

Bacteriophage deficiency characterizes respiratory virome dysbiosis in childhood asthma

Spyridon Megremis¹, Bede Constantinides², Paraskevi Xepapadaki³, Claus Bachert⁴, Susetta Neurath-Finotto⁵, Tuomas Jartti^{6, 7}, Marek L Kowalski⁸, Alexandros Georgios Sotiropoulos⁹, Avraam Tapinos¹, Tytti Vuorinen¹⁰, Evangelos Andreakos¹¹, David Robertson^{1, 12}, Nikolaos Papadopoulos^{1, 3}

1 Division of Evolution and Genomic Science, University of Manchester, UK

2 Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, UK

3 Department of Allergy and Clinical Immunology, University of Athens, Greece

4 Upper Airways Research Laboratory, ENT-Department, University Hospital Ghent, Belgium

5 Department of Molecular Pneumonology, University of Erlangen-Nürnberg, Germany

6 Department of Paediatrics, University of Turku, Finland

7 Department of Paediatrics, University of Oulu, Finland

8 Department of Immunology and Allergy, Medical University of Lodz, Poland

9 Department of Plant and Microbial Biology, University of Zurich, Switzerland

10 Department of Clinical Virology, Turku University Hospital, Finland

11 Centre for Clinical, Experimental Surgery and Translational Research of the Biomedical Research Foundation of the Academy of Athens, Greece

12 Institute of Infection, Immunity and inflammation, University of Glasgow, UK

Keywords

Virome, asthma, bacteriophages, respiratory viruses, Anelloviruses

Abstract:

Asthma development and persistence is tightly linked to respiratory viruses. Viral presence is usually interrogated with targeted approaches during periods of disease activity and/or infections, thus neglecting viral occurrence during steady states. We investigate the virome in the upper respiratory system of healthy and asthmatic preschool children during asymptomatic/non-infection periods using metagenomics. Children with asthma have a characteristically dysbiotic virome that correlates to disease severity and control. The major component of dysbiosis is bacteriophage deficiency, while eukaryotic viral presence is increased. At the metacommunity level, differential virus species co-occurrence patterns suggest a decrease of the microbiota community resilience in asthma. Viral dysbiosis is therefore a key characteristic of asthma pathophysiology.

Introduction:

More than 300 million people suffer from asthma worldwide¹, affecting their quality of life, and reducing educational attainment in paediatric and adolescent populations². Asthma remains a major driver of morbidity throughout the productive years, while also affecting healthy aging³. Asthma susceptibility to viral and bacterial respiratory infections may influence the divergence from a healthy trajectory and is one of the main priorities for research and intervention^{1,4-8}.

The respiratory tract harbours heterogenous microbiota decreasing in biomass and richness from the ‘source’ upper respiratory tract (URT) towards the lower airway based on the

microbial immigration and elimination model^{9,10}. It has been demonstrated that bacterial communities are disrupted in disease or infectious conditions¹¹⁻¹⁵, while the bacterial component of the respiratory microbiome is increasingly recognised to have an important role in the susceptibility and severity of acute respiratory illness and asthma^{16,17}. The microbial colonisation of the nasal cavity and nasopharynx in early life has been linked with wheeze episodes and progression to asthma, potentially mediated through resistance or susceptibility to acute respiratory infections during childhood¹⁸⁻²².

Compared to cellular microbes, respiratory viruses are currently considered the most important drivers of asthma development, exacerbation and persistence²³⁻²⁷. Upper respiratory viruses, most importantly rhinoviruses (RV), are strongly implicated in the induction of airway hyperresponsiveness and airway remodelling^{1,4,28-31}. Impairment of innate immune responses in children with asthma results in defective pathogen recognition, impaired interferon release and suboptimal antiviral responses^{32,33}, and drives the development of biased T2 inflammation^{32,33}.

Notably, acute respiratory viruses have historically been evaluated in isolation from the respiratory microbial ecosystem³⁴. Most importantly, our knowledge on prokaryotic viruses infecting bacteria (bacteriophages or phages) is extremely limited, even though they are the most straightforward link between viruses and bacteria in the respiratory system³⁵. Only a handful of studies have investigated the respiratory prokaryotic virome^{36,37} and to our knowledge none in asthma.

We hypothesized that a respiratory virome, including both eukaryotic and prokaryotic viruses, is present in the upper airway and is affected in asthma. The aim of this study was to identify

DNA and RNA virus species in the upper respiratory system of healthy and asthmatic preschool children during a period of stable disease activity using a metagenomic approach³⁸. We provide key characteristics of viral dysbiosis in asthma, in the context of the respiratory virome.

Results:

Reduced bacteriophage presence in asthma

We processed nasopharyngeal samples (NPS) obtained from a subgroup of children with asthma and healthy controls recruited in the PreDicta cohort²⁷ (Figure S1 & S2A). The children did not have symptoms of a respiratory infection or asthma exacerbation for at least a month prior to sampling. Subject characteristics are shown in Supplementary table 1. The viral metagenome assembled genomes (vMAGs) identified in the discovery cohort (Figure S2B-S2D) were taxonomically organised in the order of the prokaryotic Caudovirales viruses (phages) and eukaryotic viruses of the Anelloviridae and Picornaviridae families. The phage genome frequency of occurrence was significantly less in asthma ($p:0.045$) (Figure 1A). We observed that 87.5% (7 out of 8) of these vMAGs were identified in healthy samples compared to 37.5% (3 out of 8) in asthma (Figure 1A), along with an elevated number of sequencing reads mapping to them (Figure 1B). We estimated the number of prokaryotic vMAGs as a function of increasing sample size and observed a clear divergence between the two groups (Figure 1C). Further sample-level metagenomic analysis revealed that phages were significantly underrepresented in asthma ($p:0.02$) (Figure 1D).

To validate our findings, we repeated the comparison in a second cohort of age-matched healthy and asthmatic children with increased Illumina read length and depth (validation cohort, Supplementary table 1) (Figure S2A). Compared to the discovery MiSeq output, this approach increased the number of contigs 5-fold, the number of vMAGs 2-fold, the viral

sequencing reads 22-fold, while including samples from different geographical locations (Figure S2A, S2B & S2E). As in the discovery cohort, the identified prokaryotic viruses belonged exclusively to the Caudovirales order of phages (Figure S2F). Most eukaryotic viral contigs mapped to acute respiratory viruses of the Picornaviridae (Rhinoviruses), Paramyxoviridae (Parainfluenza) and Anelloviridae vMAGs (Figure S2F). There was no significant difference in the prokaryotic vMAG frequency of occurrence ($p:0.065$) (Figure 1A). However, similar to the discovery cohort, a lower number phage vMAGs were observed in asthma (55.5%: 5/9) compared to health (100%: 9/9) (Figure 1A), along with an elevated number of sequencing reads per vMAG, regardless of geography (Figure 1B). The health and asthma sample-based rarefaction curves diverged showing reduced number of different phages in asthma (Figure 1E). The reduction was also clearly present independent of the geographical location, confirming a generalised observation (Figure S3). To describe phage ecology within each sample, we focused on samples with two or more prokaryotic vMAGs: The prokaryotic viral richness was significantly reduced in asthma compared to healthy controls ($p:0.003$) (Figure 1F). This was also observed when the geographical origin of the samples was considered (Figure 1G). The within-sample Shannon diversity of prokaryotic viruses was significantly decreased in asthma ($p:0.045$) (Figure 1H). Asthmatics were more similar to each other compared to healthy controls based on the prokaryotic virome unweighted composition (presence/absence), measured using the Jaccard ($p:0.0008$) and Sorensen ($p:0.0008$) indices of beta diversity (Figure 1I).

Overall, the majority of the prokaryotic sequencing reads were observed in samples from healthy children (Figure 1J). Furthermore, there were fewer sequencing reads in asthma for most of the identified phage vMAGs ($n=12$, 75%) (Figure 1K). Phages were also over-represented in the subset of vMAGs identified only in healthy children ($n=9$, 64.29%)

compared to the asthma-specific viral genomes (n=1, 5.88%) (Figure 1L & 1M). Within the ‘core’ virome, defined as vMAGs with more than 30% detection rate and more than 0.01% read abundance, bacteriophages accounted for 21% in healthy samples (Figure 1N), in contrast to 12% in asthma (Figure 1O). Collectively, these data show that the virome of preschool children with asthma is less rich and diverse in bacteriophages compared to healthy controls. This was observed in samples from three different geographical locations and at two different sequencing depths.

Increased eukaryotic virus presence in asthma

We then asked whether eukaryotic viruses followed a similar trend. 14/15 (93%) eukaryotic vMAGs were observed in the asthma samples, versus 60% (n=9) in samples from the healthy group (Figure 2A & 2B). The frequency of occurrence of eukaryotic vMAGs was higher in asthma in the HiSeq cohort (p:0.016) and the integrated dataset (p:0.002). The healthy and asthmatic sample-based rarefaction curves of the estimated number of eukaryotic vMAGs diverged in both the discovery and validation cohorts demonstrating increased eukaryotic vMAG presence in asthma (Figure 2C & 2D), regardless of geography (Figure S3). There were no significant differences when considering eukaryotic vMAG presence within each sample, nevertheless, in the discovery cohort we observed a marginal increase (p: 0.059) in the abundance-based Shannon diversity (Figure 2E) and a significant (p: 0.026) reduction in the Simpson evenness of asthma samples (Figure 2G). These differences were also present in the validation cohort when comparing healthy and asthma samples from Poland: Shannon diversity was significantly increased (p: 0.045) (Figure 2F) while the Simpson evenness was significantly decreased in asthma (p: 0.029) (Figure 2H). This however was not observed in samples from Finland. Collectively, these data suggest that the asthmatic virome is often more

diverse in eukaryotic viruses which acquire a more even distribution within each individual's virome compared to the healthy viromes.

Increased presence of Anelloviruses in asthma

Anelloviruses accounted for the majority of eukaryotic vMAGs in both health (88.8%) and asthma (85.7%) in the MiSeq cohort and more than 45% in the HiSeq cohort (48.5% and 46.5%, respectively) (Figure 2A & 2B), representing a significant fraction of the respiratory eukaryotic virome (Figure S2D & S2F). Anelloviruses were the most frequent eukaryotic virus identified in the discovery cohort as well as in the validation cohort (Figure 2B). Most of the identified Anelloviridae genomes were observed in asthma (Figure 2A & 2B). Anelloviruses were enriched (14/20 of Anelloviridae vMAGs) among the eukaryotic vMAGs which were observed with higher frequency in asthmatics compared to healthy donors (Figure 2B), as well as among the vMAGs observed only in asthma (Figure 1M). The estimated number of Anelloviridae vMAGs as a fraction of sample size was higher in asthma compared to the healthy group across both cohorts (Figure 2I & 2J). The within-sample Shannon diversity of Anelloviridae was higher in asthma ($p:0.024$), while the Simpson index was significantly lower ($p:0.008$) in the MiSeq cohort, suggesting that healthy children had a few dominating species while asthmatic patients had a richer, more diverse and even distribution of Anelloviruses (Figure 2K & 2L); However, the difference was not significant in the HiSeq cohort ($p>0.05$).

Divergent virome profiles linked with asthma and health

To further investigate the observed differentiation of the nasopharyngeal virome profiles in health and asthma we identified a set of 8 virome features in both discovery and validation cohorts (Figure 3A) and an additional 2 features in the latter (Figure 3B) that provided optimal hierarchical co-clustering of the healthy and asthma individuals; 70% (7/10) of healthy samples

in the discovery cohort and 81% (9/11) of healthy samples in the validation cohort co-clustered (Figure 3A-3D). Then we focused on the sub-clusters that were obtained under the above configurations (Figure 3E & 3F). Based on the virome characteristics of these clusters (Figure S4 & S5) 3 virome profile groups (VPGs) were assigned: Prokaryotic VPG; (PVPG, n=32), contained samples with high prokaryotic richness and low/intermediate richness of eukaryotic viruses and Anelloviridae, Eukaryotic VPG (EVPG, n=12) included samples with high eukaryotic richness and low/intermediate Anelloviridae richness, and Anelloviridae VPG (AVPG, n=12) contained samples with high Anelloviridae richness (Figure 3E & 3F); We then investigated the relevance of the VPGs in asthma presence, severity, control and treatment: 81% of the healthy samples were grouped in PVPG compared to 42.8% of asthma samples (Figure 3G). We observed a gradient distribution of children with different levels of asthma control in the three VPGs (Figure 3H), while 57% and 24% of children with moderate and mild asthma, respectively, were grouped in AVPG, (Figure 3I). Treatments were evenly distributed among the three VPGs (Brown-Forsyth and Welch ANOVA tests p:0.66, p:0.63), suggesting that asthma treatment does not affect the composition of the VPGs. Overall, our data show a significant link between the nasopharyngeal virome, and the presence, control and severity of asthma in pre-school children.

Dysbiotic structure of the asthmatic virome

To investigate the microbial community structure in health and asthma, we then studied the viral-viral and viral-bacterial co-occurrence. From our metagenomic sequencing, 355 bacterial metagenome assembled genomes (bMAGs) in the discovery and 522 in the validation cohorts have been identified. Their presence in health and asthma are depicted in Figure S6 and S7, respectively. We constructed two undirected microbial network meta-graphs representing an overview of microbial co-occurrences related to viruses in health and asthma (Pearson test

$p < 0.05$). The healthy meta-graph was composed of 328 nodes (MAGs) connected by 1180 edges (Figure 4A). The asthma meta-graph consisted of 326 MAGs with 1014 edges (Figure 4B). The normalised degree centrality, i.e. the number of interactions of each MAG, was significantly higher in the healthy meta-graph compared to the asthmatic ($p < 0.0001$), suggesting reduced connectivity in asthma (Figure 4C). The Caudovirales, Anelloviridae and Picornaviridae viral families were the largest hubs in both the healthy and asthma meta-graphs, while taxa of the *Bacillales*, *Pseudomonadales*, *Enterobacterales*, and *Lactobacillales* order were the most highly paired bacteria with viruses (Figure S8A & S8B). Since prokaryotic virus richness was reduced in asthma, we investigated the co-occurring pairs of bacteriophages with bacteria (Figure 4D); in asthma, bacteriophages co-occurred with a significantly ($p: 0.0001$) lower number of bacteria (Figure 4E) per bacterial order, compared to the healthy meta-graph, for most of the identified bacterial orders (Figure 4D). In the healthy meta-graph 71% (23/ 37) of virus-virus pairs involved taxa from different virus families (cross-family co-occurrence) (Figure S8C). In asthma, this was reduced to 40% (13/32) while 69% of them involved Enterovirus and Rhinovirus species A, B or C (Figure S8D). The above suggest that asthma is characterised by decreased connectivity (co-occurrence) within the virome-related microbiome, decreased number of bacteriophage-bacterial pairs, and increased cross-family co-occurrence of picornaviruses within the virome.

