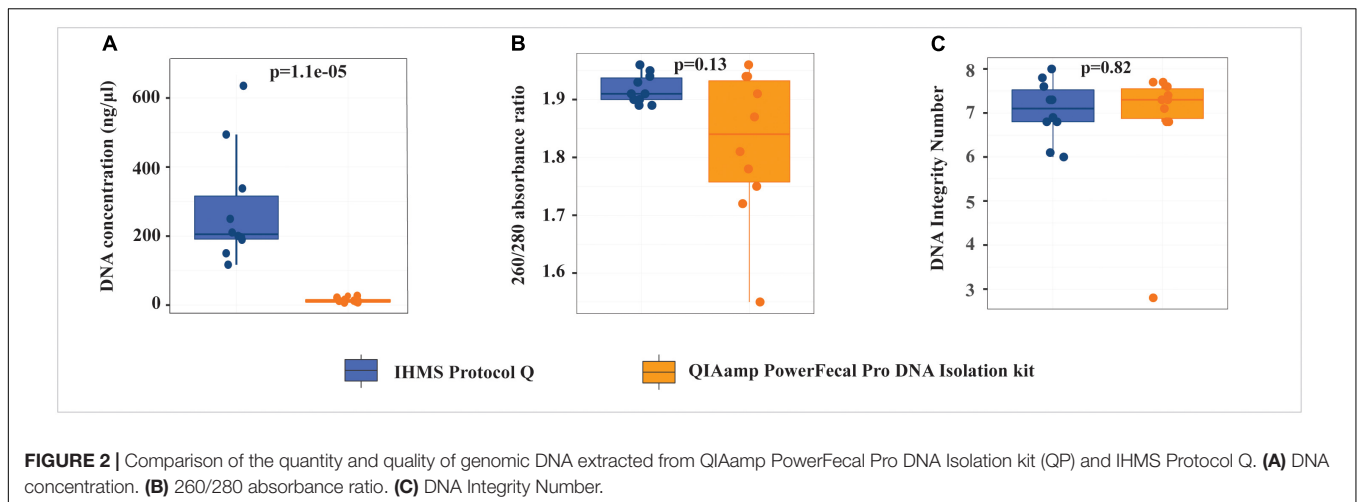
Fecal samples were randomly selected from cases ($n = 3$) and controls ($n = 7$) to evaluate two gDNA extraction protocols. Using the same fecal material mass, IHMS Protocol Q produced significantly higher gDNA yield ($p < 0.0001$, median DNA

concentration 17.26 times higher) than the QP kit (Figure 2 and Supplementary Table S2). The 260/280 absorbance ratio was used to assess the gDNA purity, within an acceptable range of 1.8–2.0 as good-quality DNA. gDNA from the IHMS Protocol Q had 260/280 ratio ranging from 1.89 to 1.96, whereas four out of 10 samples extracted from the QP kit had 260/280 ratio below 1.8. The gDNA integrity was assessed with DNA Integrity Number (DIN) ranging from 1 to 10, where 1 indicates highly degraded gDNA and 10 represents highly intact gDNA (Nguyet et al., 2014). No significant difference in DIN values was found between the two protocols ($p = 0.82$); however, one sample with lower 260/280 ratio of 1.55 from the QP kit had also appreciably low gDIN value of 2.8. Our study indicated that the IHMS Protocol Q yielded significantly higher gDNA quantity and more stable gDNA quality than the QP kit. The remaining 14 samples were therefore extracted by the IHMS Protocol Q. All samples ($n = 24$) extracted by the IHMS Protocol Q were further subjected to shotgun metagenome sequencing and downstream analysis.

Gut Microbial Composition Between HIV-1-ART Cases and HIV-1-Negative Controls

Surprisingly, the average proportion of filtered reads assigned to known bacterial species accounted for 39.9% of total reads, less than 0.1% of reads were assigned to archaea and virus, whereas 60% of reads could not be mapped to any known reference genome of different taxonomic groups by MetaPhlAn 2.0 (Supplementary Figure 1a and Supplementary Table S3). Given that the focus of this study was on bacterial communities, the downstream analysis was performed on bacteria only. Intriguingly, the bacterial species-level richness rapidly reached a plateau for all 24 samples at less than 5 million reads (Supplementary Figure 1b). MetaPhlAn 2.0 identified 288 bacterial species in the 24 fecal samples (Supplementary Table S4), belonging to 109 genera, 53 families, 23 orders, 14 classes, and seven phyla (Figure 3A). *Clostridiales* and *Bacteroidales* were the most abundant orders (Supplementary Figure 2a). *Ruminococcaceae* and *Bifidobacteriaceae* were the most abundant families; *Faecalibacterium* and *Bifidobacterium* were the most abundant genera (Figures 3B,C). The most abundant bacterial species observed in all fecal samples was *Faecalibacterium prausnitzii* (average relative abundance 15.97% in case group vs. 20.16% in control group), followed by *Bifidobacterium adolescentis* (11.17% case vs. 7.11% control), *Collinsella aerofaciens* (5.96% case vs. 6.51% control), and *Ruminococcus bromii* (4.16% case vs. 6.10% control). Sixteen species showed average relative abundance above 1% (Figure 3D). However, no significant difference of these abundant species was found between HIV-1-ART cases and HIV-1-negative controls.

Despite the small sample size, three families, five genera, and 13 species biomarkers were identified between the case and control groups (Figure 3E and Supplementary Figures 2b,c) using LEfSe biomarker discovery tool. Among the 13 species



biomarkers, seven were identified as biomarkers for the case group and six for the control group (**Figure 3E**). Wilcoxon rank sum test was consistent with LDA with two exceptions (*Corynebacterium pseudogenitalium* and *Streptococcus anginosus*) (**Supplementary Table S5A**).

