

RESEARCH

Open Access

Seasonal dynamics of DNA and RNA viral bioaerosol communities in a daycare center



Aaron J. Prussin II^{1*†} , Pedro J. Torres^{2†}, John Shimashita³, Steven R. Head³, Kyle J. Bibby⁴, Scott T. Kelley² and Linsey C. Marr¹

Abstract

Background: Viruses play an important role in ecosystems, including the built environment (BE). While numerous studies have characterized bacterial and fungal microbiomes in the BE, few have focused on the viral microbiome (virome). Longitudinal microbiome studies provide insight into the stability and dynamics of microbial communities; however, few such studies exist for the microbiome of the BE, and most have focused on bacteria. Here, we present a longitudinal, metagenomic-based analysis of the airborne DNA and RNA virome of a children's daycare center. Specifically, we investigate how the airborne virome varies as a function of season and human occupancy, and we identify possible sources of the viruses and their hosts, mainly humans, animals, plants, and insects.

Results: Season strongly influenced the airborne viral community composition, and a single sample collected when the daycare center was unoccupied suggested that occupancy also influenced the community. The pattern of influence differed between DNA and RNA viromes. Human-associated viruses were much more diverse and dominant in the winter, while the summertime virome contained a high relative proportion and diversity of plant-associated viruses.

Conclusions: This airborne microbiome in this building exhibited seasonality in its viral community but not its bacterial community. Human occupancy influenced both types of communities. By adding new data about the viral microbiome to complement burgeoning information about the bacterial and fungal microbiomes, this study contributes to a more complete understanding of the airborne microbiome.

Keywords: Virome, Metagenomics, Built environment, Bioaerosol

Background

There is an emerging emphasis on understanding the complex interactions among humans, the built environment (BE), and the microbiomes of both because of their effects on human health [1]. This thrust has been facilitated by applications for molecular methods that have dramatically expanded our understanding of the diversity and dynamism of microbes [2–4]. While numerous studies have characterized bacterial and fungal microbiomes in the BE [2, 5–9], far fewer have focused on the viral microbiome (virome). Williams [10] described viruses as “the forgotten siblings of the microbiome family” and argued that the human virome is probably at

least as important for human health as is the “bacteriome.” Viruses play critical roles in many ecosystems. For instance, many bacteriophages act as predators to control bacterial populations [11–14], while others determine bacterial pathogenicity by directly inserting pathogenicity islands and virulence factors into bacterial genomes [15]. Numerous eukaryotic viruses impact human health, and some can spread rapidly through the BE as airborne particles, known as “aerosols” [16–20]. We have shown previously that viruses in indoor air are as numerous as bacteria [21].

One reason for the paucity of data on the airborne virome is the inherent technical difficulty of studying it [22]. Some of the challenges of studying airborne virome include no conserved viral gene for marker studies (e.g., 16S for bacteria and ITS/18S for fungi), sampling challenges particularly for aerosols, low amounts of biomass in the air, limited representation of viruses in reference

* Correspondence: aprussin@vt.edu

†Aaron J. Prussin II and Pedro J. Torres contributed equally to this work.

¹Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA

Full list of author information is available at the end of the article



databases, and the often short-term persistence of viruses in the environment, especially RNA viruses [22]. While several studies have investigated the spread of pathogenic viruses (e.g., norovirus and rotavirus) on surfaces in the BE [23–25], few have attempted to apply broad-scale deep-sequencing approaches to analyzing entire viromes in the BE [26].

As a result, studies on the virome of the BE and/or the airborne virome have been sparse. One study that examined the DNA virome on restroom surfaces reported that enterophages, human papilloma virus, and herpesviruses were the most abundant viruses [26]. We are aware of five studies that have characterized the airborne virome [27–31]. Three of these (Hall et al. [28], Brisebois et al. [30], and Rosario et al. [31]) included samples from the BE, and none examined longitudinal effects. Longitudinal data are important because they can provide insight into the stability and dynamics of microbial communities. Longitudinal studies can also reveal temporal or seasonal trends while assisting with causal inference (e.g., viruses associated with certain seasonal diseases are more prevalent in the winter).