To investigate inter-viral and inter-bacterial correlation with regard to the number and type of viral-bacterial co-occurrence, but also in relation to the fraction of interactions within the meta-graph, we expressed the number of specific virus-bacterial pairs as a fraction of the total number of edges (interactions) in each meta-graph and calculated the virus-virus (Figure 4F & 4G) and bacteria-bacteria (Figure 4H & 4I) pairwise R coefficients (Pearson correlations). In asthma, viruses correlated with each-other less positively than viruses in the healthy meta-

graph ($p: 0.031$) (Figure 4J), based on their bacterial pairs. In contrast, the level of correlation amongst bacteria based on their virus-co-occurring pairs was significantly higher ($p < 0.0001$) in asthma compared to the healthy meta-graph (Figure 4K). These observations suggest a less efficient structural organisation of the virome-related microbiome meta-community in asthma, characterised by reduced connectivity, fewer co-occurring phage-bacteria pairs, increased co-occurrence of viruses from different families—especially rhinoviruses—and increased correlation among bacteria. All the above are strong indications of virome dysbiosis in preschool children with asthma.

Discussion

Despite the well-established effects of specific viral infections on the development, exacerbation and persistence of asthma²³⁻²⁶, the relationship between viral ecology of the airways and asthma remains poorly understood, and prokaryotic viruses in particular have received little attention. We show that during asymptomatic/infection-free periods, preschool children with asthma have a characteristically dysbiotic virome correlating with disease severity and control. The major component of dysbiosis is bacteriophage deficiency, while eukaryotic viral presence is increased, driven mainly by Anelloviruses and Picornaviruses. Networks of viral-viral and viral-bacterial co-occurrence were significantly loosened, suggesting that microbial communities may be less resilient in asthma. Viral dysbiosis therefore appears to be not only a key characteristic of asthma pathophysiology, but one that may be amenable to intervention³⁹⁻⁴¹.

Viral exposures—and subsequent interactions with the host—are constant in the respiratory tract⁴². Furthermore, at least some viruses may persist⁴³. Among these, prokaryotic viruses have been barely considered until now, despite their role in regulating bacterial abundance and

promoting competition, stability and resilience^{23,44-47} within microbial ecosystems, which in turn may direct asthma development, activity and persistence^{21,22,48,49}.

Bacteriophages are active regulators of bacterial populations⁵⁰⁻⁵² and changes in phage composition or diversity have been linked to various diseases including Inflammatory Bowel disease, Parkinson's disease and Type 1 Diabetes⁵³. Moreover, bacteriophages can protect the epithelium from bacterial infections in a mucus-dependent manner, or can modulate the innate and adaptive arms providing non-host immunity^{37,52,54-58}. In addition to reduced diversity and frequency of phage occurrence in asthmatics, the asthmatic metacommunity was characterised by a low number of phage-bacteria pairs, suggesting limited predator-prey ecological interactions⁵⁹⁻⁶². Phage-bacteria dynamics is complex⁶³ yet unexplored in the human airway and/or in relation to asthma. Microbial dysregulation through reduced phage presence is a novel concept in asthma pathophysiology. Changes in the bacteriome composition associated with risk of future asthma development and the loss of asthma control^{16,21,48,64} can potentially be linked to alterations in phage presence and/or dynamics. Our findings suggest that these scenarios are plausible, since virome profiles with reduced bacteriophages were associated with worsening of asthma control and severity.

The asthmatic virome was characterised by increased occurrence, richness and diversity of eukaryotic viruses. Since there is no reason to assume that exposure of children with asthma to viruses is governed by different epidemiological characteristics than healthy individuals, this could reflect differential levels of viral control through immune competence⁶⁵⁻⁶⁷; it is possible that asthmatics fail to clear eukaryotic viruses as efficiently as healthy children, allowing viruses to persist in the airways at low levels during asymptomatic periods. This may also explain a differential threshold for susceptibility to viral infection in asthma⁶⁸. This may be

further underpinned by the reduced evenness index of eukaryotic viruses in asthma, i.e. more than two species often contributing equally to the observed eukaryotic virome abundance, whereas the healthy eukaryotic virome was dominated by a single species even though more than two species could be detected.

Another association of the virome with the immune status can be drawn through Anelloviruses, which have been linked to functional immune competence⁶⁹⁻⁷³. We report that Anelloviruses are a stable component of the upper airway virome, as in other systems^{69,74-76}, and in some cases they are observed with increased diversity and reduced evenness in asthma. The epidemiology of Anelloviruses is not entirely understood, nor are the factors that underlie their wide taxonomic diversity. It is possible that reduced immune competence may facilitate the colonisation of an individual's airways with tight clusters of closely related Anelloviruses around founding clones (sometimes termed 'quasispecies'). Moreover, in the asthmatic metacommunity, most of the viral-viral co-occurrence pairs involved Rhinovirus and Anellovirus species. Anelloviruses interact with the human host and hijack the host's defence through direct suppression of NF κ B, the regulation of TLR9 by viral-CpGs, and the regulation of IFN signalling pathways by virus-encoded micro-RNAs⁷⁵. Notably, we have previously demonstrated that rhinoviruses are also encoded by genomes of extremely low CpG content which can influence TLR9-dependent stimulation⁷⁸⁻⁸². This suggests that the immunomodulatory effect of Anelloviruses could act in favour for respiratory infections or condition the presence of sub-symptomatic low-virulence viruses to maintain a background level of chronic inflammation^{70,71,83,84}. In support, we observed that the majority of asthma donors had an Anellovirus-rich (AVPG) or eukaryotic-rich (EVPG) virome and that compromised asthma control was associated with these virome profiles.

The reduced connectivity within the asthmatic virome-related metacommunity suggests a reduction of species co-occurrence and/or increased interpersonal variability. Increased microbiome heterogeneity has been suggested to be a hallmark of dysbiotic microbiomes^{14,85}, and our virome data strongly point in this direction. Interestingly, virus-based interbacterial correlations were greater in the asthmatic metacommunity, and lacked negative correlations. This suggests reduced modularity among microbes co-occurring with viruses, and thus a reduced ability to compartmentalise virus-induced microbiome perturbations in asthma, such as during the addition or elimination of a virus from the system⁸⁶. This type of microbiome dysregulation could lead to reduced stability, resilience and shifts in the microbial composition upon perturbations^{44,87,88}, for example during respiratory infections and/or virus-induced asthma exacerbations. Notably, differential pathogenic bacteria-to-bacteria correlations have also been identified in tonsils of atopic individuals⁸⁹.

In this study, a number of choices had to be made regarding the design, analysis and interpretation of the metagenomic data. First, metagenomic sequencing allows the identification of viral and microbial genomes or genome ‘traces’, without however concluding presence of infectious pathogens⁹⁰⁻⁹². Viral enrichment through gradient ultra-centrifugation can bias the sequencing output as it depends on previous knowledge of the physical properties of the virions or virus like particles^{93,94}. To tackle the above, we followed a sample-processing strategy filtering out ‘naked’ nucleic acids while retaining encapsulated viral sequences of DNA and RNA genomes^{95,96}, in order to be as close as possible in the identification of virus-like particles or viruses. Second, the sequencing strategy, as demonstrated, can have a significant effect in the identification of viral genomes and the subsequent meta-analysis. Sequencing too deep can blur a differentiating signal while sequencing at a low depth reduces diversity. To address this, we have used two different sequencing strategies and asked whether

we can observe the same or similar differences between health and asthma. This was also expanded to different geo-locations including potentially variable virus exposure characteristics. Third, we chose to focus at the ecological (community) level. The ecological indices used are either incidence- or abundance-based. Viruses do not follow uniform replication/infection kinetics and have different hosts, i.e. eukaryotic versus prokaryotic viruses. In contrast to bacteriome studies where the relative abundance of each species implies some level of time-dependent stable presence in the system, the same cannot be assumed for viruses. Even, at a longitudinal setting, unequal time-intervals would be needed to study stable viral presence. Thus, our data provide a snapshot view of virus presence in the nasopharynx. Finally, the cohorts comprised of preschool-age children, an age during which the diagnosis of asthma is challenging and prognosis uncertain. Nevertheless, asthma persists in a considerable proportion of such children providing the opportunity for evaluating further outcomes. As with other microbiome niches, the respiratory virome is expected to change with time; whether the observed pathophysiology is age-specific or not remains to be elucidated. This is also the case for different asthma phenotypes and severity levels. Nevertheless, the PreDicta cohort was shown to be representative of preschool asthma²⁷. Finally, the identified viromes and bacteriomes refer to the specific tissue niche of the nasopharynx and further investigation is needed to study viral presence along the respiratory tract.

Overall, we report novel findings in the virome of children with asthma with potential for intervention. Reduced phage presence and/or diversity might lead to dysregulated bacterial colonisation through the absence of appropriate ecological interactions compromising the system's resilience to environmental^{6,97-99} or immune-driven perturbations like asthma exacerbations^{49,100}. In the same direction, increased co-occurrence of eukaryotic viruses might increase the chance of a symptomatic infection or sustained inflammation. These events could

induce stochastic changes⁸⁵ in the respiratory metacommunity triggering a transition to compositional states^{16,21,49,64,101} with increased susceptibility for viral and bacterial infections, which in turn further compromise the system's stability and resilience. The temporal trajectories of these complex dynamics and their association with immune competence and disease persistence at different ages, are currently explored by the CURE consortium¹⁰².

Funding: The study acquired funding from the European Union's Horizon 2020 research and innovation programme CURE under grant agreement No 767015, and the European FP7-Health programme PREDICTA under Grant agreement ID: 260895. CURE: "Constructing a 'Eubiosis Reinstatement Therapy' for Asthma". PREDICTA: "Post-infectious immune reprogramming and its association with persistence and chronicity of respiratory allergic diseases".

Acknowledgements: We thank Dr Anna Sobanska for her valuable contribution in patients' recruitment and care at Lodz center, Poland.

Author contributions

Conceptualization, SM, NP; Methodology, SM, BC, AT, DR, NP; Software, SM, BC, AT, AGS; Investigation, SM, DR, NP.; Resources, VX, MK, CB, SNF, TJ, TV; Writing – Original Draft, SM, DR, NP; Writing – Review & Editing; SM, BC, AT, AGS, VX, MK, CB, SNF, TJ, TV, EA, DR, NP; Visualization, SM, BC.; Supervision, DR, NP.

Methods

Cohort description

Donors were recruited as part of the Predicta cohort¹⁰³. The PreDicta paediatric cohort was a two-year prospective multicentre study (five European regions) under the EU FP7 program.

The cohort was designed to prospectively evaluate wheeze/asthma persistence in pre-schoolers associated with viral/microbial exposures and immunological responses¹⁰³. Children with an asthma diagnosis¹⁰⁴ of mild to moderate severity according to GINA¹⁰⁵ within the last two years were recruited. Donors did not present an episode of an URT and/or asthma exacerbation during the preceding four weeks prior to inclusion and baseline evaluation. Since all donors were retrospectively monitored for two years, detailed knowledge of upper (URTIs) and lower (LRTIs) respiratory tract infection events following baseline evaluation. In this study we sought to characterise the nasopharyngeal metagenome of preschool children at a baseline, non-infectious and non-exacerbation state, with emphasis on the viral component (virome). Thus, the timing of sample collection was at least one month prior to and one month following a reported respiratory tract infection. Nasopharyngeal samples were included from two groups of donors: a discovery group from southern Europe (Athens, Greece), and a validation group comprising individuals from Lodz (Poland) and Turku (Finland) (Central and North Europe). Except for geographical origin, clinical characteristics were similar among donors from the two groups (Supplementary table 1). The discovery cohort included 10 healthy donors and 24 asthma patients. The validation cohort included 11 asthmatic patients and 11 healthy donors (Supplementary table 1).

Supplementary table 1: Characteristics of study participants.