The Altered Microbial Diversity in HIV-1-ART Cases Compared to Negative Controls

A decreased tendency of microbial richness and Shannon diversity at species level was observed in HIV-1-ART cases compared to HIV-1-negative controls, though differences were not statistically significant ($p > 0.05$) (**Figure 4A**). Beta diversity assessed with Bray–Curtis, unweighted UniFrac, and weighted UniFrac dissimilarities was significantly higher in the case group compared to that in the control group ($p = 0.0025$, $p = 0.0031$, $p = 0.047$, respectively) (**Figure 4B**). NMDS with Bray–Curtis distance (**Figure 4C**) and UniFrac distances (data not shown) showed no significant separation between cases and controls, although samples from cases were more diversely distributed as compared to controls. Based on IGC analysis, no statistically significant difference was found in microbial gene richness between cases and controls.

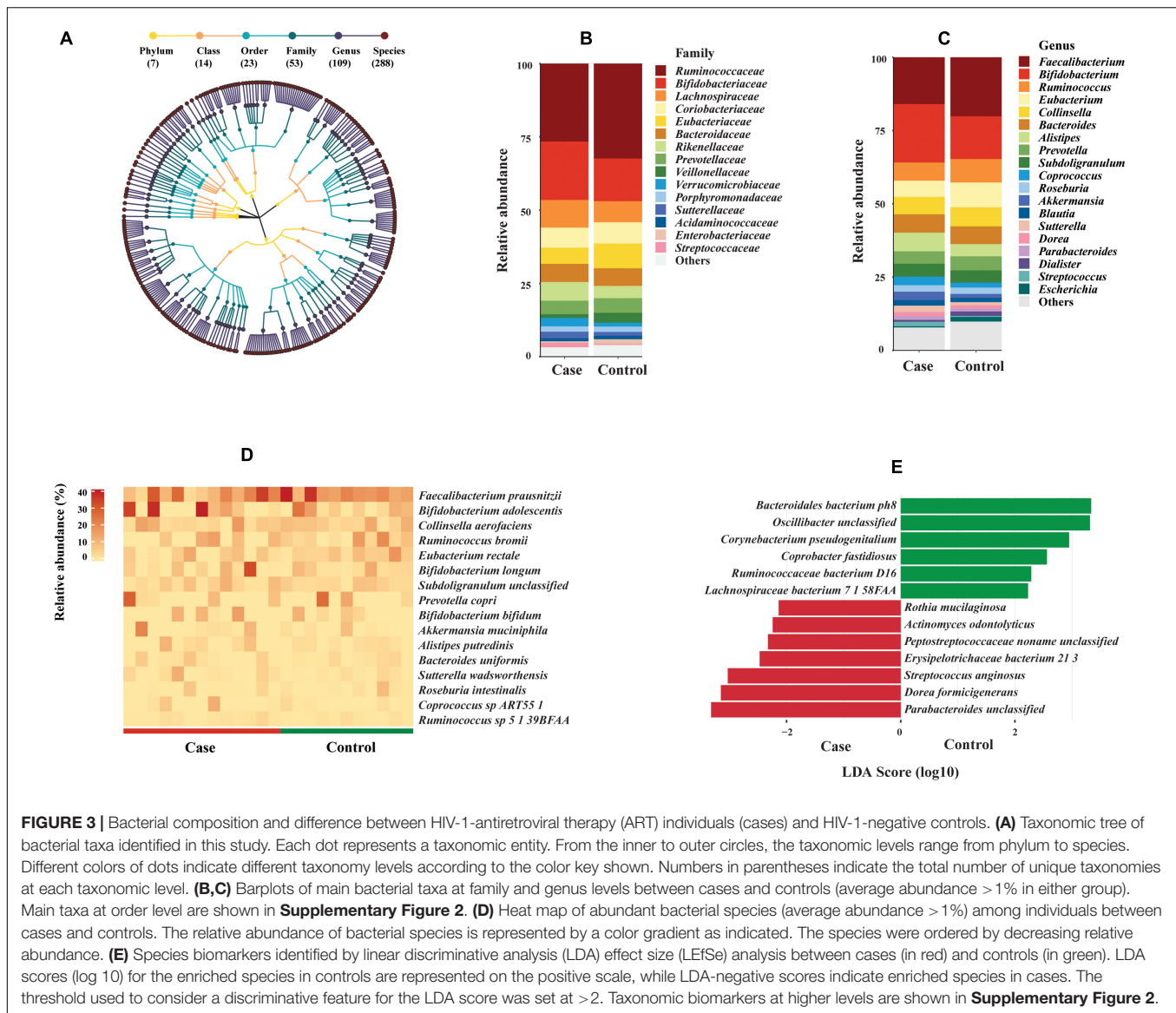
Microbial Functional Genes and Pathways in Cases and Controls

To assess the potential functionality of gut microbiota in HIV-1 infection, microbial genes were annotated and analyzed using different databases and functional systems: KEGG and MetaCyc pathways, KO. We found that 94.6% of filtered reads were unmapped or unintegrated against the KEGG database. The most abundant KEGG pathways identified in this study were biosynthesis of ansamycins (ko01051), valine, leucine and isoleucine biosynthesis (ko00290), ribosome (ko03010), D-Glutamine, and D-glutamate metabolism (ko00471) (**Supplementary Figure 3a**). Similarly, 95.4% of filtered reads were unmapped or unintegrated against

the MetaCyc pathway database; the top MetaCyc pathways were adenosine ribonucleotide *de novo* biosynthesis (PWY-7219) (MetaCyc pathways class: nucleoside and nucleotide biosynthesis), UMP biosynthesis (PWY-5686) (nucleoside and nucleotide biosynthesis), UDP-N-acetylmuramoyl-pentapeptide biosynthesis II (lysine-containing) (PWY-6386) (cell structure biosynthesis), L-isoleucine biosynthesis I (from threonine) (ILEUSYN-PWY) (amino acid biosynthesis), pyruvate fermentation to isobutanol (PWY-7111) (generation of precursor metabolites and energy), and L-valine biosynthesis (VALSYN-PWY) (amino acid biosynthesis) (**Supplementary Figure 3b**). No statistically significant difference of these main pathways was found between the case and control groups. We found that one MetaCyc pathway, Bifidobacterium shunt (P124-PWY), also known as fructose-6-phosphate pathway, was statistically significantly enriched in the case group compared to the control group (**Supplementary Table S5B**). Additionally, the microbial communities in the case group were significantly enriched in genes encoding for enzymes such as xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (K01621), which is key enzyme in fructose-6-phosphate pathway, as well as pyridoxamine 5'-phosphate oxidase (K00275), and dihydropterolate synthase (K00796) (**Supplementary Table S5C**).