Results of studies on the bacterial and fungal communities indoors offer clues but not a complete picture about the dynamics of the airborne virome. Seasonal variations in the bacterial community in outdoor air [32] can theoretically propagate to indoor air, as bioaerosols are readily exchanged between the two volumes [33]. A study in Finland observed seasonal patterns of bacterial diversity in indoor dust [34], while one in the San Francisco Bay Area found only weak evidence of seasonality in bacterial diversity [35]. In both cases, differences were greater between buildings or even between rooms within the same building than between seasons, indicating that building characteristics and/or occupants are more important than season in shaping the bacterial microbiome. In terms of BE fungal communities, Adams et al. [35] found that fungal diversity in settled dust was strongly influenced by season and appeared to be dominated by outdoor sources. These studies have shown that a combination of outdoor air, occupants, and building characteristics contributes to shaping the microbiome of the BE [2, 36].

The seasonality of human-associated viruses, such as influenza virus and rotavirus [37–39], suggests that there may also be seasonal patterns in at least part of the virome in the BE. Viruses that infect bacteria, known as bacteriophage or phage, are primary drivers of the structure and function of bacterial communities in many environments [13, 14, 40]. Given that seasonal trends have been observed in the bacterial microbiome in some BEs, it is plausible their phage predators exhibit parallel seasonal dynamics.

Because of the long hours that some children spend at daycare centers, these buildings are of special interest in

understanding linkages between the microbiome of the BE and human health. Furthermore, daycare centers could potentially harbor a large diversity and number of viruses, thanks in part to high occupancy by young children who have yet to develop hygienic behaviors [41, 42]. Daycare centers also represent an ideal test environment to understand the spread of viral diseases in the classrooms of schools or universities. Thus, daycare centers are an excellent model system in which to study both the diversity of airborne viruses and their possible seasonality. In our prior study of the seasonal dynamics of airborne bacteria in a daycare center, we found no significant seasonal differences in community structure by season [43]. Herein, we analyze samples that were collected continuously over the course of a year in a daycare center to test the hypothesis that the airborne viral microbiome is influenced by both seasonality and human occupancy. We extracted DNA and RNA from the samples and performed deep-sequencing and a metagenomics analysis. We found that both season and human occupancy had a strong influence on the DNA and RNA viral community composition, as well as specific viral taxa. Furthermore, human-associated viruses were more abundant in the winter while plant-associated viruses were more abundant in the summer. These results complement knowledge about airborne bacterial and fungal microbiomes and contribute to a more complete understanding of the airborne microbiome in the BE.

Results

Most abundant viruses in a daycare center

To characterize the airborne viral community in a daycare center, we collected filter samples from the heating, ventilation, and air conditioning (HVAC) system between January 2014 and February 2015, as previously described [43]. We analyzed a total of three ~1-month-long samples from each of spring, summer, and fall and four samples from winter. Additionally, we collected a sample when the daycare center was closed in late December and early January. The daycare center only closes for an extended period of time once per year, and thus, we were only able to collect one sample while it was closed in this study. We also analyzed two negative controls (unexposed filter and molecular biology grade water, for DNA and RNA samples), which showed an extremely low total BLAST hit output of DNA (mean = 100) and RNA (mean = 2) viral hits (Additional file 1: Figure S1 a-b) compared to the daycare samples (DNA mean = 3892 and RNA mean = 2320).