MiSeq cohort (Discovery Group)		
	Healthy donors (n=10)	Asthma patients (n=24)
Centre, n (%)	Athens, 10 (100%)	Athens, 24 (100%)
Age (years), (mean ± SD)	5.01 ± 0.74	4.95 ± 0.62
Age range (years)	3.26 - 5.82	4.03-6.04
Male, n (%)	6 (60%)	14 (58.33%)
Atopy, n (%)	0 (0%)	14 (58.33%)
Rhinitis, n (%)	1 (10%)	6 (25%)
# number of URTIs last 12 months, (mean ± SD)	5.6 ± 1.91	6.3 ± 4.0
Maximum duration URTIs, (mean ± SD)	7.5 ± 1.58	9.2 ± 2.7

# number of LRTIs last 12 months, (mean \pm SD)	0	0.0 \pm 0.2
Antibiotic courses last 12 months	2.3 \pm 3.10	1.8 \pm 1.6
Maximum duration antibiotic courses last 12 months	7.43 \pm 3.29	8.2 \pm 3.3
Asthma control (controlled, partly controlled, uncontrolled)		10 (41.67%), 12 (50.00%), 2 (8.33%)
Asthma severity (intermittent, mild persistent, moderate persistent), n (%)		2(8.33%), 20 (83.33%), 2 (8.33%)
# number of respiratory episodes last 3 months, (mean \pm SD)		2.37 \pm 1.86
# number of respiratory episodes last 12 months, (mean \pm SD)		6.17 \pm 3.97
HiSeq cohort (Verification Group)		
	Healthy donors (n=11)	Asthma patients (n=11)
Centre	Lodz, 5 (45.4%), Turku 6, (54.5%)	Lodz, 7 (63.6%), Turku 4, (36.3%)
Age (years), (mean \pm SD)	5.3 \pm 0.68	5.12 \pm 0.70
Age range (years)	4.04-6.29	4.13-5.98
Male, n (%)	3 (27.2%)	5 (45.4%)
Atopy, n (%)	1 (9.1%)	5 (45.4%)
Rhinitis, n (%)	0 (0%)	9 (81.8%)
# number of URTIs last 12 months, (mean \pm SD)	3.1 \pm 1.75	5.54 \pm 5.83
Maximum duration URTIs, (mean \pm SD)	7.7 \pm 2.90	9.44 \pm 4.12
# number of LRTIs last 12 months, (mean \pm SD)	0.18 \pm 0.6	2.54 \pm 3.77
Maximum duration LRTIs, (mean \pm SD)	1.27 \pm 4.22	5.1 \pm 7.06
Antibiotic courses last 12 months	1.18 \pm 1.40	3.54 \pm 3.17
Maximum duration antibiotic courses last 12 months	3.18 \pm 3.65	5.63 \pm 6.62
Asthma control (controlled, partly controlled, uncontrolled)		2 (18.18%), 6 (54.54%), 3 (27.27%)
Asthma severity (intermittent, mild persistent, moderate persistent), n (%)		6 (54.54%), 5 (45.45%), 0 (0 %)
# number of respiratory episodes last 3 months, (mean \pm SD)		2.1 \pm 1.44
# number of respiratory episodes last 12 months, (mean \pm SD)		7.45 \pm 4.74

Sample processing

Nasopharyngeal samples were obtained using flocked nasopharyngeal swabs (ESwab™ collection and transport system, COPAN)¹⁰⁶. Specimens were collected using the same standard operational procedure (SOP) across the recruitment centres. The nasopharyngeal

specimen was eluted in 1000 µl of transport medium (Amies buffer), split into 200 µl aliquots and stored at -80°C. One aliquot was sent to the University of Manchester for metagenomic analysis. Samples obtained from asthma patients and healthy participants were analysed in parallel to reduce bias. Samples were slowly defrosted and centrifuged under chill conditions at 2500g for 5 minutes to spin down cells and debris¹⁰⁷. Caesium chloride purification was not used because it can lead to bias towards specific virion like particles (VLPs)^{93,94}. To enrich for encapsulated nucleic acids, supernatants were split into four aliquots (50 µl) and treated with a mixture containing 5 units of Turbo DNase (Ambion-ThermoFisher Scientific) and 2.5 µl RNase inhibitor (RNaseOUT™, Invitrogen-ThermoFisher Scientific) in DNase reaction buffer (Ambion-ThermoFisher Scientific) and incubated for 45 minutes at 37°C¹⁰⁸. Ten microliters of DNase inactivation reagent were added to each aliquot and incubated for 5 minutes at room temperature. Samples were centrifuged at 10,000g for 1.5 minutes and supernatants were pooled back together per sample (total of 180 µl) and used for subsequent nucleic acid isolation. All samples were processed in a class 2 biosafety cabinet to reduce environmental contamination. A negative control sample (sterile filtered water) was introduced to each experiment, followed the whole protocol and the DNA concentration was measured before library preparation using a Qubit™ fluorometer (ThermoFisher Scientific). Negative controls did not produce measurable readings (estimated concentration <10 pg/µL).

Metagenomic nucleic acid extraction

Nucleic acids were extracted using the AllPrep DNA/RNA Micro kit (Qiagen) for simultaneous purification of minute amounts of DNA and RNA from the same sample without making use of the carrier RNA. An additional on-column DNase treatment step (RNase free DNase set, Qiagen) was added to remove possible traces of remaining DNA co-extracted with the RNA. The quality of DNA and RNA was evaluated in a NanoDrop™1000 spectrophotometer

(ThermoFisher Scientific) and the yield in the QubitTM fluorometer (ThermoFisher Scientific) using dsDNA and ssRNA high sensitivity assays (ThermoFisher Scientific). Synthesis of the first strand of cDNA was carried out using the SuperScriptTM II reverse transcriptase system and random primers (ThermoFisher Scientific) with default conditions. The second strand of the cDNA was synthesised using Klenow fragment polymerase (New England Biolabs). Ten microliters of cDNA were incubated for 2 min at 95°C and chilled on ice for 2 min before addition of 5 units of Klenow fragment and incubation at 37°C for 1 h. Enzyme inactivation was performed at 75°C for 10 min. The resulting product was a double-stranded copy of the initial single-stranded cDNA. To amplify the ds-cDNA and genomic DNA, a whole genome amplification strategy was carried out using multiple displacement amplification^{109,110}, TruePrimeTMWGA kit (SYGNIS)¹¹¹⁻¹¹³. Samples were incubated for 3 hours at 30°C followed by polymerase inactivation at 65°C for 10 minutes. DNA amplification yield was confirmed by fluorometric measurement as described before. The whole process negative control samples did not produce any measurable readings (estimated concentration < 10 pg/μL). Overall, each swab specimen produced two amplified dsDNA samples, one derived from the initial isolated DNA (metagenomic DNA sequences) and a second from the initial isolated RNA (metagenomic RNA sequences).

Metagenomic high throughput sequencing

An aliquot of 10 μl (Normalisation: 1 ng of DNA per reaction) from each sample was transferred to the Genomic Technologies Core Facility (GTCF) in the University of Manchester. Illumina sequencing libraries were generated using ‘on-bead’ tagmentation chemistry with the Nextera DNA Flex Library Prep Kit (Illumina, Inc.) according to the manufacturer’s protocol. For the MiSeq cohort, DNA libraries were pair-end sequenced using the MiSeq high throughput sequencing instrument (Illumina) in 2x75 base pair (bp) runs. For

the HiSeq cohort DNA libraries were pair-end sequenced using the HiSeq4000 high throughput sequencing platform (Illumina) in 2x150 bp. Each flow cell contained libraries from both asthma patients and healthy children to avoid sequencing bias. The output data was de-multiplexed and BCL-to-FASTQ conversion was performed using Illumina's bcl2fastq software, version 2.17.1.14.

De novo assembly, microbial annotation and taxonomy assignment

We used customised computational strategies to achieve *de novo* assembly of contigs from short read sequences: In the MiSeq cohort (MiSeq, low read output), reads with alignments to host sequences (hg38) were discarded (BBMap 36.92) and remaining reads were trimmed of adapter sequences and low-quality regions (BBDuk 36.92). Filtered paired and orphan reads were pooled across all individuals and co-assembled (MetaSPAdes 3.9.1; k-mers 17,21,25,31,43,55,67; doi:10.1101/gr.213959.116) before being queried against the NCBI NR protein database dated 2017-01 (Diamond 0.8.34; doi:10.1038/nmeth.3176). Taxonomic ranks were assigned to each contig from protein homology search results using the weighted lowest common ancestor algorithm (wLCA) <https://arxiv.org/abs/1511.08753> as implemented in MEGAN 6 (doi:10.1101/gr.5969107) using a 50% match weighting. Contigs assigned known non-microbial (namely host) taxa were discarded. Finally, reads from individuals were mapped to combined meta-assembly contigs (BBMap 36.92; seed k=7) representative of i) entire microbiome and ii) viruses, enabling inference of relative taxonomic abundances for each individual. More than 70% (78.92%) of contigs were classified, mapping to 380 distinct microbial taxa including 25 viral metagenomic assembled genomes (vMAGs). A total of 214,533,315 sequencing reads from 34 complete metagenomic specimens (Mean log₁₀: 5.785±SEM 0.154) were aligned back into MAGs. 5,351,667 sequencing reads aligned to 25 vMAGs (Mean log₁₀: 2.156±SEM 0.434).

In the HiSeq cohort (HiSeq, high read output), BBTools (38.23) were used to decontaminate libraries from human sequences and remove artefacts (BBDuk and BBMap, sliding window based quality filtering at Q10). Filtered paired and orphan reads were co-assembled and individual-based assembled. For co-assembly, sequencing reads from DNA & RNA-derived samples were assembled with MetaSpades (3.12.0) at k-mers k 21, 33, 55, 77, 99, 127. Megahit (1.1.3) was used for individual-based assembly with default K-mers (k 21, 29, 39, 59, 79, 99, 119, 141). MetaQUAST was used to merge co-assembled and individual-based assemblies. LCA was performed in DIAMOND 0.9.22 (NR database and taxon mapping downloaded 2018-09-13). Sequencing reads remapped (BBMap, default settings with scafstats output) to all MetSPAdes metagenomic assembled genomes. Overall, we produced a total of 34,031 classified contigs, organised in 576 MAGs, including 54 viral MAGs (vMAGs). About 700 million sequencing reads (712,176,554) from 22 complete metagenomic specimens aligned onto the microbiome MAGs (Mean \log_{10} : $6.684 \pm \text{SEM } 0.113$). Of those, 121,694,232 sequencing reads aligned to vMAGs (Mean \log_{10} : $4.782 \pm \text{SEM } 0.299$). A summary of the complete microbiome MAGs identified in healthy and asthmatic children can be seen in Figure S6 and Figure S7, respectively. Viral genomes were identified within the ENA sequence viral database using EBI's BLASTN REST API¹⁴ and 7mer MASH distances were generated with Sourmash¹⁵ (Figure S1G).

Estimation of vMAG ecological indexes

To avoid confounding effects of sample size and underestimation of true species richness, we have performed sample-based randomisation and rarefaction with extrapolation. We have evaluated the asymptotic richness estimator at all levels of taxonomic rank accumulation (rarefaction) up to the size of the reference sample: MiSeq cohort (healthy: $n = 10$, asthma:

$n = 24$), HiSeq cohort (healthy: $n = 11$, asthma: $n = 11$) using 100 randomizations. The means (and conditional standard deviations) among resamples for each level of accumulation were reported. Extrapolation was performed by a factor of 1 (EstimateS, version 9.1.0)¹¹⁶. For comparison of the sample-based rarefaction curves (richness divergence between clinical groups), non-overlap of 95% confidence intervals constructed from the unconditional variance estimators is used as a simple but conservative criterion of statistical difference¹¹⁶. To define within-donor (individual-based) vMAG alpha diversity we have used the number of vMAGs observed per analysed specimen, the abundance-based Shannon diversity and the abundance-based Simpson evenness¹¹⁷. For abundance-based ecological indexes, vMAG abundance was computed using the relative read abundance; the number of sequencing reads per vMAG was normalised against the total number of sequencing reads observed per taxonomical group or any other type of grouping, e.g. eukaryotic and prokaryotic viruses. We computed the between-samples diversity for each assemblage using incidence-based (Jaccard and Sorensen) measures of relative compositional similarity (EstimateS, version 9.1.0)¹¹⁶. For incidence-based similarity, the Jaccard index compares the number of shared species to the total number of species in the combined samples of each assemblage (global view), while the Sorensen index compares the number of shared species to the mean number of species in a single assemblage (local view)¹¹⁸.

Species co-occurrence networks

We removed singletons and doubletons per each clinical group and sequencing strategy and retained MAGs with at least 3 occurrence events. We calculated the pairwise R coefficients of the MAGs based on their relative sequencing read abundance within the metagenome. The pairwise comparison matrices were transformed into interaction graphs: each node represented a microbial lineage in the LCA (MAG) and an edge represented a positive or negative

correlation between two lineages (MAGs). To dichotomize the presence or absence of a co-occurrence link, only significant (Pearson test $p < 0.05$) pairwise correlations involving a vMAG were selected; these included 40 co-occurring pairs in the healthy and 36 in the asthma MiSeq cohort, and 1765 pairs in the healthy and 1412 in the asthma HiSeq cohort. All graphs were visualised using Network Analysis, Visualization, & Graphing TORonto (NAViGaTOR)¹¹⁹. Networks were organised in order to separate different components using the GRIP algorithm or the concentric view option. Node and edge descriptives were analysed using NAViGaTOR.

Statistics

Continuous variables were tested for normal distribution using the Shapiro-Wilk test. Pairwise quantitative differences between two groups were tested using two-tailed t test and Mann Whitney (Wilcoxon rank sum test) for variables not following a normal distribution. F test was used to test for equal variance. Pairwise comparisons between three or more groups was performed with the Brown-Forsythe and Welch ANOVA tests corrected for multiple comparisons by controlling the False Discovery Rate (FDR) for normally distributed variables and the Kruskal-Wallis test corrected for multiple comparisons (Dunn's test) for variables not following a normal distribution. Principal component analysis, hierarchical agglomerative clustering (average and weighted linkage), K-means clustering and feature reduction based on variance was performed on normalised variables with mean of 0 and variance of 1 (Qlucore Omics explorer, version 3.4); We used principal component analysis and exploratory visualization to find a representation from which we can extract interpretable and potentially interesting information, following the evolution of the projection score in real time during variance filtering (Qlucore Omics explorer, version 3.4). The Pearson test was used to test for pairwise correlations among continuous variables. Statistical tests and graphical visualisations were produced in GraphPad prism 8.02.