Association Between Gut Microbiome and CD4 + T-Cell Counts

The relative abundance of bacterial species, functional genes, and pathways in relation to the CD4 + T-cell counts were analyzed by LDA, Wilcoxon rank sum test, as well as Spearman's correlation analysis. LDA identified five species biomarkers (shown in **Supplementary Figure 4** including *Lactococcus lactis*) for the lower CD4 + T-cell counts (LC) group and one (*Alistipes putredinis*) for the higher CD4 + T-cell counts (HC) group. Wilcoxon rank sum test was consistent with LDA with two exceptions (**Supplementary Table S6A**). Consistently, *A. putredinis* was positively correlated to both nadir CD4 + T-cell count ($\rho = 0.603$, $p = 0.0290$) and current CD4 + T-cell count ($\rho = 0.779$, $p = 0.0017$), while *Lactococcus*



lactis was inversely correlated to nadir CD4 + T-cell count ($\rho = -0.733$, $p = 0.0044$) and current CD4 + T-cell count ($\rho = -0.717$, $p = 0.0058$) using Spearman's correlation analysis (**Figure 5A**). The total ART duration was positively correlated to *Ruminococcus lactaris* ($\rho = 0.580$, $p = 0.0375$) while inversely correlated to *Streptococcus sanguinis* ($\rho = -0.638$, $p = 0.0191$), *Streptococcus gordonii* ($\rho = -0.554$, $p = 0.0497$), *Megamonas unclassified* ($\rho = -0.620$, $p = 0.0237$), and *Aggregatibacter unclassified* ($\rho = -0.559$, $p = 0.0470$) (**Figure 5A**).

LDA identified one MetaCyc pathway biomarker, i.e., phosphopantothenate biosynthesis I (PANTO-PWY) among patients in the lower current CD4 + T-cell counts (LC) group compared to those with higher current CD4 + T-cell counts (HC) group (data not shown). Using Wilcoxon rank sum test, 22 additional MetaCyc pathways and one KEGG pathway chlorocyclohexane and chlorobenzene degradation (ko00361) were found to be significantly enriched in the LC group

(**Supplementary Tables S6B,C**). At the gene family level, 57 KOs with characterized function were found to be significantly different between LC and HC groups (**Supplementary Table S6D**). The Spearman's correlation analysis showed that three MetaCyc pathways were inversely correlated to both nadir and current CD4 + T-cell counts (**Figure 5B**). Of these, phosphopantothenate biosynthesis I (PANTO-PWY) showed strongly inverse correlations with both nadir CD4 + T-cell count ($\rho = -0.784$, $p = 0.0015$) and current CD4 + T-cell count ($\rho = -0.780$, $p = 0.0017$). Consistently, phosphopantothenate biosynthesis I (PANTO-PWY) and glycerol degradation to butanol (PWY-7003) pathways were significantly enriched in the LC group ($p = 0.0012$ and $p = 0.0082$, respectively) by Wilcoxon rank sum test (**Supplementary Table S6B**). These data suggested that MetaCyc pathways related to phosphopantothenate biosynthesis I and glycerol degradation to butanol were associated with low CD4 + T-cell