According to the mean relative abundance of all viruses, the top 20 most abundant DNA viruses belonged to six different taxonomic families (summarized to family level when possible) (Fig. 1a). One of the six viral families was Herpesviridae, a human-associated virus that made up 7.8% of the viruses in the spring, 6.1% in the

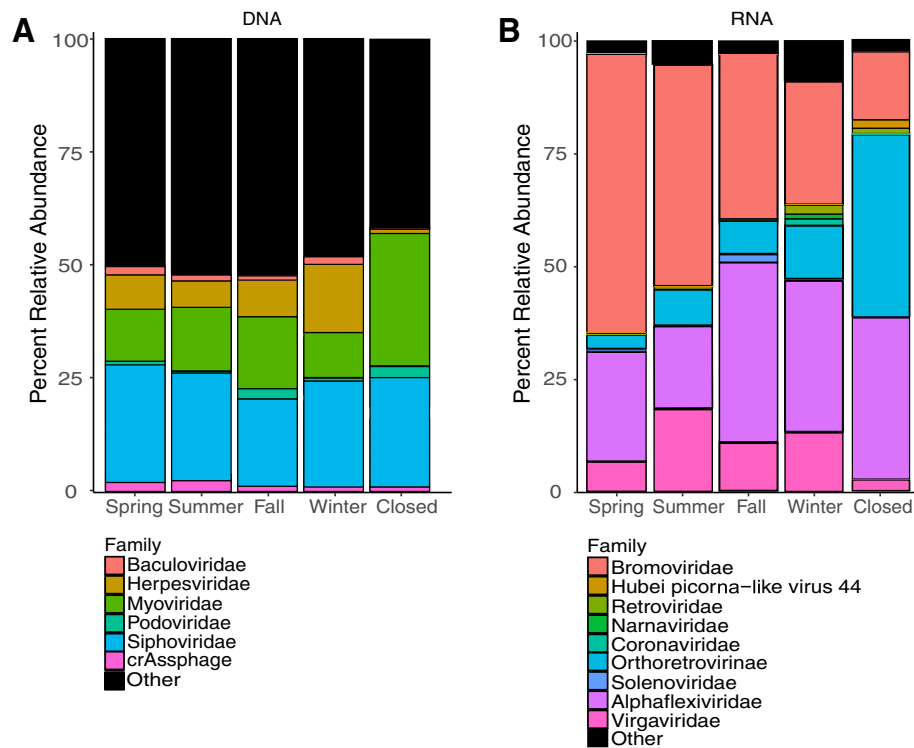


Fig. 1 Mean relative abundance of airborne DNA and RNA viruses at the level of family by season. Stacked barplot showing the percent relative abundance of the top 20 most abundant **a** DNA and **b** RNA airborne viruses at the family level, when possible, summarized by season (spring, $n = 3$; summer, $n = 3$; fall, $n = 3$; winter, $n = 4$; closed $n = 1$)

summer, 8.3% in the fall, 15.4% in the winter, and 1.2% when the facility was closed. An arthropod-associated virus, Baculoviridae, comprised 2.1% of the viral sequences in the spring, 1.6% in the summer, 1.1% in the fall, 1.9% in the winter, and 0.6% when the daycare center was closed. Four out of six viral families were bacteriophages, including Myoviridae, Podoviridae, Siphoviridae, and crAssphage. Together, these bacteriophage families represented 42% of the identified viral sequences in the spring, 43% in the summer, 41% in the fall, 37% in the winter, and 59% when the daycare center was closed.