References:

1. Holgate ST, Wenzel S, Postma DS, Weiss ST, Renz H, Sly PD. Asthma. *Nature Reviews Disease Primers* 2015;1:15025.
2. Nunes C, Pereira AM, Morais-Almeida M. Asthma costs and social impact. *Asthma Res Pract* 2017;3:1.
3. Bousquet J, Barbara C, Bateman E, et al. AIRWAYS-ICPs (European Innovation Partnership on Active and Healthy Ageing) from concept to implementation. *Eur Respir J* 2016;47:1028-33.
4. Juhn YJ. Risks for infection in patients with asthma (or other atopic conditions): is asthma more than a chronic airway disease? *J Allergy Clin Immunol* 2014;134:247-57; quiz 58-9.
5. Bisgaard H, Hermansen MN, Bonnelykke K, et al. Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. *BMJ* 2010;341:c4978.
6. Busse WW, Lemanske RF, Jr., Gern JE. Role of viral respiratory infections in asthma and asthma exacerbations. *Lancet* 2010;376:826-34.
7. Mathioudakis AG, Custovic A, Deschildre A, et al. Research Priorities in Pediatric Asthma: Results of a Global Survey of Multiple Stakeholder Groups by the Pediatric Asthma in Real Life (PeARL) Think Tank. *J Allergy Clin Immunol Pract* 2020;8:1953-60 e9.
8. Papadopoulos NG, Custovic A, Cabana MD, et al. Pediatric asthma: An unmet need for more effective, focused treatments. *Pediatr Allergy Immunol* 2019;30:7-16.
9. Charlson ES, Bittinger K, Haas AR, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011;184:957-63.
10. Dickson RP, Erb-Downward JR, Freeman CM, et al. Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. *Ann Am Thorac Soc* 2015;12:821-30.
11. de Steenhuijsen Piters WA, Sanders EA, Bogaert D. The role of the local microbial ecosystem in respiratory health and disease. *Philos Trans R Soc Lond B Biol Sci* 2015;370.
12. Durack J, Boushey HA, Lynch SV. Airway Microbiota and the Implications of Dysbiosis in Asthma. *Curr Allergy Asthma Rep* 2016;16:52.
13. Man WH, de Steenhuijsen Piters WA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* 2017;15:259-70.
14. Hooks KB, O'Malley MA. Dysbiosis and Its Discontents. *MBio* 2017;8.
15. Vissers M, de Groot R, Ferwerda G. Severe viral respiratory infections: are bugs bugging? *Mucosal Immunol* 2014;7:227-38.
16. Fazlollahi M, Lee TD, Andrade J, et al. The nasal microbiome in asthma. *J Allergy Clin Immunol* 2018.
17. Huang YJ, Boushey HA. The microbiome in asthma. *J Allergy Clin Immunol* 2015;135:25-30.
18. Esposito S, Principi N. Impact of nasopharyngeal microbiota on the development of respiratory tract diseases. *Eur J Clin Microbiol Infect Dis* 2018;37:1-7.
19. Biesbroek G, Tsivtsivadze E, Sanders EA, et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med* 2014;190:1283-92.
20. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, et al. Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study. *Am J Respir Crit Care Med* 2017;196:1582-90.
21. Teo SM, Mok D, Pham K, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* 2015;17:704-15.
22. Tsai MH, Huang SH, Chen CL, et al. Pathogenic bacterial nasopharyngeal colonization and its impact on respiratory diseases in the first year of life: the PATCH Birth Cohort Study. *Pediatr Infect Dis J* 2015;34:652-8.

23. Rosenthal LA, Avila PC, Heymann PW, et al. Viral respiratory tract infections and asthma: the course ahead. *J Allergy Clin Immunol* 2010;125:1212-7.
24. Jartti T, Gern JE. Role of viral infections in the development and exacerbation of asthma in children. *J Allergy Clin Immunol* 2017;140:895-906.
25. Mikhail I, Grayson MH. Asthma and viral infections: An intricate relationship. *Ann Allergy Asthma Immunol* 2019;123:352-8.
26. Selby A, Munro A, Grimshaw KE, et al. Prevalence estimates and risk factors for early childhood wheeze across Europe: the EuroPrevall birth cohort. *Thorax* 2018;73:1049-61.
27. Xepapadaki P, Bachert C, Finotto S, et al. Contribution of repeated infections in asthma persistence from preschool to school age: Design and characteristics of the PreDicta cohort. *Pediatr Allergy Immunol* 2018.
28. Bizzintino J, Lee WM, Laing IA, et al. Association between human rhinovirus C and severity of acute asthma in children. *Eur Respir J* 2011;37:1037-42.
29. Gern JE. Rhinovirus and the initiation of asthma. *Curr Opin Allergy Clin Immunol* 2009;9:73-8.
30. Holt PG, Sly PD. Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. *Nat Med* 2012;18:726-35.
31. Guibas GV, Tsolia M, Christodoulou I, Stripeli F, Sakkou Z, Papadopoulos NG. Distinction between rhinovirus-induced acute asthma and asthma-augmented influenza infection. *Clin Exp Allergy* 2018;48:536-43.
32. Lejeune S, Deschildre A, Le Rouzic O, et al. Childhood asthma heterogeneity at the era of precision medicine: Modulating the immune response or the microbiota for the management of asthma attack. *Biochem Pharmacol* 2020:114046.
33. Brasier AR. Identification of innate immune response endotypes in asthma: implications for personalized medicine. *Curr Allergy Asthma Rep* 2013;13:462-8.
34. Skjerven HO, Megremis S, Papadopoulos NG, et al. Virus Type and Genomic Load in Acute Bronchiolitis: Severity and Treatment Response With Inhaled Adrenaline. *J Infect Dis* 2016;213:915-21.
35. Wylie KM. The Virome of the Human Respiratory Tract. *Clin Chest Med* 2017;38:11-9.
36. Mitchell AB, Oliver BG, Glanville AR. Translational Aspects of the Human Respiratory Virome. *Am J Respir Crit Care Med* 2016;194:1458-64.
37. Li Y, Fu X, Ma J, et al. Altered respiratory virome and serum cytokine profile associated with recurrent respiratory tract infections in children. *Nat Commun* 2019;10:2288.
38. Rose R, Constantinides B, Tapinos A, Robertson DL, Prospero M. Challenges in the analysis of viral metagenomes. *Virus Evol* 2016;2:vew022.
39. Potaczek DP, Unger SD, Zhang N, et al. Development and characterization of DNAzyme candidates demonstrating significant efficiency against human rhinoviruses. *J Allergy Clin Immunol* 2019;143:1403-15.
40. Papadopoulos NG, Megremis S, Kitsioulis NA, Vangelatou O, West P, Xepapadaki P. Promising approaches for the treatment and prevention of viral respiratory illnesses. *J Allergy Clin Immunol* 2017;140:921-32.
41. Divya Ganeshan S, Hosseinidoust Z. Phage Therapy with a Focus on the Human Microbiota. *Antibiotics (Basel)* 2019;8.
42. Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* 2018;12:1154-62.
43. Randall RE, Griffin DE. Within host RNA virus persistence: mechanisms and consequences. *Curr Opin Virol* 2017;23:35-42.
44. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. *Science* 2015;350:663-6.
45. Manrique P, Dills M, Young MJ. The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses* 2017;9.

46. Lim ES, Zhou Y, Zhao G, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med* 2015;21:1228-34.
47. Guibas GV, Megremis S, West P, Papadopoulos NG. Contributing factors to the development of childhood asthma: working toward risk minimization. *Expert Rev Clin Immunol* 2015;11:721-35.
48. Teo SM, Tang HHF, Mok D, et al. Airway Microbiota Dynamics Uncover a Critical Window for Interplay of Pathogenic Bacteria and Allergy in Childhood Respiratory Disease. *Cell Host Microbe* 2018;24:341-52 e5.
49. Tang HH, Teo SM, Belgrave DC, et al. Trajectories of childhood immune development and respiratory health relevant to asthma and allergy. *Elife* 2018;7.
50. De Sordi L, Lourenco M, Debarbieux L. "I will survive": A tale of bacteriophage-bacteria coevolution in the gut. *Gut Microbes* 2019;10:92-9.
51. Hsu BB, Gibson TE, Yeliseyev V, et al. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host Microbe* 2019;25:803-14 e5.
52. Duerkop BA. Bacteriophages shift the focus of the mammalian microbiota. *PLoS Pathog* 2018;14:e1007310.
53. Sinha A, Maurice CF. Bacteriophages: Uncharacterized and Dynamic Regulators of the Immune System. *Mediators Inflamm* 2019;2019:3730519.
54. Barr JJ, Auro R, Furlan M, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci U S A* 2013;110:10771-6.
55. Almeida GMF, Laanto E, Ashrafi R, Sundberg LR. Bacteriophage Adherence to Mucus Mediates Preventive Protection against Pathogenic Bacteria. *mBio* 2019;10.
56. Barr JJ, Youle M, Rohwer F. Innate and acquired bacteriophage-mediated immunity. *Bacteriophage* 2013;3:e25857.
57. Barr JJ, Auro R, Sam-Soon N, et al. Subdiffusive motion of bacteriophage in mucosal surfaces increases the frequency of bacterial encounters. *Proc Natl Acad Sci U S A* 2015;112:13675-80.
58. Van Belleghem JD, Dabrowska K, Vanechoutte M, Barr JJ, Bollyky PL. Interactions between Bacteriophage, Bacteria, and the Mammalian Immune System. *Viruses* 2018;11.
59. Mainali K, Bewick S, Vecchio-Pagan B, Karig D, Fagan WF. Detecting interaction networks in the human microbiome with conditional Granger causality. *PLoS Comput Biol* 2019;15:e1007037.
60. Faust K, Sathirapongsasuti JF, Izard J, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* 2012;8:e1002606.
61. Duran-Pinedo AE, Paster B, Teles R, Frias-Lopez J. Correlation network analysis applied to complex biofilm communities. *PLoS One* 2011;6:e28438.
62. Wilkins LGE, Leray M, O'Dea A, et al. Host-associated microbiomes drive structure and function of marine ecosystems. *PLoS Biol* 2019;17:e3000533.
63. Weitz JS, Poisot T, Meyer JR, et al. Phage-bacteria infection networks. *Trends Microbiol* 2013;21:82-91.
64. Zhou Y, Jackson D, Bacharier LB, et al. The upper-airway microbiota and loss of asthma control among asthmatic children. *Nat Commun* 2019;10:5714.
65. Holgate ST. Innate and adaptive immune responses in asthma. *Nat Med* 2012;18:673-83.
66. Ardain A, Marakalala MJ, Leslie A. Tissue-resident innate immunity in the lung. *Immunology* 2020;159:245-56.
67. Baraldo S, Contoli M, Bonato M, et al. Deficient Immune Response to Viral Infections in Children Predicts Later Asthma Persistence. *Am J Respir Crit Care Med* 2018;197:673-5.
68. Ramette A, Spycher BD, Wang J, Goutaki M, Beardsmore CS, Kuehni CE. Longitudinal Associations Between Respiratory Infections and Asthma in Young Children. *Am J Epidemiol* 2018;187:1714-20.
69. Focosi D, Antonelli G, Pistello M, Maggi F. Torquetenovirus: the human virome from bench to bedside. *Clin Microbiol Infect* 2016;22:589-93.
70. Young JC, Chehoud C, Bittinger K, et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant* 2015;15:200-9.

71. Abbas AA, Diamond JM, Chehoud C, et al. The Perioperative Lung Transplant Virome: Torque Tenoviruses Are Elevated in Donor Lungs and Show Divergent Dynamics in Primary Graft Dysfunction. *Am J Transplant* 2017;17:1313-24.
72. Siqueira JD, Curty G, Xutao D, et al. Composite Analysis of the Virome and Bacteriome of HIV/HPV Co-Infected Women Reveals Proxies for Immunodeficiency. *Viruses* 2019;11.
73. Mitchell AB, Glanville AR. Kinetics of TTV-DNA Plasma Load: A Global Measure of Immune Suppression? *Transplantation* 2019;103:660-1.
74. Maggi F, Bendinelli M. Human anelloviruses and the central nervous system. *Rev Med Virol* 2010;20:392-407.
75. Freer G, Maggi F, Pifferi M, Di Cicco ME, Peroni DG, Pistello M. The Virome and Its Major Component, Anellovirus, a Convolved System Molding Human Immune Defenses and Possibly Affecting the Development of Asthma and Respiratory Diseases in Childhood. *Front Microbiol* 2018;9:686.
76. Spandole S, Cimponeriu D, Berca LM, Mihaescu G. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol* 2015;160:893-908.
77. Umemura T, Tanaka Y, Kiyosawa K, Alter HJ, Shih JW. Observation of positive selection within hypervariable regions of a newly identified DNA virus (SEN virus)(1). *FEBS Lett* 2002;510:171-4.
78. Megremis S, Demetriou P, Makrinioti H, Manoussaki AE, Papadopoulos NG. The genomic signature of human rhinoviruses A, B and C. *PLoS One* 2012;7:e44557.
79. Antzin-Anduetza I, Mahiet C, Granger LA, Odendall C, Swanson CM. Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* 2017;14:49.
80. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res* 2014;42:4527-45.
81. Gaunt E, Wise HM, Zhang H, et al. Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *Elife* 2016;5:e12735.
82. Jimenez-Baranda S, Greenbaum B, Manches O, et al. Oligonucleotide motifs that disappear during the evolution of influenza virus in humans increase alpha interferon secretion by plasmacytoid dendritic cells. *J Virol* 2011;85:3893-904.
83. Maggi F, Pifferi M, Tempestini E, et al. Correlation between Torque tenovirus infection and serum levels of eosinophil cationic protein in children hospitalized for acute respiratory diseases. *J Infect Dis* 2004;190:971-4.
84. Li L, Deng X, Linsuwanon P, et al. AIDS alters the commensal plasma virome. *J Virol* 2013;87:10912-5.
85. Zaneveld JR, McMinds R, Vega Thurber R. Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat Microbiol* 2017;2:17121.
86. Aderem A. Systems biology: its practice and challenges. *Cell* 2005;121:511-3.
87. Gao J, Barzel B, Barabasi AL. Universal resilience patterns in complex networks. *Nature* 2016;530:307-12.
88. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol* 2012;10:538-50.
89. Ivaska LE, Hanif T, Ahmad F, et al. Tonsillar microbial diversity, abundance, and interrelations in atopic and non-atopic individuals. *Allergy* 2020.
90. Autenrieth IB. The microbiome in health and disease: a new role of microbes in molecular medicine. *J Mol Med (Berl)* 2017;95:1-3.
91. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2012;2:63-77.
92. Rascovan N, Duraisamy R, Desnues C. Metagenomics and the Human Virome in Asymptomatic Individuals. *Annu Rev Microbiol* 2016;70:125-41.