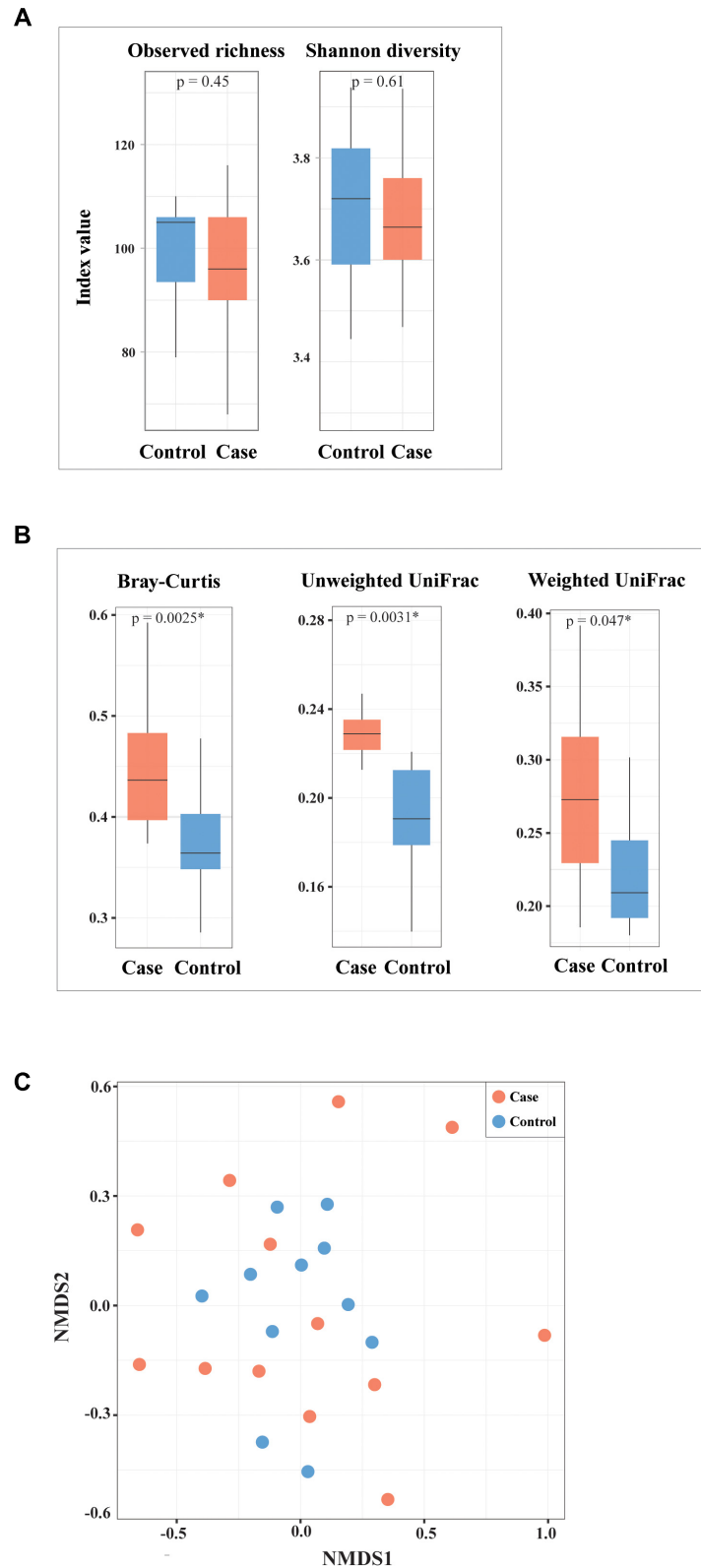
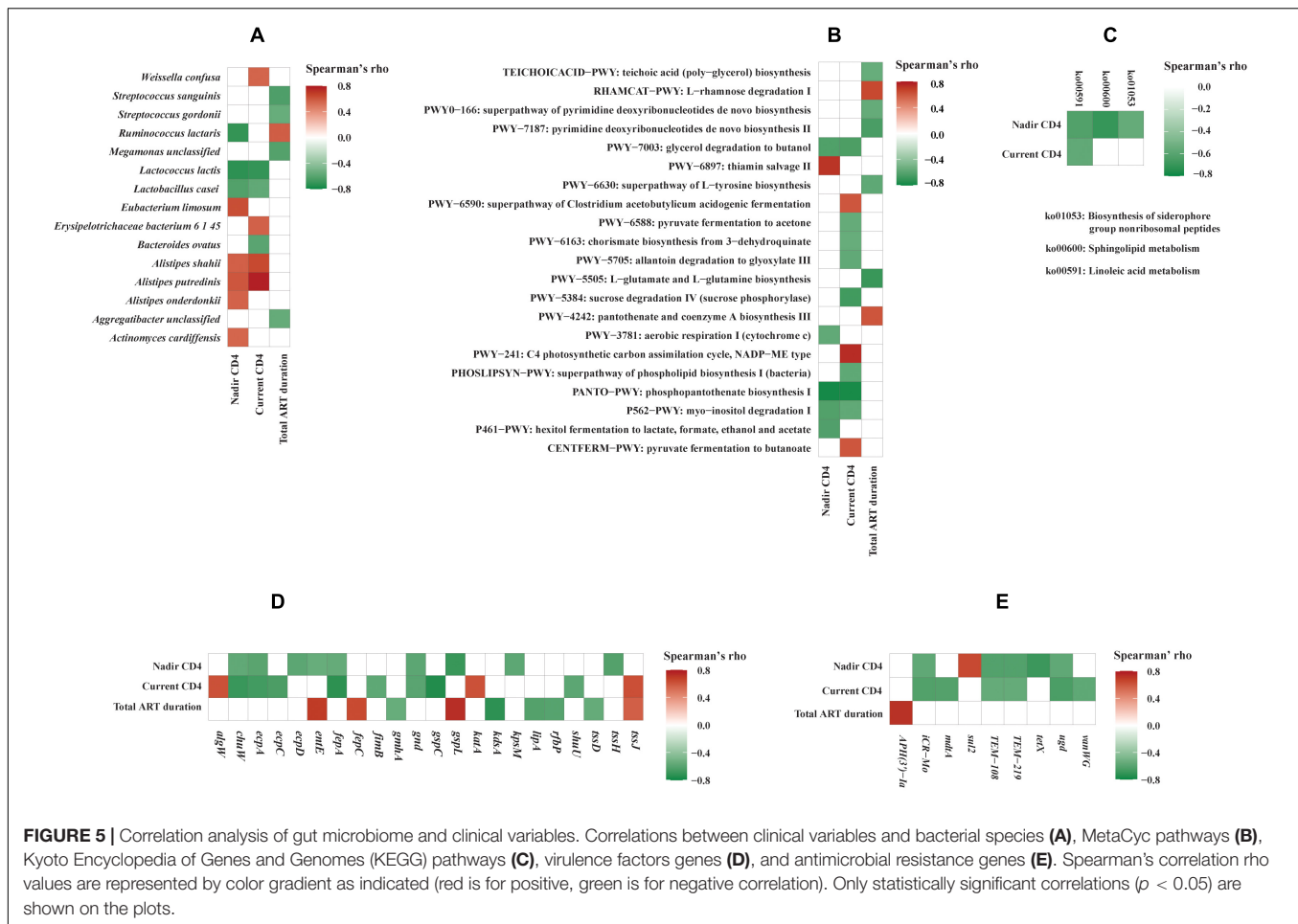


FIGURE 4 | Alpha and beta diversity of bacterial species between cases and controls. **(A)** Alpha diversity assessed by observed species richness and Shannon diversity. **(B,C)** Beta diversity assessed with Bray–Curtis, weighted UniFrac, unweighted UniFrac dissimilarities, as well as non-metric multidimensional scaling (NMDS) based on Bray–Curtis distance. *Statistically significant.



count. We also observed correlations between three KEGG pathways and CD4 + T-cell counts (Figure 5C). For instance, linoleic acid metabolism pathway (ko00591) was inversely correlated to both nadir CD4 + T-cell count ($\rho = -0.630$, $p = 0.0211$) and current CD4 + T-cell count ($\rho = -0.577$, $p = 0.0391$), which was consistent with an earlier study (Vesterbacka et al., 2017).

Difference of Bacterial Virulence Factors and Antimicrobial Resistance Genes Between Cases and Controls

We identified multiple bacterial virulence factors genes and AMR genes in the full dataset (Supplementary Tables S7, S8). It is noteworthy that more than 200 virulence factor genes were detected exclusively in HIV-1-ART cases, though their relative abundance was low. These genes encode critical bacterial virulence factors such as adhesion, toxin, hemolysin, type III secretion system, etc. (Supplementary Table S7). We found that 173 AMR genes were present in cases only; in addition, cases were significantly enriched in genes associated with tetracycline antibiotic resistance and antibiotic efflux pumps ($p < 0.05$) (Supplementary Table S8). Rarefaction analysis

showed that sequencing depth critically affected the recovery of bacterial virulence factors genes and AMR genes, which was dramatically different from taxonomic profiling. A sequencing depth of 57 million reads per sample was insufficient to capture full spectrums of bacterial virulence factors genes and AMR genes in human fecal samples in this study (Supplementary Figures 5a,b).