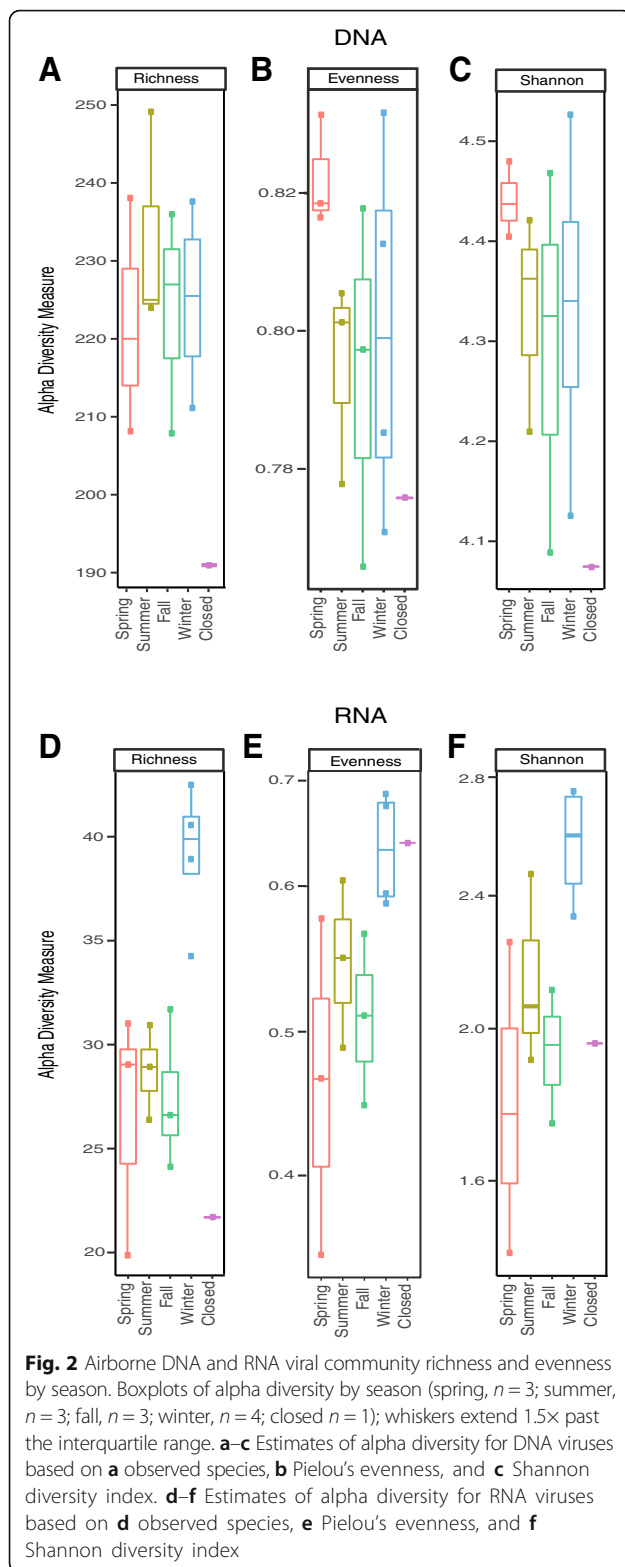
According to the mean relative abundance of all viruses, the top 20 most abundant RNA viruses belonged to nine different families (summarized to family level when possible) (Fig. 1b). Four out of nine viral families were plant-associated: Bromoviridae, Solemoviridae, Alphaflexiviridae, and Virgaviridae. Together, these viruses comprised a mean of 94% of the total RNA viruses in the spring, 86% in the summer, 89% in the fall, 74% in the winter, and 54% when the daycare center was closed. One belonged to the family Retroviridae which was largely composed of the human endogenous retrovirus, which represented 0.33% of the total in the spring, 0.65% in the summer, 0.22% in the fall, 2.1% in the winter, and 1.2% when the daycare center was closed. The Hubei-picorna-like virus (arthropod host) represented <

0.2% of the total throughout all seasons except when the daycare was closed (1.9%). The Orthoretrovirinae family represented 3% of the viruses in the spring, 7.8% in the summer, 7.2% in the fall, 12.1% in the winter, and 40.5% when the daycare center was closed. The Narnaviridae family (fungal host) represented 0.18% of the viruses in the spring, 0.13% in the summer, 0.18% in fall, 1% in the winter, and 0% when the daycare center was closed. Both animals and humans serve as a natural host to the Coronaviridae family, which represented 0% of the viruses in the spring, summer, and fall; 1.14% of the viruses in the winter; and 0.23% when the daycare center was closed.