93. Kleiner M, Hooper LV, Duerkop BA. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 2015;16:7.
94. Duerkop BA, Kleiner M, Paez-Espino D, et al. Murine colitis reveals a disease-associated bacteriophage community. *Nature Microbiology* 2018.
95. Nelson MT, Pope CE, Marsh RL, et al. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Rep* 2019;26:2227-40 e5.
96. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 2018;6:42.
97. Haahtela T, Holgate S, Pawankar R, et al. The biodiversity hypothesis and allergic disease: world allergy organization position statement. *World Allergy Organ J* 2013;6:3.
98. Liu AH. Revisiting the hygiene hypothesis for allergy and asthma. *J Allergy Clin Immunol* 2015;136:860-5.
99. Ruokolainen L. Green living environment protects against allergy, or does it? *Eur Respir J* 2017;49.
100. Andreakos E. Asthma exacerbations: a molecular dichotomy between antiviral and pro-inflammatory responses revealed. *EMBO Mol Med* 2012;4:1231-3.
101. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012;7:e30126.
102. Constructing a 'Eubiosis Reinstatement Therapy' for Asthma. at <https://www.cureasthma.eu/about-us>.)
103. Xepapadaki P, Bachert C, Finotto S, et al. Contribution of repeated infections in asthma persistence from preschool to school age: Design and characteristics of the PreDicta cohort. *Pediatr Allergy Immunol* 2018;29:383-93.
104. Bacharier LB, Boner A, Carlsen KH, et al. Diagnosis and treatment of asthma in childhood: a PRACTALL consensus report. *Allergy* 2008;63:5-34.
105. Bateman ED, Hurd SS, Barnes PJ, et al. Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J* 2008;31:143-78.
106. Luna PN, Hasegawa K, Ajami NJ, et al. The association between anterior nares and nasopharyngeal microbiota in infants hospitalized for bronchiolitis. *Microbiome* 2018;6:2.
107. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 2009;4:470-83.
108. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* 2001;98:11609-14.
109. Hosono S, Faruqi AF, Dean FB, et al. Unbiased whole-genome amplification directly from clinical samples. *Genome Res* 2003;13:954-64.
110. Lasken RS. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem Soc Trans* 2009;37:450-3.
111. Picher AJ, Budeus B, Wafzig O, et al. TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. *Nat Commun* 2016;7:13296.
112. Silander K, Saarela J. Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol Biol* 2008;439:1-18.
113. Deleye L, De Coninck D, Dheedene A, et al. Performance of a TthPrimPol-based whole genome amplification kit for copy number alteration detection using massively parallel sequencing. *Sci Rep* 2016;6:31825.
114. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 2019;47:W636-W41.
115. al Be. sourmash: a library for MinHash sketching of DNA, *Journal of Open Source Software*, 2016;1(5), 27, .

116. Colwell RK. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. 2013.
117. L.J. Entropy and diversity. *Oikos* 2006;113:363–75.
118. Lou Jost AC, and Robin L. Chazdon. Compositional Similarity and Beta Diversity. In: McGill AEMaBJ, ed. *Biological Diversity Frontiers in Measurement and Assessment* 2010.
119. Brown KR, Otasek D, Ali M, et al. NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. *Bioinformatics* 2009;25:3327-9.

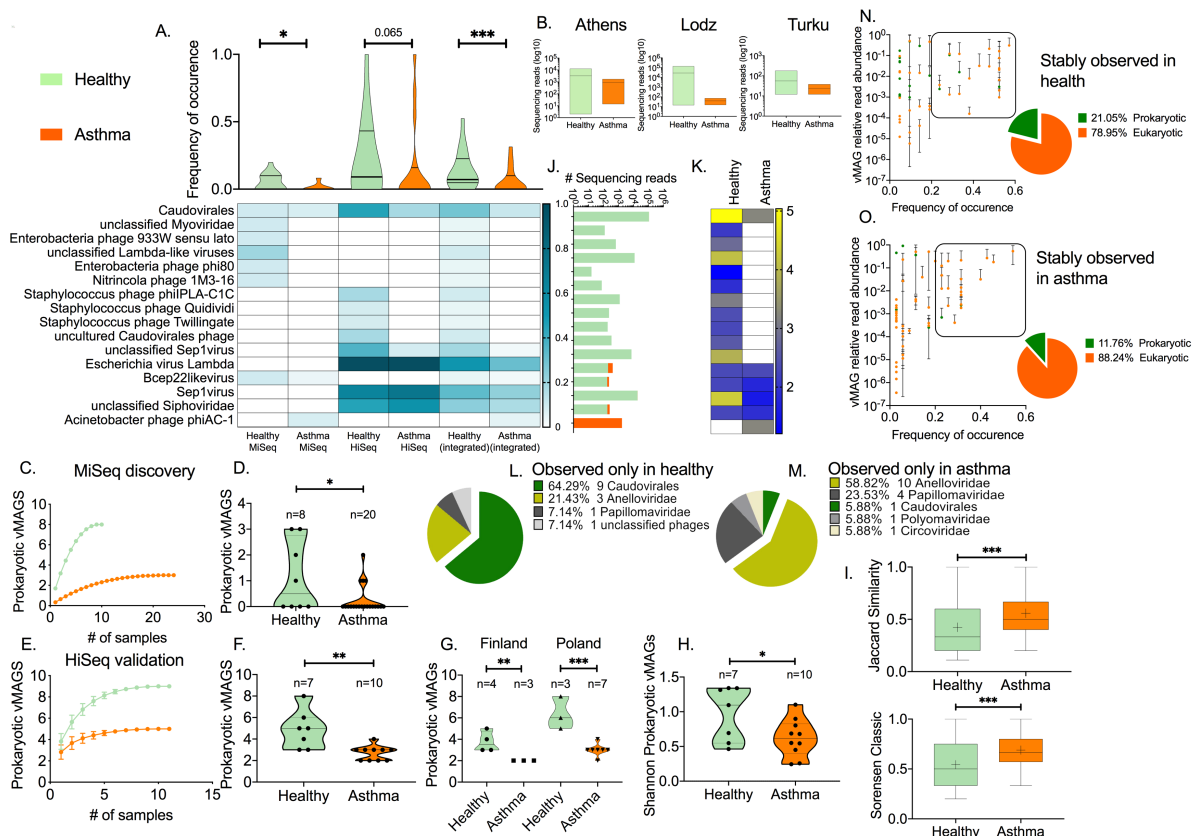


Figure 1: Prokaryotic viruses in health and asthma. (A) Heatmap of the frequency of occurrence (FrO) for each prokaryotic vMAG; Data on the discovery MiSeq and validation HiSeq cohorts are presented, along with the total vMAG FrO per clinical group (integrated data). FrO distributions are depicted as violin plots along with the median and 95% CIs. Statistical differences between viral species in health and asthma were tested using paired two-tailed t test. (B) Floating bar plots of the total number of sequencing reads that mapped to each prokaryotic vMAGs in health and asthma in Greece, Poland and Finland; Mean values and range are depicted. (C) Comparison of sample-based estimated number of prokaryotic vMAGs in health and asthma; MiSeq discovery cohort. The graph depicts the accumulation of

bacteriophage vMAGs with increasing number of samples after randomisation without replacement. Data points depict mean values calculated after 100 repetitions with standard deviations. (D) Scatter plots of the number of prokaryotic vMAGs observed in each sample in health and asthma. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test. (E) Comparison of sample-based estimated number of prokaryotic vMAGs in health and asthma; HiSeq validation cohort. Data points depict mean values calculated after 100 repetitions with standard deviations. (F) Comparison of the prokaryotic vMAG richness between healthy and asthmatic individuals; HiSeq cohort. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test. (G) Comparison of the prokaryotic vMAG richness between healthy and asthmatic individuals in Poland and Finland; HiSeq cohort. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test. (H) Comparison of the prokaryotic vMAG Shannon diversity between healthy and asthmatic individuals; HiSeq cohort. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test. (I) Prokaryotic vMAG beta diversity within health and asthma. Floating bars depict range, mean (cross) and median of the Jaccard and Sorensen indexes calculate between the samples within each group; HiSeq cohort. Statistical significance was tested using a nonparametric Mann Whitney test. (J) Stacked plot of the fraction of sequencing reads aligned onto prokaryotic vMAGs identified in health and/or asthma (Log₁₀ scale). Prokaryotic vMAGs are presented based on Figure 1A. (K) Heatmap of the total number of prokaryotic vMAG sequencing reads in health and asthma; scale (Log₁₀); gradient: yellow-high to blue-low. Prokaryotic vMAGs are presented based on Figure 1A. (L) Compositional pie charts of viral genomes observed exclusively in healthy children. (M) Compositional pie charts of viral genomes observed exclusively in children with asthma. Scatter plots of vMAG frequency of occurrence and vMAG mean relative read abundance in (N) health and (O) asthma: We defined

stably observed viruses based on their average sequencing read abundance and frequency of occurrence; the proportion of prokaryotic vMAGs within this subset was higher in health compared to asthma. Colour-coding: Orange; asthma Green; health. Significance tests: * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$, **** $p < 0.00001$.

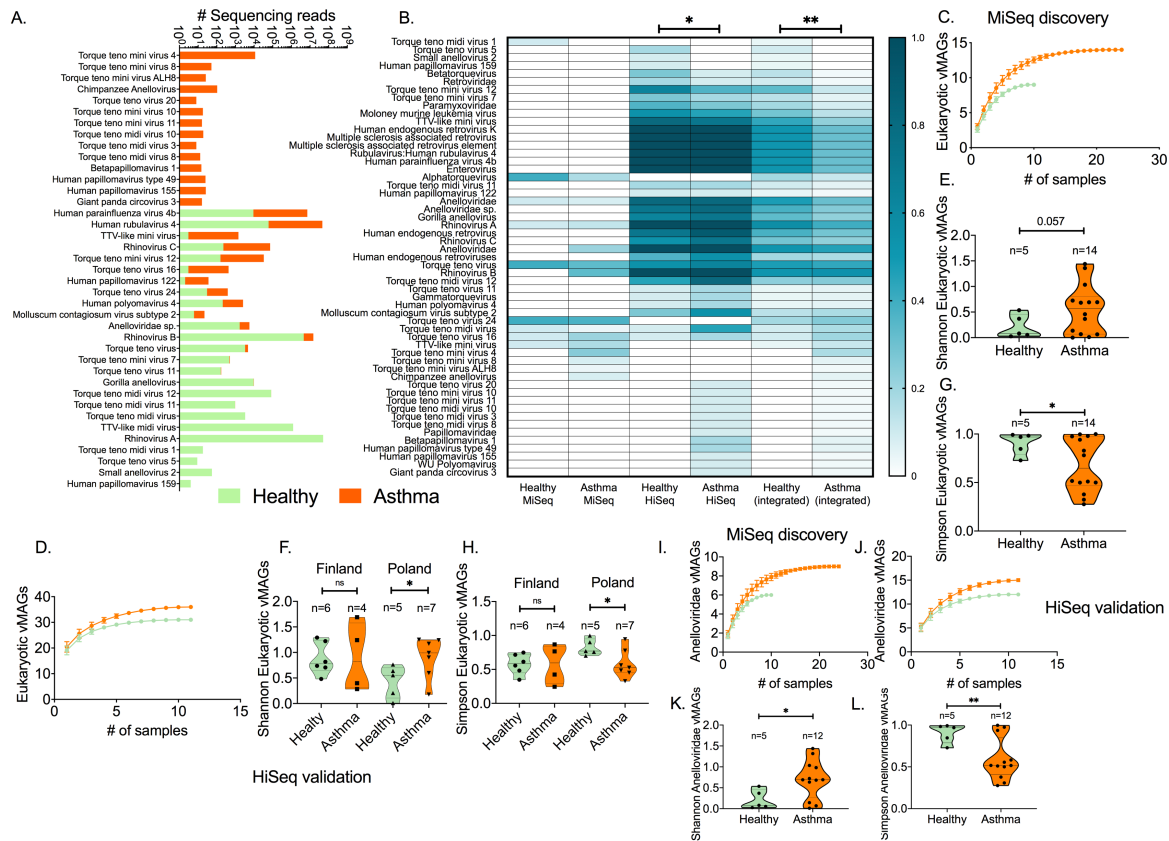


Figure 2: Eukaryotic viruses in the healthy and asthmatic virome. (A) Stacked plot of the fraction of sequencing reads aligned onto eukaryotic vMAGs identified in health and/or asthma (Log10 scale). (B) Heatmap of the frequency of occurrence (FrO) for each eukaryotic vMAG; Data on the MiSeq and HiSeq cohorts are presented, along with the FrO per clinical group (integrated dataset). Significant differences were tested using a two-tailed paired t test. Comparison of sample-based rarefaction curves of the estimated number of eukaryotic vMAGs in health and asthma; (C) MiSeq discovery, and (D) HiSeq validation cohort. Data points depict mean values calculated after 100 repetitions with standard deviations. Comparison of the eukaryotic vMAG Shannon diversity between healthy and asthmatic individuals; (E) MiSeq

discovery, and (F) HiSeq validation cohort. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test with Welch's correction. Comparison of the eukaryotic vMAG Simpson evenness between healthy and asthmatic individuals; (G) MiSeq discovery, and (H) HiSeq validation cohort. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test with Welch's correction. Comparison of sample-based rarefaction curves of the estimated number of Anelloviridae vMAGs in health and asthma; (I) MiSeq discovery, and (J) HiSeq validation cohort. Data points depict mean values calculated after 100 repetitions with standard deviations. Comparison of the Anelloviridae vMAG ecological indexes between healthy and asthmatic individuals; (K) Shannon diversity, and (L) Simpson evenness. Median and 95% CIs are depicted. Statistical significance was tested using a parametric two-tailed t test with Welch's correction. Colour-coding: Orange; asthma Green; health. Significance tests: * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$, **** $p < 0.00001$.