We observed correlations between virulence factors genes, AMR genes, and clinical variables (Figures 5D,E). Both current and nadir CD4 + T-cell counts were correlated inversely to virulence genes encoding 6-phosphogluconate dehydrogenase (*gnd*), *E. coli* common pilus structural subunit EcpA (*ecpA*), ferrienterobactin outer membrane transporter (*fepA*), and oxygen-independent coproporphyrinogen III oxidase (*chuW*) (Figure 5D). While nadir CD4 + T-cell count was inversely correlated to several important virulence genes such as type VI secretion system ATPase TssH (*tssh*) ($\rho = -0.625$, $p = 0.0224$), fimbria adhesin EcpD (*ecpD*) ($\rho = -0.596$, $p = 0.0317$), type II secretion system protein L (*gspL*) ($\rho = -0.702$, $p = 0.0074$), polysialic acid transport protein KpsM (*kpsM*) ($\rho = -0.588$, $p = 0.0346$), and enterobactin synthase component E (*entE*) ($\rho = -0.556$, $p = 0.0483$). Additionally, current CD4 + T-cell count was inversely

correlated to genes encoding outer membrane usher protein EcpC (*ecpC*) ($\rho = -0.652$, $p = 0.0157$), type II secretion system protein C (*gspC*) ($\rho = -0.765$, $p = 0.0023$), permease of iron compound ABC transport system (*shuU*) ($\rho = -0.592$, $p = 0.0330$), and type I fimbriae regulatory protein FimB (*fimB*) ($\rho = -0.584$, $p = 0.0362$). Our study also indicated that HIV-1 infection is correlated to changes in AMR genes. Both current and nadir CD4 + T-cell counts were inversely correlated to genes associated with beta-lactam antibiotic resistance (*TEM-108* and *TEM-219*) and peptide antibiotic resistance (*icr-Mo* and *ugd*) (Figure 5E). While nadir CD4 + T-cell count was inversely correlated to tetracycline antibiotic resistance gene *tetX* ($\rho = -0.694$, $p = 0.0085$), current CD4 + T-cell count was inversely correlated to *mdtA* ($\rho = -0.624$, $p = 0.0226$), which encodes for resistance-nodulation-cell division (RND) antibiotic efflux pump, and glycopeptide resistance gene *vanWG* ($\rho = -0.584$, $p = 0.0362$). Total ART duration was strongly positively correlated to aminoglycoside antibiotic resistance gene *APH(3')-Ia* ($\rho = 0.725$, $p = 0.0050$).

DISCUSSION

In this study, we set up a standard shotgun metagenomics workflow through methodologies assessment including sample preparation in order to get a precise gut microbiome profile. We did a literature review and selected and evaluated two fecal DNA extraction protocols that performed better than other methods used in gut microbiome studies (Angebault et al., 2018; Lim et al., 2018; Fiedorova et al., 2019). We found the IHMS Protocol Q yielded significantly higher gDNA quantity and more stable gDNA quality than the commercial kit QP. It is possible that different gDNA quantity and quality may have an impact on the microbial structures and downstream analysis; further studies are warranted to investigate this. Notably, we found that only ~40% of the total reads were assigned to known microbial taxa, whereas 60% of the reads could not be mapped to any known taxonomic group by MetaPhlan 2.0. This was in line with previous studies estimating that approximately 40–60% of human gut microbes cannot be captured by current genome-based methods, despite the considerable efforts that have been made to culture and sequence members of the gut microbiome (Sunagawa et al., 2013; Nayfach et al., 2019). We assessed the shotgun metagenome sequencing depth required for characterizing human gut microbiome at microbial species and gene level. We found that less than 5 million reads per sample were sufficient to achieve almost a full bacterial composition, whereas an increasing number of microbial genes, including virulence factors genes and AMR genes, were still being recovered at a sequencing depth of 57 million reads. This finding is consistent with a recent metagenomics study performed on animal and environmental samples, indicating that taxonomic profiling is much more stable to sequencing depth than AMR gene content (Gweon et al., 2019). To the best of our knowledge, this is the first study depicting the shotgun metagenome

sequencing depth requirement for simultaneously characterizing gut microbiome, virulome, and resistome in human fecal samples. Our findings will facilitate the selection of appropriate sequencing depth for human gut microbiome studies to obtain reliable results.

Accumulating studies, including ours, have shown that HIV-1 infection is associated with gut dysbiosis, reduced alpha diversity, and increased beta diversity of microbiota (Nowak et al., 2015; Kang and Cai, 2019; Flygel et al., 2020), which may persist after ART assumption (Vujkovic-Cvijin et al., 2013; Vazquez-Castellanos et al., 2015; Ji et al., 2018). Despite the study limits, e.g., small sample size and lack of treatment-naive HIV-1 patients, we observed similar microbiota alteration in HIV-1-ART individuals compared to negative controls. It should be noted that previous studies were mostly based on 16S rRNA gene sequencing, which can profile microbiota at the genus level. Our data showed similar changes of microbial diversity in HIV-1-infected individuals at the species level. We found that several bacterial species were significantly enriched in HIV-1-ART individuals. Three of these species, at their higher taxonomic level (i.e., *Peptostreptococcus*, *Erysipelotrichaceae*, and *Dorea*), have earlier been reported to be enriched in HIV-1-infected subjects (McHardy et al., 2013; Vujkovic-Cvijin et al., 2013; Dinh et al., 2015; Ling et al., 2016), indicating the potential role of these bacteria in HIV-1 pathogenesis.