Although this study did not focus on pathogens, we searched for influenza virus and other common pathogens (e.g., adenovirus, RSV, rotavirus). These accounted for <0.005% of the total relative abundance in all seasons.

Alpha diversity

We investigated how the number of viral species (richness) and their distribution (evenness) in the indoor air virome varied with season and human occupancy by using the number of unique BLAST hits in each sample as an estimate of species abundance for DNA and RNA viruses (Fig. 2). DNA communities tended to have a higher richness as estimated by observed species in the summer

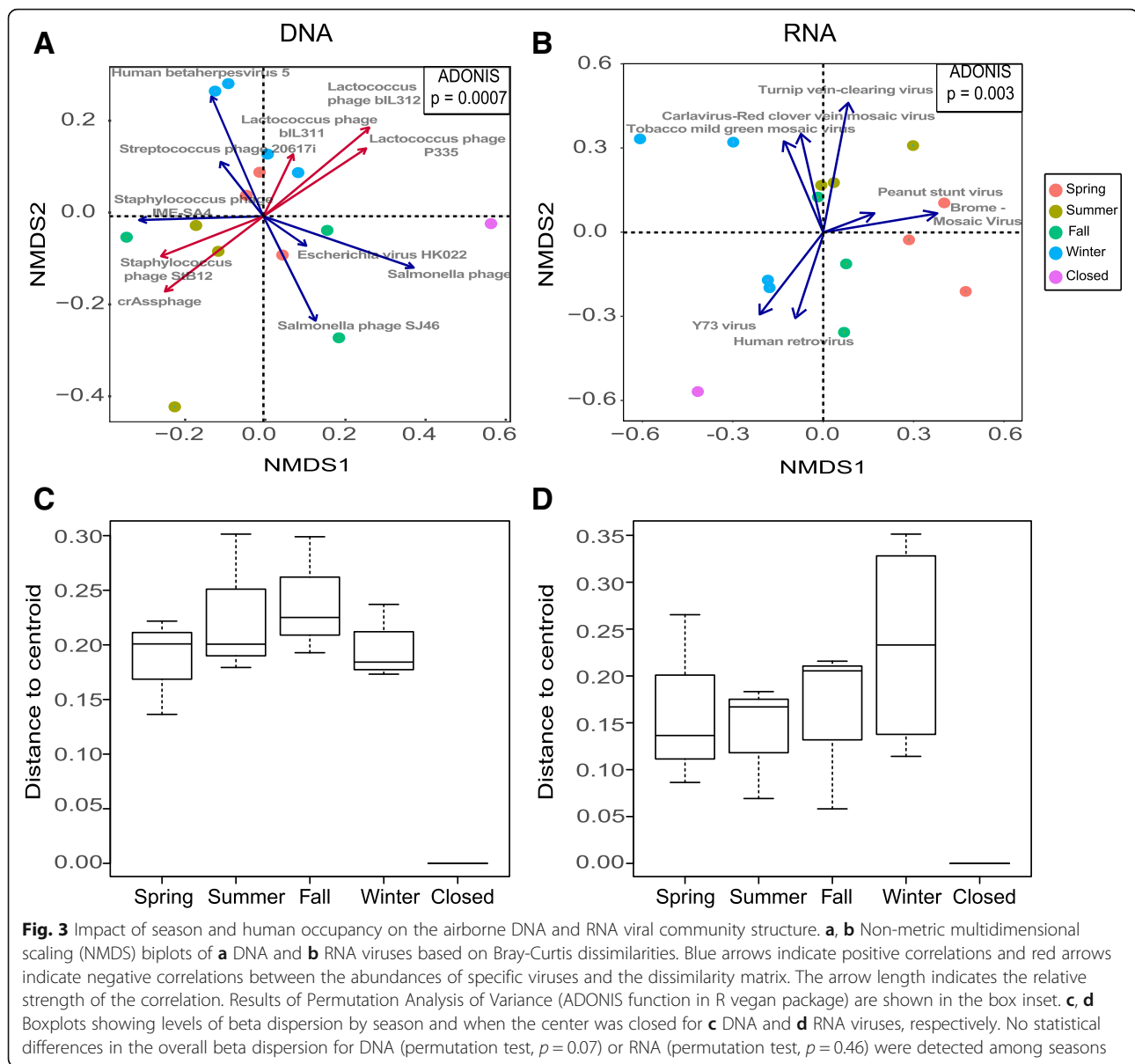


(Fig. 2a), while the evenness of the viral community was higher in the spring as measured by the Pielou and Shannon indices (Fig. 2c, d), although this difference was not significant. Notably, the DNA communities exhibited the

lowest diversity and evenness when the daycare center was closed (Fig. 2a–c). Seasonality had a more pronounced effect on the diversity of RNA communities. RNA communities tended to have a lower richness and evenness in the spring and higher richness and evenness in the winter (Fig. 2d–f). While not as noticeable as in the DNA communities, there was reduced diversity when the daycare center was closed (Fig. 2d). Unlike the DNA viral communities, RNA viral communities showed very high evenness when the daycare center was closed (Fig. 2e, f).

Beta diversity

We used non-metric multidimensional scaling (NMDS) plots based on Bray-Curtis dissimilarity to compare the similarity of the airborne DNA and RNA viral community composition between different seasons. There was noticeable clustering between samples from different seasons in the DNA (Fig. 3a) and RNA viral communities (Fig. 3b). Permutation Multivariate Analysis of Variance (PERMANOVA - ADONIS function in R vegan package) showed that season had a significant effect on the viral community structure for both DNA ($p = 0.0007$) and RNA ($p = 0.003$) viruses. We then identified species that best described the overall biological pattern of the dissimilarity matrix and fitted them to the existing NMDS plot, forming a biplot. Arrows were colored to indicate positive correlation (the highest relative