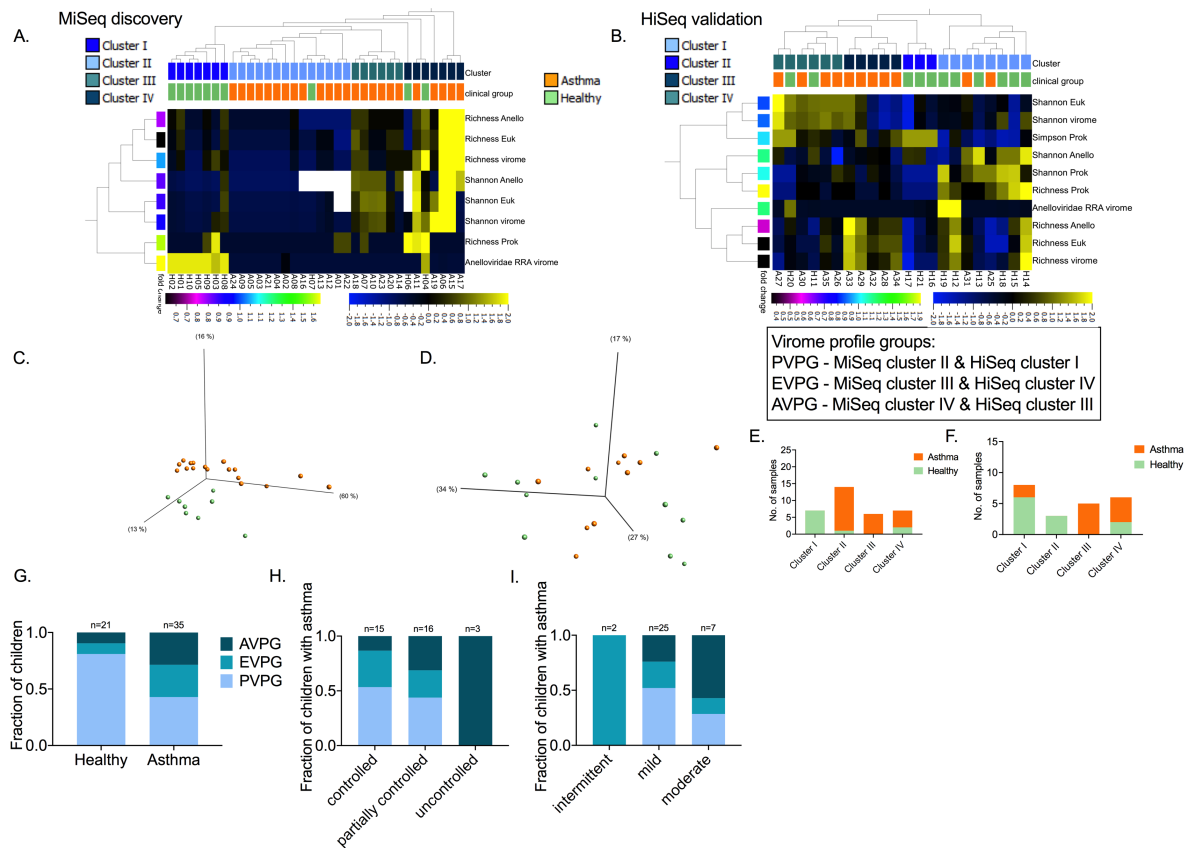


Figure 3: Virome profile groups identified in the MiSeq and HiSeq cohorts. Heat maps of the virome features that contributed to the co-clustering of healthy and asthma individuals in the (A) MiSeq cohort and (B) HiSeq cohorts; Eighteen virome features were filtered (feature reduction) based on their variance up to the level that provided the best health and asthma grouping; MiSeq: variance 0.11, projection score 0.18, and HiSeq: variance 0.04, projection score 0.13. The features that contributed to this separation are presented. All features were normalised to mean=0 and variance=1. Standardised heatmap color-coding; Yellow +2, Blue -2. First row of sample annotation; Green: healthy, Orange: asthma. Second row of sample annotation: virome clusters; sample subgroupings based on the retained virome features. The cladograms over the heat maps depict the agglomerative hierarchical clustering of individuals based on their virome characteristics (weighted linkage). Annotation of healthy and asthma individuals from the (C) MiSeq and (D) HiSeq cohorts in principal components. Stacked plots present the number and proportion of healthy and asthma individuals in each cluster identified

in the (E) MiSeq and (F) HiSeq cohorts. Based on the virome properties of the samples in each MiSeq and HiSeq cluster, three overall virome profile groups were assigned: VPG1 (PVPG), VPG2 (EVPG) and VPG3 (AVPG). (G) Stacked plots of the fractions of healthy and asthma samples in the PVPG, EVPG and AVPG virome profile groups. (H) Stacked plots of the fractions of children with asthma in PVPG, EVPG and AVPG and different levels of asthma control: controlled, partially controlled and uncontrolled asthma. All cases of uncontrolled asthma were observed in AVPG along with 31.2% of partially controlled and 13.3% of controlled asthma. On the contrary, 43.7% of partially controlled and 53.3% of controlled asthma cases were observed in PVPG. (I) Stacked plots of the fractions of children with asthma in PVPG, EVPG and AVPG and different levels of asthma severity: intermittent, mild and moderate asthma.

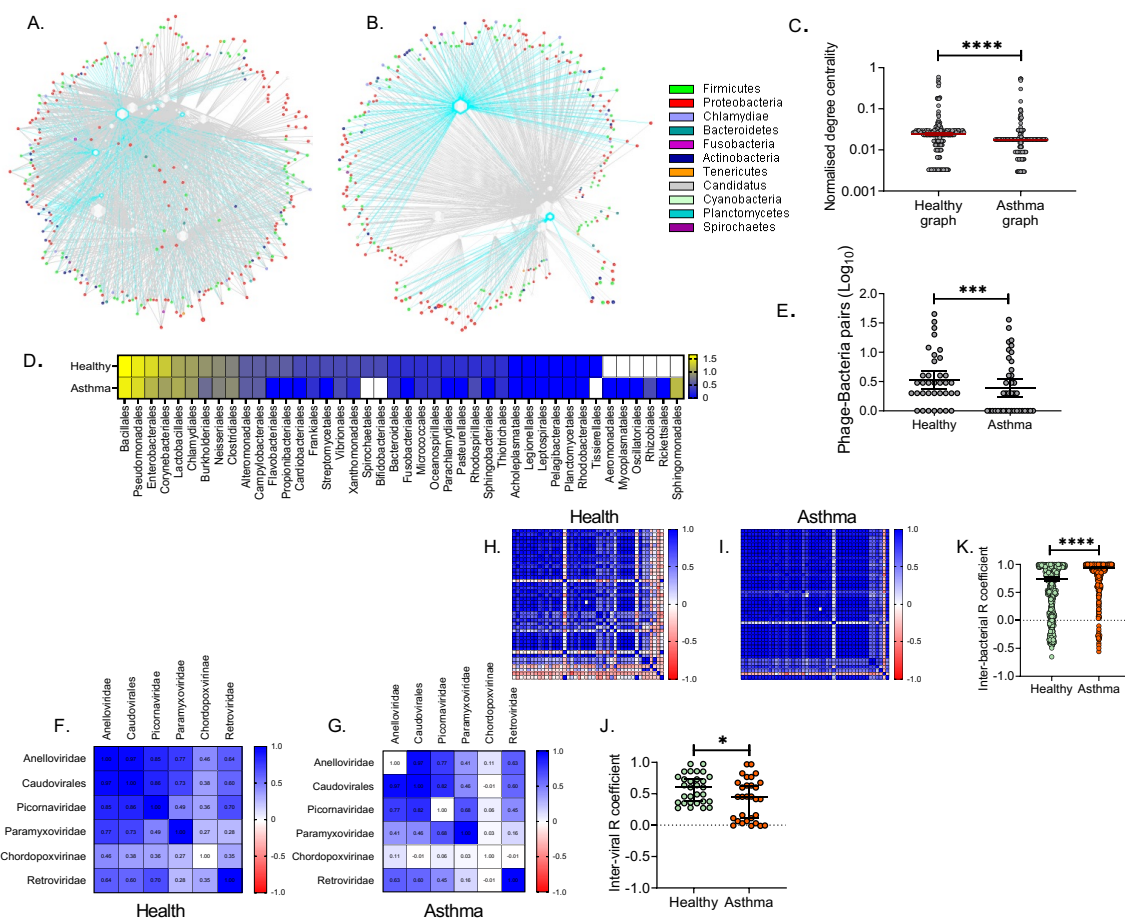
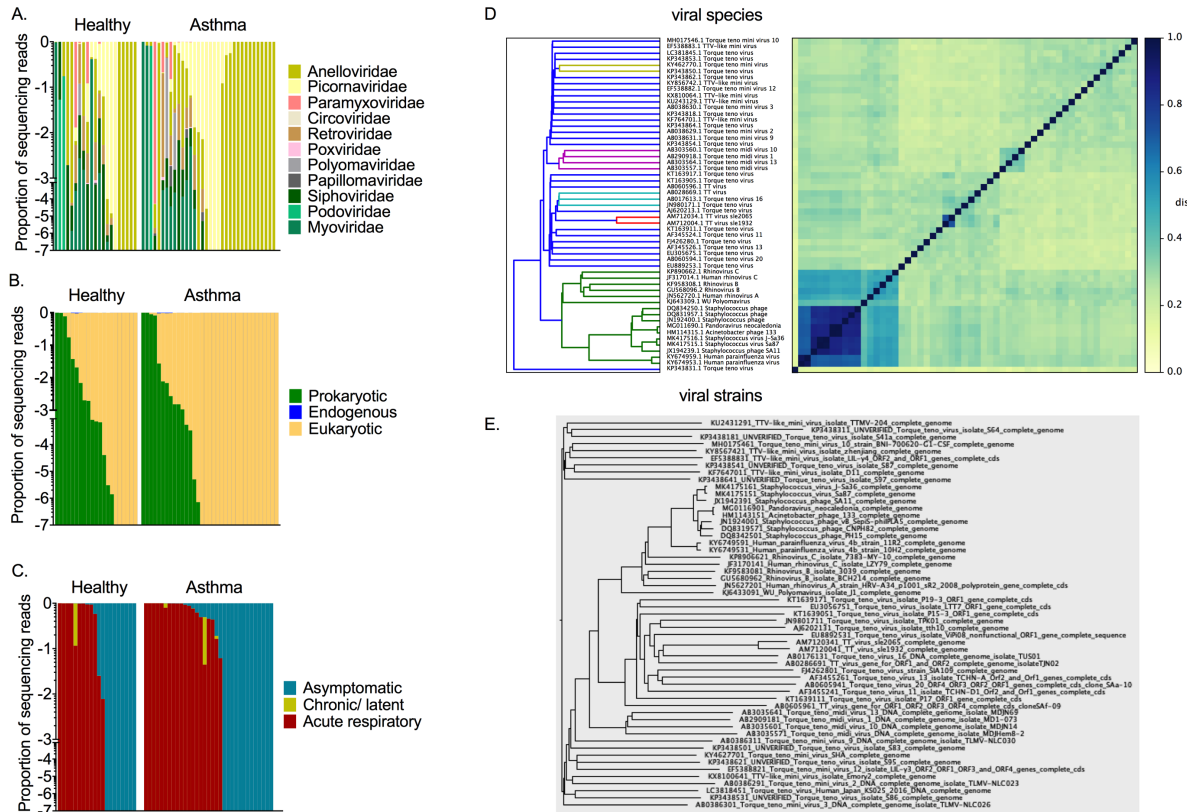
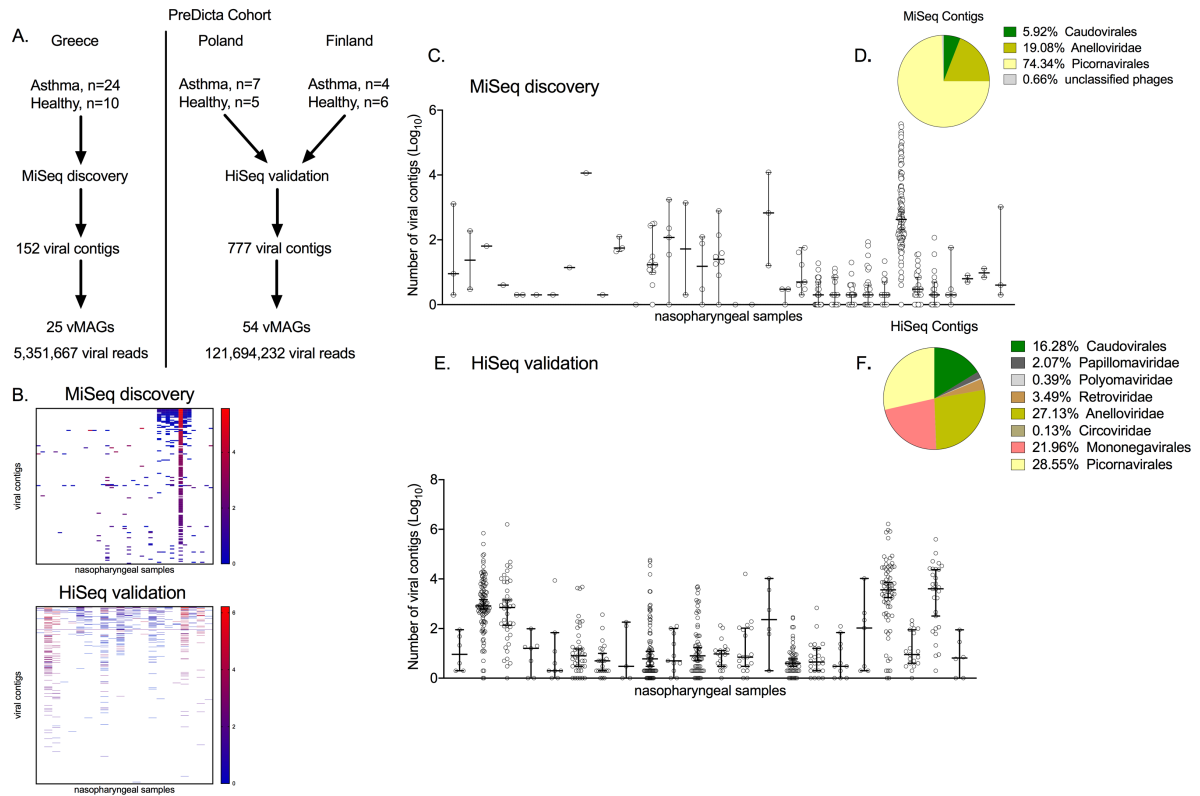