To date, only a few studies have investigated functional genes and pathways in microbial communities and their roles in HIV-1 infection using whole metagenome sequencing (Vazquez-Castellanos et al., 2015; Lu et al., 2018; Guillen et al., 2019; Rocafort et al., 2019). None of these studies has assessed the proportion of functionally characterized genes in the samples. Remarkably, we found that more than 94% of filtered reads were unmapped or unintegrated against two established pathway databases, which was in line with a previous study using similar methods (Abu-Ali et al., 2018). These data suggest that most microbial genes in human feces remain functionally unknown. We found that “Bifidobacterium shunt,” also known as fructose-6-phosphate pathway, was significantly enriched in the case group. Consistently, gene encoding for the key enzyme in this pathway, i.e., xylulose-5-phosphate/fructose-6-phosphate phosphoketolase, which plays a key role in carbohydrate metabolism in a number of bacteria, was also enriched in the case group. This may indicate the potential role of the fructose-6-phosphate pathway in HIV-1 infection. Further larger studies with matched groups are essential to confirm this and unravel the potential underlying mechanisms.

We observed correlations between CD4 + T-cell count and gut microbiome in HIV-1-ART patients. For instance, both current and nadir CD4 + T-cell counts were inversely correlated to linoleic acid metabolism pathway, which is corroborated by previous findings that the fatty acid metabolism has an important role in the regulation of immune responses and immunological diseases (Yaqoob, 2003; Hosomi et al., 2020). Additionally, both current and nadir CD4 + T-cell counts were inversely correlated to pathways related to phosphopantothenate biosynthesis I, myo-inositol degradation I, glycerol degradation to butanol, implying that these pathways might play a role in

the mechanism of immune cell regulation in HIV-1 infection. A previous study indicated that the microbial shift associated with immune deficiency in HIV-1 infection was related to increased bacterial virulence factors and AMR genes in the gut microbiome (Guillen et al., 2019). Consistently, we found important bacterial virulence factors such as adhesion, toxin, hemolysin, type III secretion system encoding genes that were exclusively present or enriched in the HIV-1-ART patients. AMR genes associated with tetracycline antibiotic resistance were significantly enriched in the case group. Several virulence factors genes and AMR genes were found to be correlated to CD4 + T-cell count and total ART duration. It is notable that the total ART duration was strongly positively correlated to abundance of aminoglycoside antibiotic resistance gene *APH(3')-Ia*. We did not have access to earlier antibiotic treatment in HIV-1-ART individuals, but it is not unlikely that a more frequent use of antibiotics preceding the study in these patients could have contributed to these findings. Further studies are needed to validate the associations between HIV-1 infection, degree of immunodeficiency, and gut microbiome observed in this study.

We acknowledge limitations in this study. First, it is an exploratory microbiome study using deep shotgun metagenome sequencing on a small sample size, as its purpose was mainly to set up a standard whole metagenomics pipeline, thereby characterizing the gut microbiome profile associated with HIV-1 infection at the bacterial species and functional levels. Second, previous studies indicated that antiretroviral drugs themselves may impact the gut microbiota (Nowak et al., 2015; Ray et al., 2020). Since all HIV-1-infected participants in this study were on ART at the time of sampling, we could not differentiate the effect of ART and HIV-1 infection itself, respectively, on gut microbiome dysbiosis. Third, the associations between gut microbiome alteration and HIV-1 infection might be confounded by factors such as gender, age, and sexual preference, which were not evaluated due to the small sample size. It should be noted that the gender difference in cases and controls might also be a confounder. In addition, it is difficult to identify microbial differences in composition and functions associated with disease status, as such differences can be partly due to genetic diversity among the human samples (Goodrich et al., 2014; Tierney et al., 2019). Furthermore, statistical correction for multiple testing was not applied because of the low sample size and the exploratory focus of the study. Therefore, further longitudinal studies on a larger scale of treatment-naïve patients before and after ART, with matched groups by other confounders, are essential to validate our findings.

In conclusion, we set up whole metagenomics workflows and characterized gut microbiome in HIV-1-infected individuals on ART. The results indicate that gut dysbiosis in HIV-1 infection is potentially associated with changes in bacterial composition, functional genes and pathways, as well as gut microbial virulome and resistome. Further studies are essential to validate these findings and to evaluate potentially unmeasured confounders that might impact the gut microbiome. The methodological assessment shows that different DNA extraction protocols significantly affect the gDNA quantity and quality, and

metagenome sequencing depth critically affects the recovery of microbial genes including bacterial virulence factors genes and AMR genes in human metagenomic samples. These findings will support the design of future human gut microbiome studies, particularly when virulome and resistome are of study interest.

DATA AVAILABILITY STATEMENT

The raw shotgun metagenome sequencing data and metadata from this study have been submitted to GenBank Sequence Read Archive under the BioProject ID PRJNA692830. All the codes are available at github: <https://github.com/Asw614/Whole-genome-metagenomic-analysis-of-the-gut-microbiome>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Regional Ethics Committee, Stockholm (2009/1485-31, 2013/1944-31/4, and 2014/920-3). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AS and UN initiated the study and supervised the sample and data collection. XB performed the laboratory work, analyzed the data, and wrote the manuscript. AN performed the bioinformatics and statistical analyses with the supervision of UN and XB and contributed to the manuscript preparation. PN and SR contributed to the sample collection, data interpretation, and manuscript preparation. All authors reviewed, edited, and approved the final manuscript.

FUNDING

This study was supported by the Swedish Research Council (AS: 2020-02129, 2017-05848, and 2016-01675) and ALF-Stockholm County Council no. 2019 (AS). It is also based upon work from CARE Consortium, that has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 825673 and the Ministry of Science and Higher Education of the Russian Federation, federal target program 2014-2020, unique project identifier RFMEFI61019X0020. UN is supported by Swedish Research Council grants (2017-01330 and 2018-06156).