abundance in that direction - blue) and negative correlation (the lowest relative abundance in that direction - red). Through our analysis, we found that viral diversity was strongly affected by season. Human-associated viruses were relatively more abundant in the winter (Fig. 3a), while plant pathogenic viruses were relatively more abundant in the spring, summer, and fall (Fig. 3b). Among DNA viruses, Lactococcus phage, Human betaherpesvirus 5 (Cytomegalovirus), and Streptococcus phage increased in relative abundance in the winter; Staphylococcus phage and crAssphage decreased in relative abundance during the fall; and Escherichia virus and Salmonella phage increased in relative abundance in the summer and fall (Fig. 3a). Among RNA viruses, Turnip vein-clearing virus, Brome mosaic virus, and peanut stunt virus increased in relative abundance in the spring and summer (Fig. 3b). There were some viruses whose correlation did not point to a particular cluster such as the Tobacco mild green mosaic virus and Carlavirus-Red clover vein mosaic virus, which seemed to relatively increase in between the winter and summer season (closer to summer). The Y73 virus and Human retrovirus relatively increased around the winter season and when the daycare center was closed. Confidence



in differences when the daycare center was closed is limited because we had only one such sample.

We also examined the variability in community structure (assessed by the analysis of beta dispersion) of DNA and RNA viruses within each individual season (Fig. 3c, d). Interestingly, DNA and RNA communities showed different dispersion trends based on the season. For DNA viruses, the variability of the within-season community structure increased from spring to fall and then fell in the winter. For RNA viruses, the variability of the within-season community structure was lowest in the summer and highest in the winter. Though there were modest differences in the beta dispersion, we did not detect statistically significant heterogeneity in the DNA (permutation

test, $p=0.07$) or RNA (permutation test, $p=0.28$) beta dispersion. This indicates that there is an effect of season on viral communities, and these communities display similar community variance.