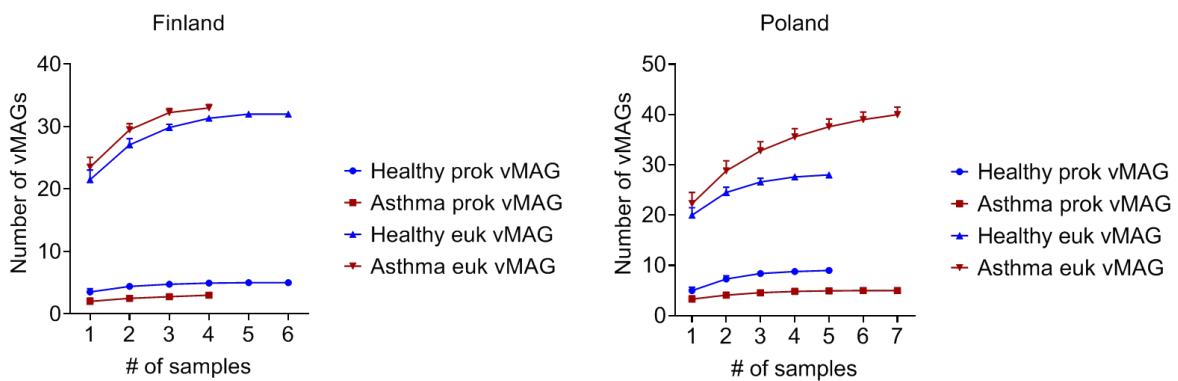
Figure 4: Viral-bacterial co-occurrence networks. (A) Healthy meta-graph, (B) Asthma meta-graph. Viruses are depicted as polygons in the centre of the graphs and bacteria as spheres in the periphery (virus concentric view). The size of each node is analogous to the normalised node degree centrality. Edges represent statistically significant correlations (Pearson test $p < 0.05$) between viral and bacterial MAGs. Caudovirales nodes are highlighted with light blue as well as their edges. (C) Comparison of normalised degree centrality between the healthy and the asthma meta-graph; Statistical significance was assessed by the Mann-Whitney nonparametric t test. (D) Heatmap showing the log₁₀ transformed number of phage-bacteria pairs in health and asthma and their distribution in bacterial orders (Yellow: high, Blue: low). (E) Comparison of the log₁₀ transformed number of phage-bacteria pairs per bacterial order; Statistical significance was assessed by the Wilcoxon matched-pairs rank test. Color-coded pairwise correlation matrix amongst viruses identified in (F) health or (G) asthma in relation to their specific bacteria pairs (High: blue, Low: red). Color-coded pairwise correlation matrix amongst bacteria identified in (H) health or (I) asthma in relation to their specific virus pairs (High: blue, Low: red). (J) Comparison of the level of virus-virus correlation in respect to their bacteria pairs within the healthy and asthmatic meta-communities; Statistical significance was assessed by the Mann-Whitney nonparametric t test. (K) Comparison of the level of bacteria-bacteria correlation in respect to their virus pairs within the healthy and asthmatic meta-communities; Statistical significance was assessed by the Mann-Whitney nonparametric t test. Significance tests: * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$, **** $p < 0.00001$. Scatter plots depict median values with 95% CIs.



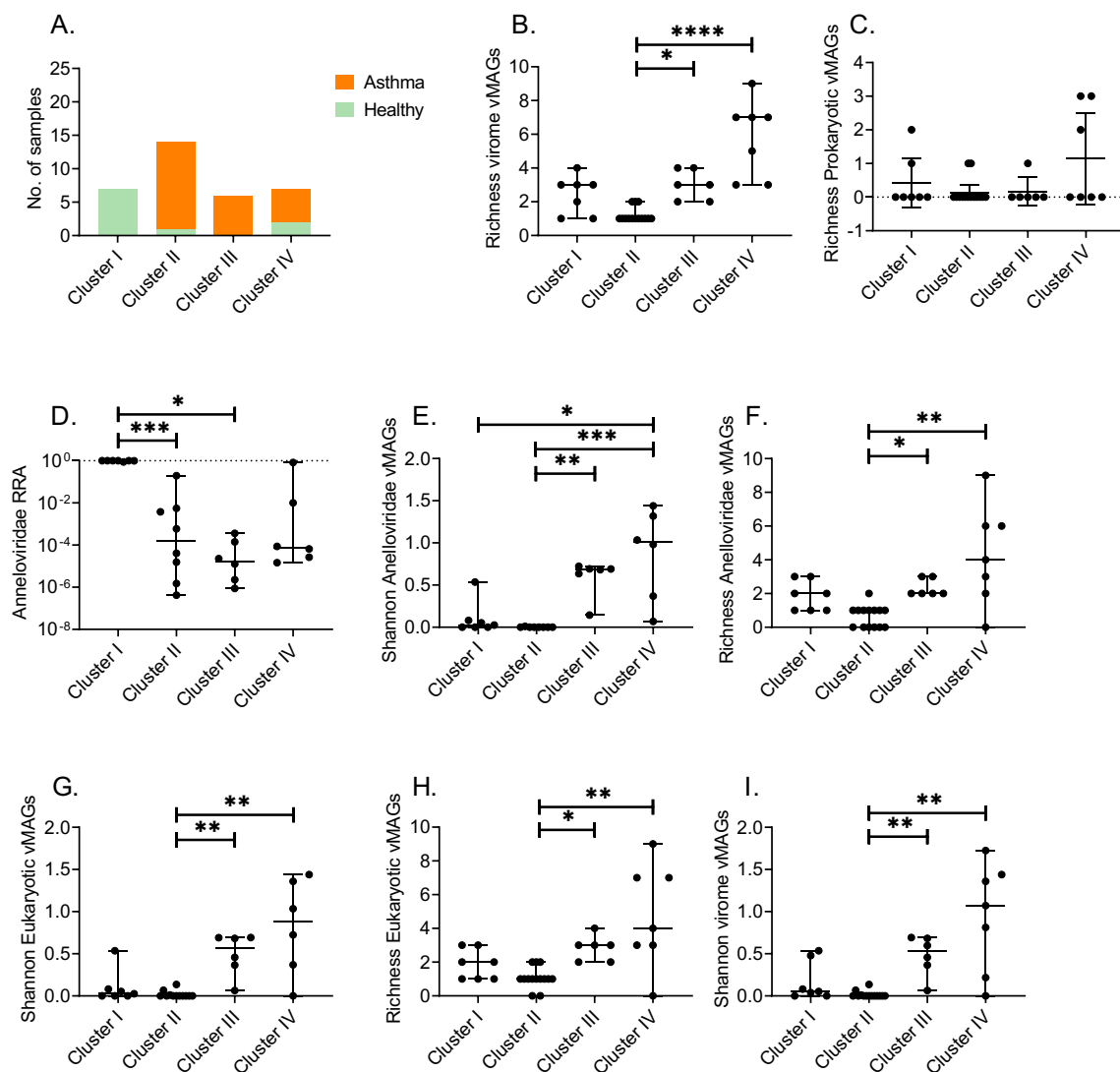
Supplementary figure 1: Virome description. Virus abundance plots per healthy and asthmatic individuals. Viruses are organised based on different properties: (A) taxonomic family, (B) host-type, and (C) type of pathogenicity. (D) Clustered Heatmap and dendrogram of 7mer MASH distances between vMAG representative genomes at the species, and (E) strain level. Nonredundant representative genomes for each vMAG were identified within the ENA Sequence Viral database using nucleotide BLAST.



Supplementary figure 2: Viral metagenome assembled genomes identified in the study. (A) Study design and metagenomic sequencing output. (B) Heat maps of the number of sequencing reads aligned into viral contigs identified in the nasopharyngeal samples of children screened with MiSeq and HiSeq; Columns: samples, rows: viral contigs; High: red, low: blue. Scatter plots of the number of different viral contigs identified in each sample sequenced in (C) MiSeq, and (E) HiSeq; Median values with 95% CIs are depicted. Pie charts of the fraction of viral contigs classified at different taxonomy levels in (D) MiSeq, and (F) HiSeq.

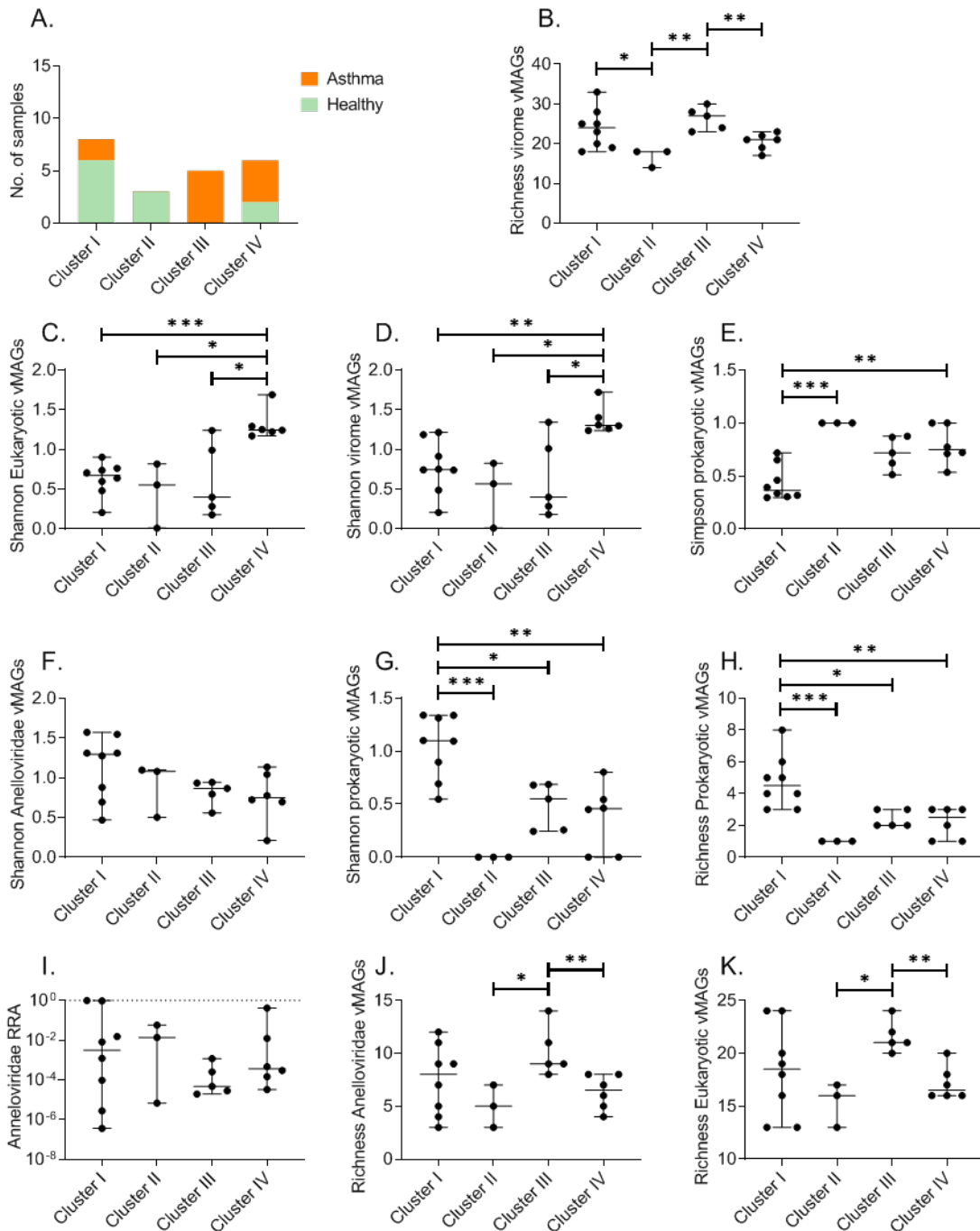


Supplementary figure 3: Estimation of viral richness in donors from Finland and Poland. The number of vMAGs was estimated using sample-based rarefaction curves without replacement (100 iterations) with increasing number of subsamplings. The plots depict the estimated number of vMAGs for the prokaryotic and eukaryotic viruses in healthy and asthma donors recruited in Turku (Finland) and Lodz (Poland). The mean and standard deviation (SD) is depicted for each data point. Dark red: Asthma, Blue: healthy.