ACKNOWLEDGMENTS

We acknowledge support from the National Genomics Infrastructure in Genomics Production Stockholm funded by the Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced

Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We thank the core facility at Novum, BEA, Bioinformatics and Expression Analysis, which is supported by the board of research at the Karolinska Institute and the research committee at the Karolinska hospital. The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at [SNIC CENTRE] partially funded by the Swedish Research Council through grant agreement no. 2018-05973. We thank Lokeshwaran Manoharan from National Bioinformatics Infrastructure Sweden (NBIS),

who provided the bioinformatics support during data analysis. We also would like to thank Robert van Domselaar from the Department of Medicine Huddinge, Karolinska Institutet, for laboratory assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.667718/full#supplementary-material>

REFERENCES

- Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., et al. (2018). Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* 3, 356–366. doi: 10.1038/s41564-017-0084-4
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935
- Altman, T., Travers, M., Kothari, A., Caspi, R., and Karp, P. D. (2013). A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics* 14:112. doi: 10.1186/1471-2105-14-112
- Ananthkrishnan, A. N., Bernstein, C. N., Iliopoulos, D., Macpherson, A., Neurath, M. F., Ali, R. A. R., et al. (2018). Environmental triggers in IBD: a review of progress and evidence. *Nat. Rev. Gastroenterol. Hepatol.* 15, 39–49. doi: 10.1038/nrgastro.2017.136
- Angebault, C., Ghazlane, A., Volant, S., Botterel, F., d'Enfert, C., and Bougnoux, M. E. (2018). Combined bacterial and fungal intestinal microbiota analyses: impact of storage conditions and DNA extraction protocols. *PLoS One* 13:e0201174. doi: 10.1371/journal.pone.0201174
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480. doi: 10.1093/nar/gkv1164
- Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., et al. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* 35, 1069–1076. doi: 10.1038/nbt.3960
- Dillon, S. M., Frank, D. N., and Wilson, C. C. (2016). The gut microbiome and HIV-1 pathogenesis: a two-way street. *AIDS* 30, 2737–2751. doi: 10.1097/QAD.0000000000001289
- Dillon, S. M., Kibbie, J., Lee, E. J., Guo, K., Santiago, M. L., Austin, G. L., et al. (2017). Low abundance of colonic butyrate-producing bacteria in HIV infection is associated with microbial translocation and immune activation. *AIDS* 31, 511–521. doi: 10.1097/QAD.0000000000001366
- Dinh, D. M., Volpe, G. E., Duffalo, C., Bhalchandra, S., Tai, A. K., Kane, A. V., et al. (2015). Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *J. Infect. Dis.* 211, 19–27. doi: 10.1093/infdis/jiu409
- D'Souza, A. W., Moodley-Govender, E., Berla, B., Kelkar, T., Wang, B., Sun, X., et al. (2019). Cotrimoxazole prophylaxis increases resistance gene prevalence and alpha-diversity but decreases beta-diversity in the gut microbiome of HIV-exposed, uninfected infants. *Clin. Infect. Dis.* 71, 2858–2868. doi: 10.1093/cid/ciz1186
- Fiedorova, K., Radvansky, M., Nemcova, E., Grombirikova, H., Bosak, J., Cernochova, M., et al. (2019). The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery. *Front. Microbiol.* 10:821. doi: 10.3389/fmicb.2019.00821
- Flygel, T. T., Sovershaeva, E., Claassen-Weitz, S., Hjerde, E., Mwaikono, K. S., Odland, J. O., et al. (2020). Composition of gut microbiota of children and adolescents with perinatal human immunodeficiency virus infection taking antiretroviral therapy in Zimbabwe. *J. Infect. Dis.* 221, 483–492. doi: 10.1093/infdis/jiz473
- Franzosa, E. A., McIver, L. J., Rahnvard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. doi: 10.1038/s41592-018-0176-y
- Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., et al. (2014). Human genetics shape the gut microbiome. *Cell* 159, 789–799. doi: 10.1016/j.cell.2014.09.053
- Guillen, Y., Noguera-Julian, M., Rivera, J., Casadella, M., Zevin, A. S., Rocafort, M., et al. (2019). Low nadir CD4+ T-cell counts predict gut dysbiosis in HIV-1 infection. *Mucosal Immunol.* 12, 232–246. doi: 10.1038/s41385-018-0083-7
- Gweon, H. S., Shaw, L. P., Swann, J., De Maio, N., AbuOun, M., Niehus, R., et al. (2019). The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environ. Microb.* 14:7. doi: 10.1186/s40793-019-0347-1
- Handley, S. A., Thackray, L. B., Zhao, G., Presti, R., Miller, A. D., Droit, L., et al. (2012). Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* 151, 253–266. doi: 10.1016/j.cell.2012.09.024
- Hosomi, K., Kiyono, H., and Kunisawa, J. (2020). Fatty acid metabolism in the host and commensal bacteria for the control of intestinal immune responses and diseases. *Gut Microbes* 11, 276–284. doi: 10.1080/19490976.2019.1612662
- Hunt, P. W., Sinclair, E., Rodriguez, B., Shive, C., Clagett, B., Funderburg, N., et al. (2014). Gut epithelial barrier dysfunction and innate immune activation predict mortality in treated HIV infection. *J. Infect. Dis.* 210, 1228–1238. doi: 10.1093/infdis/jiu238
- Ji, Y., Zhang, F., Zhang, R., Shen, Y., Liu, L., Wang, J., et al. (2018). Changes in intestinal microbiota in HIV-1-infected subjects following cART initiation: influence of CD4+ T cell count. *Emerg. Microbes Infect.* 7:113. doi: 10.1038/s41426-018-0117-y
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kang, Y., and Cai, Y. (2019). Altered gut microbiota in HIV infection: future perspective of fecal microbiota transplantation therapy. *AIDS Res. Hum. Retroviruses* 35, 229–235. doi: 10.1089/AID.2017.0268
- Karlsson, F., Tremaroli, V., Nielsen, J., and Backhed, F. (2013). Assessing the human gut microbiota in metabolic diseases. *Diabetes* 62, 3341–3349. doi: 10.2337/db13-0844
- Kelley, C. F., Kraft, C. S., de Man, T. J., Duphare, C., Lee, H. W., Yang, J., et al. (2017). The rectal mucosa and condomless receptive anal intercourse in HIV-negative MSM: implications for HIV transmission and prevention. *Mucosal Immunol.* 10, 996–1007. doi: 10.1038/mi.2016.97
- Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018). Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422. doi: 10.1038/s41579-018-0029-9
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

- Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* 32, 834–841. doi: 10.1038/nbt.2942
- Lim, M. Y., Song, E. J., Kim, S. H., Lee, J., and Nam, Y. D. (2018). Comparison of DNA extraction methods for human gut microbial community profiling. *Syst. Appl. Microbiol.* 41, 151–157. doi: 10.1016/j.syapm.2017.11.008
- Ling, Z., Jin, C., Xie, T., Cheng, Y., Li, L., and Wu, N. (2016). Alterations in the fecal microbiota of patients with HIV-1 infection: an observational study in a Chinese population. *Sci. Rep.* 6:30673. doi: 10.1038/srep30673
- Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080
- Lu, W., Feng, Y., Jing, F., Han, Y., Lyu, N., Liu, F., et al. (2018). Association between gut microbiota and CD4 recovery in HIV-1 infected patients. *Front. Microbiol.* 9:1451. doi: 10.3389/fmicb.2018.01451
- Mathee, K., Cickovski, T., Deoraj, A., Stollstorff, M., and Narasimhan, G. (2020). The gut microbiome and neuropsychiatric disorders: implications for attention deficit hyperactivity disorder (ADHD). *J. Med. Microbiol.* 69, 14–24. doi: 10.1099/jmm.0.001112
- McHardy, I. H., Li, X., Tong, M., Ruegger, P., Jacobs, J., Borneman, J., et al. (2013). HIV Infection is associated with compositional and functional shifts in the rectal mucosal microbiota. *Microbiome* 1:26. doi: 10.1186/2049-2618-1-26
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., and Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature* 568, 505–510. doi: 10.1038/s41586-019-1058-x
- Nguyet, K., Whitney, N., Lucy, C., Alvin, L., and Bart, C. (2014). *Integrating the DNA Integrity Number (DIN) to Assess Genomic DNA (gDNA) Quality Control Using the Agilent 2200 TapeStation System*. Santa Clara, CA: Agilent Technologies, 5991EN–5442EN.
- Noguera-Julian, M., Rocaforat, M., Guillen, Y., Rivera, J., Casadella, M., Nowak, P., et al. (2016). Gut microbiota linked to sexual preference and HIV infection. *EBio Med.* 5, 135–146. doi: 10.1016/j.ebiom.2016.01.032
- Nowak, P., Troseid, M., Avershina, E., Barqasho, B., Neogi, U., Holm, K., et al. (2015). Gut microbiota diversity predicts immune status in HIV-1 infection. *AIDS* 29, 2409–2418. doi: 10.1097/QAD.0000000000000869
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Ray, S., Narayanan, A., Giske, C. G., Neogi, U., Sonnerborg, A., and Nowak, P. (2020). Altered Gut microbiome under antiretroviral therapy: impact of efavirenz and zidovudine. *ACS Infect. Dis.* 7, 1104–1115. doi: 10.1021/acsinfectdis.0c00536
- Rocaforat, M., Noguera-Julian, M., Rivera, J., Pastor, L., Guillen, Y., Langhorst, J., et al. (2019). Evolution of the gut microbiome following acute HIV-1 infection. *Microbiome* 7:73. doi: 10.1186/s40168-019-0687-5
- Schwabe, R. F., and Jobin, C. (2013). The microbiome and cancer. *Nat. Rev. Cancer* 13, 800–812. doi: 10.1038/nrc3610
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60
- Severance, E. G., Yolken, R. H., and Eaton, W. W. (2016). Autoimmune diseases, gastrointestinal disorders and the microbiome in schizophrenia: more than a gut feeling. *Schizophr. Res.* 176, 23–35. doi: 10.1016/j.schres.2014.06.027
- Sonnerborg, J. L., and Backhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. *Nature* 535, 56–64. doi: 10.1038/nature18846
- Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., et al. (2013). Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* 10, 1196–1199. doi: 10.1038/nmeth.2693
- Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H., and UniProt, C. (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31, 926–932. doi: 10.1093/bioinformatics/btu739
- Svard, J., Paquin-Proulx, D., Buggert, M., Noyan, K., Barqasho, B., Sonnerborg, A., et al. (2015). Role of translocated bacterial flagellin in monocyte activation among individuals with chronic HIV-1 infection. *Clin. Immunol.* 161, 180–189. doi: 10.1016/j.clim.2015.08.018
- Tierney, B. T., Yang, Z., Luber, J. M., Beaudin, M., Wibowo, M. C., Baek, C., et al. (2019). The landscape of genetic content in the gut and oral human microbiome. *Cell Host Microbe* 26, 283–295.e8. doi: 10.1016/j.chom.2019.07.008
- Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903. doi: 10.1038/nmeth.3589
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031. doi: 10.1038/nature05414
- Vazquez-Castellanos, J. F., Serrano-Villar, S., Jimenez-Hernandez, N., Soto Del Rio, M. D., Gayo, S., Rojo, D., et al. (2018). Interplay between gut microbiota metabolism and inflammation in HIV infection. *ISME J.* 12, 1964–1976. doi: 10.1038/s41396-018-0151-8
- Vazquez-Castellanos, J. F., Serrano-Villar, S., Latorre, A., Artacho, A., Ferrus, M. L., Madrid, N., et al. (2015). Altered metabolism of gut microbiota contributes to chronic immune activation in HIV-infected individuals. *Mucosal Immunol.* 8, 760–772. doi: 10.1038/mi.2014.107
- Vesterbacka, J., Rivera, J., Noyan, K., Parera, M., Neogi, U., Calle, M., et al. (2017). Richer gut microbiota with distinct metabolic profile in HIV infected elite controllers. *Sci. Rep.* 7:6269. doi: 10.1038/s41598-017-06675-1
- Vujkovic-Cvijin, I., Dunham, R. M., Iwai, S., Maher, M. C., Albright, R. G., Broadhurst, M. J., et al. (2013). Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Sci. Transl. Med.* 5:193ra91. doi: 10.1126/scitranslmed.3006438
- Yaqoob, P. (2003). Fatty acids as gatekeepers of immune cell regulation. *Trends Immunol.* 24, 639–645. doi: 10.1016/j.it.2003.10.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Bai, Narayanan, Nowak, Ray, Neogi and Soñnerborg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.