Most important viruses driving seasonality

We used a random forest machine learning classifier to determine how well the season could be predicted based on the viral hits. The random forest model predicted the season with 62% accuracy for DNA viruses and 58% for RNA viruses. Variable importance by a mean decrease in accuracy was then calculated for the random forest model. Figure 4 and Fig. 5 illustrate 18 DNA and RNA viruses, respectively, whose removal caused the greatest decrease in model accuracy. Most DNA viruses

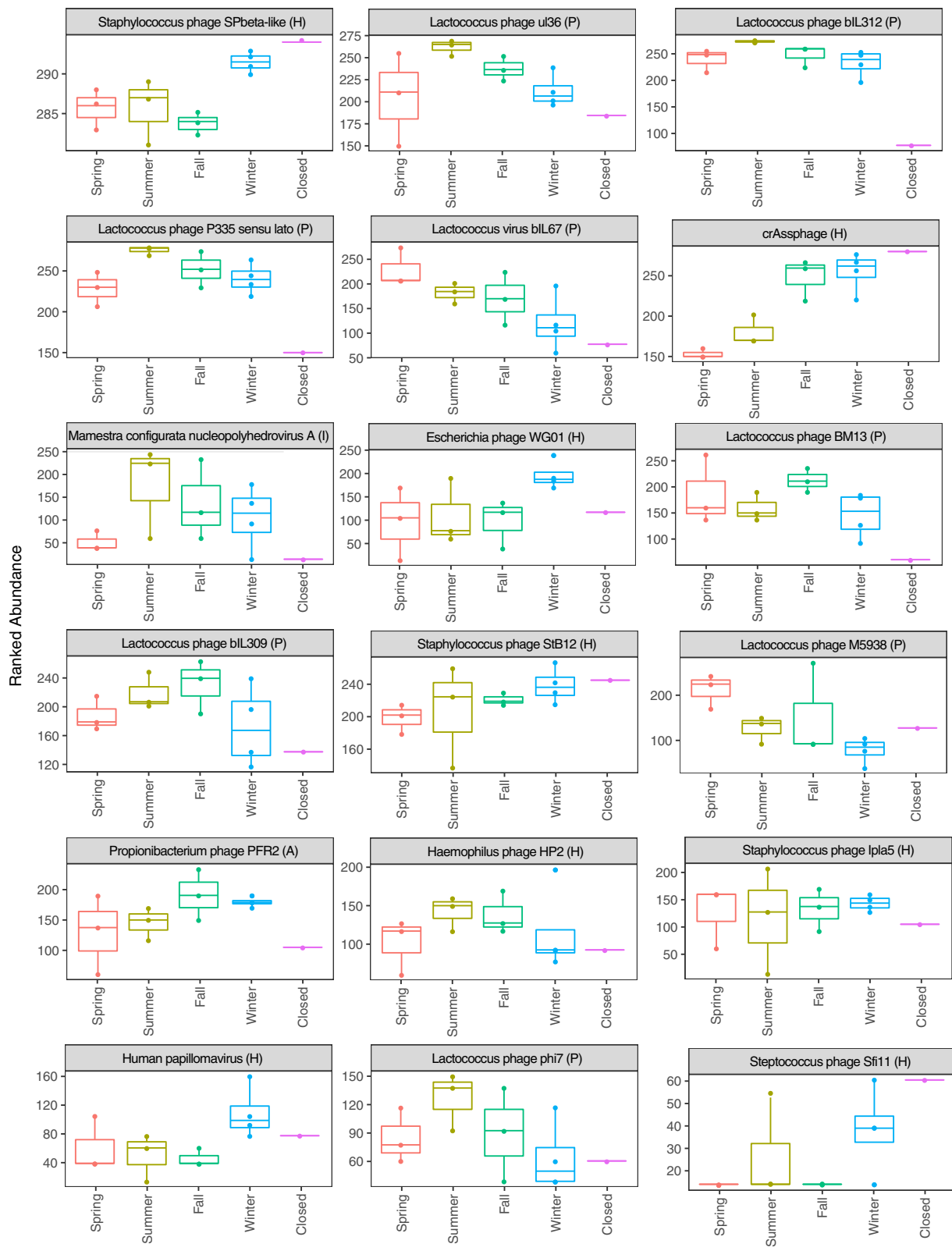


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Ranked abundances of DNA viral species that had the strongest seasonality. Random forest modeling was used to identify viruses that best distinguish between the different seasons: spring, summer, fall, and winter and when the daycare center was closed. The ranked abundance of the top 18 most discriminant DNA viruses is shown identified down to the strain level, if possible. *P* = plant-associated virus, *H* = human-associated virus, *A* = animal-associated virus, *I* = insect-associated virus, and *S/OP* = soil-associated/opportunistic pathogen

identified showed clear seasonal trends in the normalized ranked abundance. Out of the 18 viral DNA strains, eight were plant-associated, eight were human-associated, one was animal-associated, and one was insect-associated (respectively denoted by the *P*, *H*, *A*, and *I* next to the virus name in Fig. 4). Seven of the eight plant-associated viruses showed higher abundances on average in the summer when compared to the winter, while seven out of the eight human-associated viruses had higher numbers on average in the winter when compared to the summer. Interestingly, all viruses that displayed high abundances when the daycare center was closed were human-associated viruses. The *Staphylococcus* phage SPbeta-like, which is associated with human skin, displayed the highest relative abundance of all the viruses, likely due to constant shedding of skin of the human occupants.

RNA viruses also showed clear seasonal trends in their normalized relative abundance. Out of the 18 viral RNA strains, 10 were plant-associated, four were animal-associated, three were human-associated, and one was fungal-associated (respectively denoted by the *P*, *A*, *H*, and *F* next to the virus name Fig. 5). Five of the 10 plant-associated viruses showed higher abundances in the summer than the winter, and two were higher in the winter (clover yellow mosaic and tobacco green mosaic virus). The three human-associated viruses were more abundant in the winter compared to summer. Except