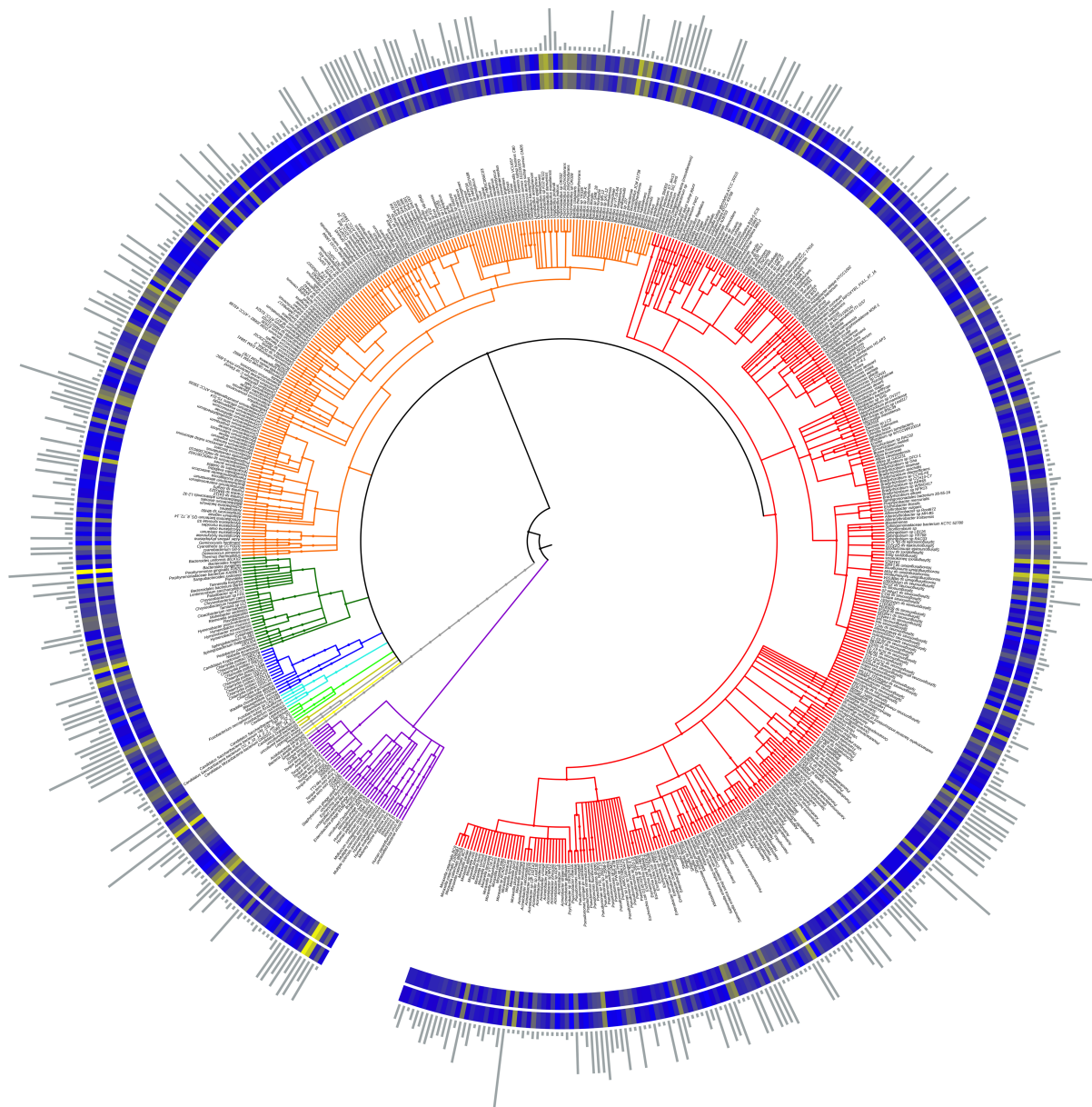
Supplementary figure 4: Clusters of donors with different virome properties identified in the MiSeq cohort. Separation of healthy and asthma donors was based on eight virome features

(Figure 4A). Based on these features 4 subgroups of donors were identified. The fraction of healthy and asthma samples in each cluster is depicted in (A). The quantitative differences between the samples of each cluster for the 8 virome features are depicted: (B) Richness of the total virome, (C) Richness of the prokaryotic virome (bacteriophages), (D) Relative read abundance of the Anelloviridae family, (E) Shannon abundance-based diversity of Anelloviruses, (F) Richness of Anelloviruses, (G) Shannon diversity of the eukaryotic virome, (H) Richness of the eukaryotic virome, and (I) Shannon diversity of the virome. Kruskal-Wallis test corrected for multiple comparisons (Dunn's test); adjusted p values are reported: Significance tests: * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$, **** $p < 0.00001$. Scatter plots depict median values with 95% CIs.



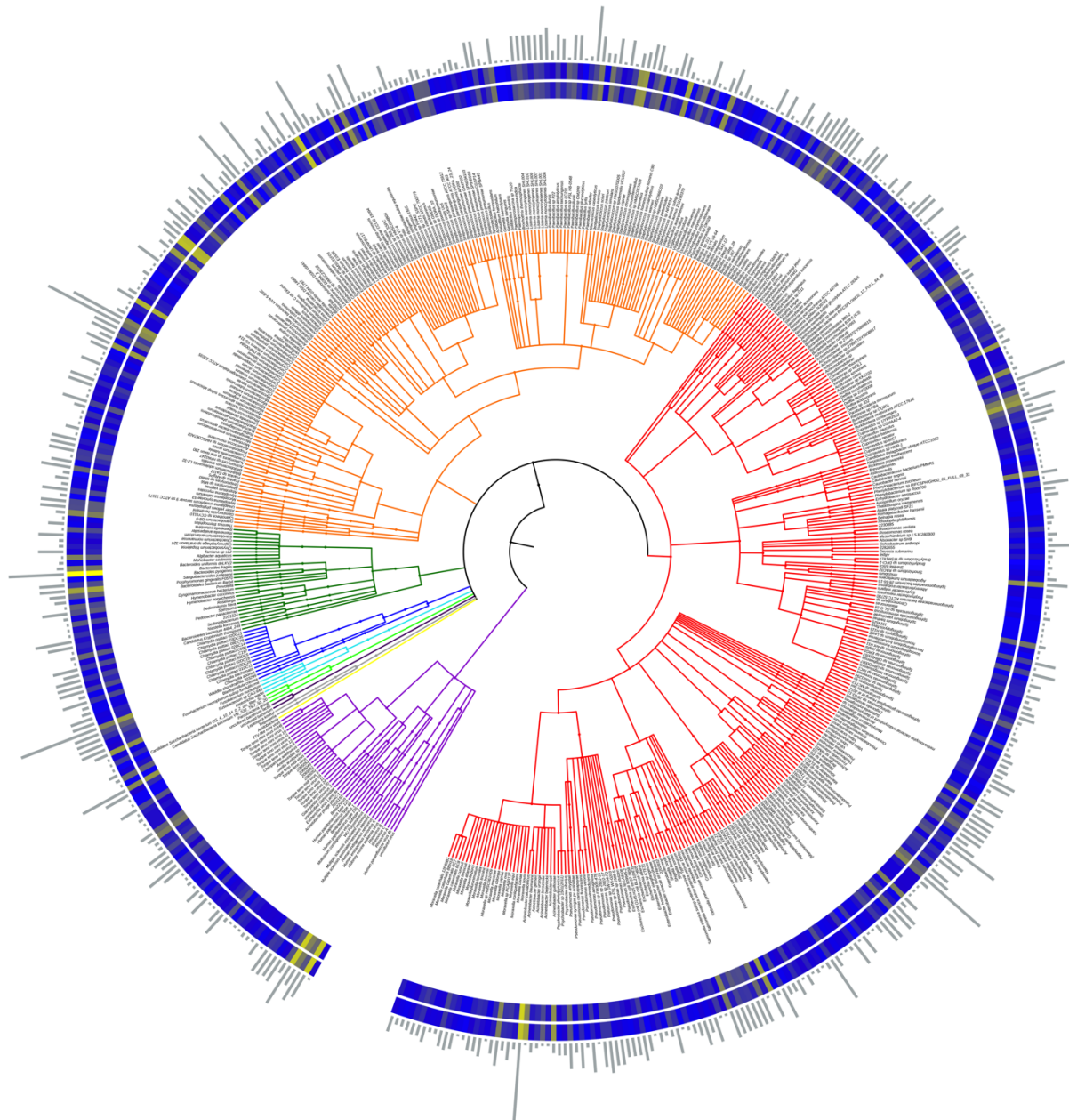
Supplementary figure 5: Clusters of donors with different virome properties identified in the HiSeq cohort. Separation of healthy and asthma donors was based on ten virome features. Based on these features 4 subgroups of donors were identified. The fraction of healthy and asthma samples in each cluster is depicted in (A). The quantitative differences between the samples of each cluster for the 10 virome features are depicted: (B) Richness of the total

virome, (C) Shannon abundance-based diversity of the eukaryotic virome, (D) Shannon diversity of the virome, (E) Simpson abundance-based evenness of the prokaryotic virome, (F) Shannon diversity of Anelloviruses, (G) Shannon diversity of the prokaryotic virome, (H) Richness of the prokaryotic virome, (I) Relative read abundance of Anelloviruses, (J) Richness of Anelloviruses, and (K) Richness of the eukaryotic virome. (B, C, D, J, K) Brown-Forsythe and Welch ANOVA test corrected for multiple comparisons by controlling the False Discovery Rate (FDR). (E, F, G, H, I) Kruskal-Wallis test corrected for multiple comparisons (Dunn's test). Adjusted p values are reported: Significance tests: * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$, **** $p < 0.00001$. Scatter plots depict median values with 95% CIs.



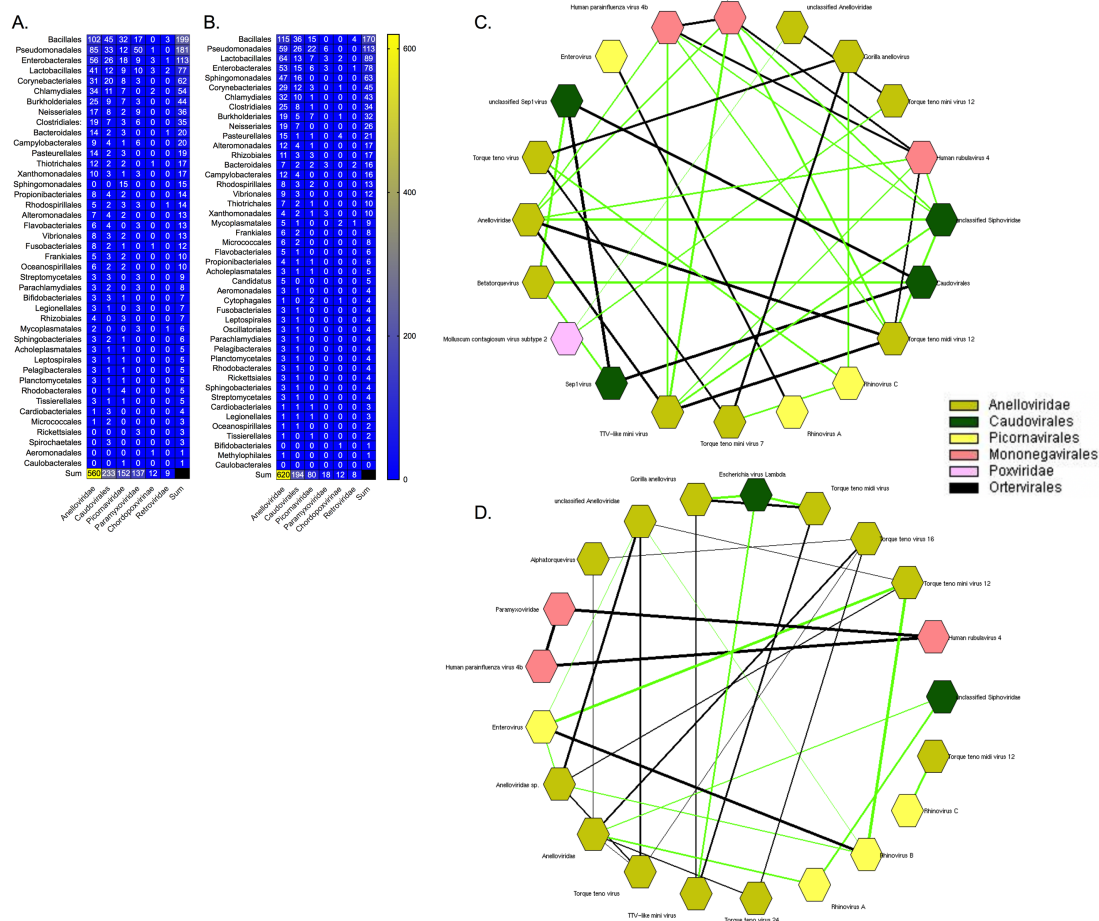
Supplementary figure 6: Respiratory metagenome assembled genomes (MAGs) of healthy children. MAGs are organised into hierarchical clusters based on their taxonomic similarity

(taxonomic lineage up to species level). A total of 813 (95%CI: 141-241) microbial taxa were identified in healthy children; 757 out of 813 taxa belonged to bacteria and 56 to viruses. We observed 236 (29%) microbial taxa, including 14 viral taxa, present exclusively in the healthy group. About 369,655,276 (mean log₁₀: 6.610±0.972, n=21) sequencing reads aligned to microbial MAGs isolated from healthy samples. The mean and total relative read abundance of each MAG are depicted as colour gradient concentric circles around the cladogram (log₁₀ transformed), respectively. The frequency of occurrence for each lineage (up to species level) is visualised as bar plots in the outer ring of the circular cladogram. Cladogram colouring: Purple; Viruses, Red; Proteobacteria, Orange; Firmicutes, Green; Bacteroidetes, Dark Blue; Chlamydiae, Light Blue; Fusobacteria, Light Green; Candidatus bacteria, Gold; Unclassified bacteria, Light Yellow; Acidobacteria



Supplementary figure 7: Respiratory metagenome assembled genomes (MAGs) in children with asthma. MAGs are organised into hierarchical clusters based on their taxonomic similarity

(taxonomic lineage up to species level). A total of 722 (95%CI: 102-177) microbial taxa were identified in asthma; 662 out of 722 taxa belonged to bacteria and 60 taxa to viruses. We observed 144 (20%) taxa, including 18 viral taxa, unique for asthma patients. 490,164,409 (mean log₁₀: 6.273±1.164, n=35) sequencing reads aligned to MAGs isolated from asthma samples. The mean and total relative read abundance of each MAG is viewed as colour gradient concentric circles around the cladogram (log₁₀ transformed), respectively. The frequency of occurrence for each lineage (up to species level) is visualised as bar plots in the outer ring of the circular cladogram. Purple; Viruses, Red; Proteobacteria, Orange; Firmicutes, Green; Bacteroidetes, Dark Blue; Chlamydiae, Light Blue; Fusobacteria, Light Green; Candidatus bacteria, Gold; Unclassified bacteria, Light Yellow; Acidobacteria.



Supplementary figure 8: Viral-bacterial and viral-viral co-occurrence pairs. Total number of edges (pairs) between viral families and bacterial orders in the (A) healthy and (B) asthma meta-graphs. Both bacteria and viruses are arranged according to the number of total cross-kingdom connections starting from the one with the highest number (high hubness) towards the one with the lowest (low hubness). Heat map colours; Red: high, Green: low. Virus-virus co-occurrence in the healthy and asthmatic virome: Viral taxa that significantly co-occurred (Pearson $p < 0.05$) are presented as a circular co-occurrence network. All correlations were positive with (C) Pearson $R > 0.666$ in the healthy and (D) $R > 0.617$ in the asthmatic graph. The level of correlation is analogous to the width of the edges of the network. Edges representing correlations amongst taxa from different viral families are highlighted with green. Viral MAGs are colour-coded; Gold: Anelloviridae, Caudovirales: Green, Picornavirales: Light yellow, Mononegavirales: Pink, Poxviridae: Light pink, Black: Ortervirales.