Low eukaryotic viral richness is associated with faecal microbiota transplantation success in ulcerative colitis patients

Nádia **Conceição-Neto**[1,2], Ward **Deboutte**[1], Tim **Dierckx**[2], Kathleen **Machiels**[3], Jun **Wang**[4,5], Kwe Claude **Yinda**[1,2], Piet **Maes**[2], Marc **Van Ranst**[2], Marie **Joossens**[4,5], Jeroen **Raes**[4,5], Séverine **Vermeire**[3], Jelle **Matthijnssens**[1#]

Supplementary information

Wet-lab procedure description

Faecal suspensions (15% m/v) were homogenised using a tissue homogeniser (MINILYS, Bertin Technologies) for 1 min at 3000rpm. This homogenisation step was then followed by a brief centrifugation at 17000 g for 3 min and filtered through a 0.8 µm filter (PES filter, Sartorious). The filtrate was then treated in a homemade buffer (1 M Tris, 100 mM CaCl2, and 30 mM MgCl2) together with Benzonase (Novagen) and Micrococcal Nuclease (New England Biolabs) enzymes for 2 h at 37°C to allow digestion of free-floating nucleic acids. Both RNA and DNA were extracted using the QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer's instructions but without addition of carrier RNA to the lysis buffer. First and second strand synthesis and random PCR amplification for 17 cycles were performed using a slightly modified Whole Transcriptome Amplification 2 (WTA2) Kit procedure (Sigma-Aldrich), with a denaturation temperature of 95°C instead of 72°C to allow for the denaturation of dsDNA and dsRNA. Amplification products were purified and were prepared for Illumina sequencing using the Nextera XT DNA library preparation kit (Illumina), with 4-min tagmentation time and 45-s PCR extension. Libraries were quantified with the KAPA Library Quantification kit (Kapa Biosystems), and sequencing of the samples was performed on a NextSeq500 platform (Illumina) for 300 cycles (150-bp paired ends).

Bioinformatics pipeline

Raw reads were trimmed for quality and adapters using Trimmomatic. After quality control, a cross assembly using all reads was done using metaSPAdes.

Phageome analysis

For the phageome pipeline scaffolds were ran through VirSorter on decontamination mode and scaffolds assigned to categories 1 and 2 were extracted for further analysis. These phage contigs were filtered for redundancy at 95% nucleotide identity over 70% of the length using Cluster Genomes.

Individual sample corrected reads were aligned against the remaining phage contigs using Bowtie 2 and generated BAM files were filtered to remove reads that aligned for <95% identity using BamM. Coverage per genome position was obtained using Bedtools, and a cutoff of 70% coverage per scaffold was used to establish the presence of a viral OTU in each sample. Abundance tables were obtained and normalised for total number of reads excluding human genome contamination.

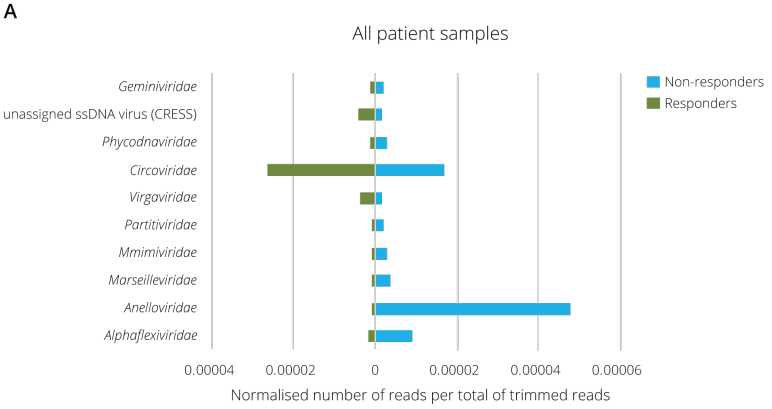
Eukaryotic virome analysis

For the eukaryotic virome, scaffolds were annotated using DIAMOND (blastx parameter) against the non-redundant (nr) database, on sensitive mode. Individual sample magnitudes were obtained by aligning corrected reads to the obtained eukaryotic virus scaffolds using BWA-MEM. Abundance tables for viral familes were obtained and normalised for total number of reads excluding human genome contamination.

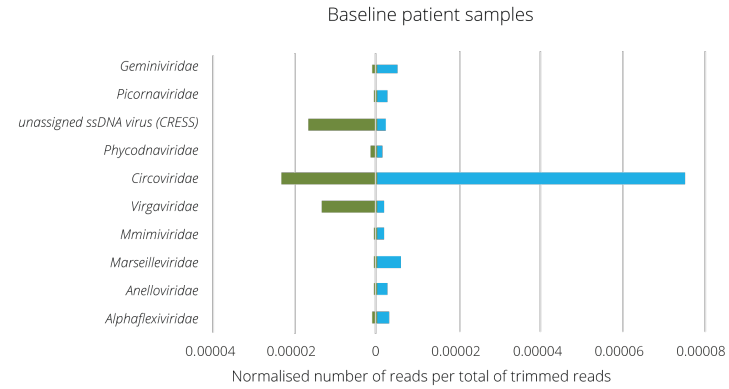
Ecological metrics analysis

All ecological metrics were calculated in R using the vegan package and comparisons between groups were doing using Mann-Whitney tests.

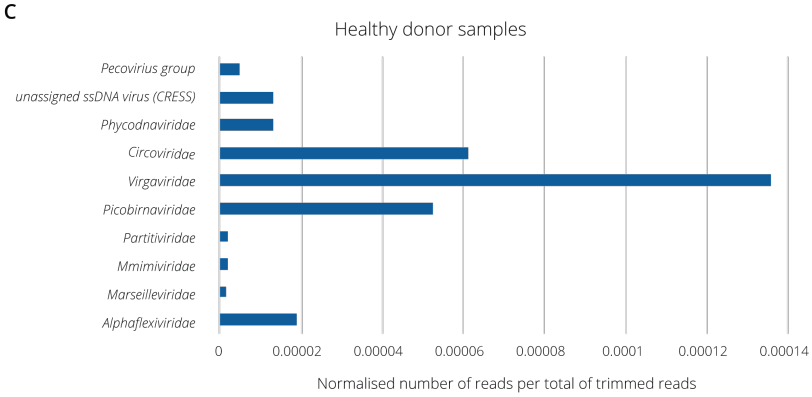
A



B



C



Supplementary figure 1 – (A) Overview of the eukaryotic viral families in responders and non-responders at all time points, **(B)** eukaryotic viral families in responders and non-responders at baseline, and **(C)** eukaryotic viral families in donors. Panel A includes 50 samples, panel B includes 9 samples and panel C includes 8 samples.