for the thrips-associated genomovirus, which was most abundant in the summer, all animal-associated viruses showed the highest abundance when the daycare center was closed. The RNA viruses consisted of positive sense ssRNA viruses, ssRNA-RT viruses, and dsRNA viruses. All ssRNA-RT viruses (Alpharetrovirus, Y73 virus, and Human retrovirus) were most abundant during the winter and when the daycare center was closed (which was during the winter, Table 1).

Discussion

Both human occupancy and seasonality influenced the airborne virome in the daycare center in this study. When the building was closed and unoccupied, the community structure was significantly different than at all other times of the year, highlighting the apparent role humans play in the viral ecology of this building. A limitation of this study is that we were only able to collect a single sample when the daycare center was unoccupied,

as it is closed for an extended period of time only once per year. Future research is needed to confirm the effect of human occupancy through either a multi-year and/or multi-location study.

The importance of human occupancy in shaping the airborne microbiome has been established previously for bacteria [43, 44]; this study shows that humans also shape the airborne virome in the BE. Our results show that human-associated viruses were more abundant in winter, while plant-associated viruses were more abundant in summer in this building. The reason for the higher proportion of human-associated viruses in winter was likely a combination of the children spending a larger fraction of their time indoors (when the weather was warmer, they usually spent at least 2 h per day outside), a lower air-exchange rate due to closed windows, and a higher recirculation rate by the heating system. In contrast, windows were open more often during the warmer months, leading to a higher air-exchange rate and greater influence of viruses from outdoor air, such as plant-associated ones. This pattern could differ in a warmer climate if, for instance, windows remain open more during the cooler months and closed with the air conditioning running during the summer.

There was no significant difference in the richness of DNA viruses between different seasons, and while evenness was higher in the spring, it was not significantly so. Our observation of a lower diversity of DNA viruses when the daycare center was closed is consistent with the idea that humans contribute to the viral richness in the BE. In contrast to DNA viruses, there were seasonal differences in the richness and evenness of RNA viruses; measures of richness and evenness were higher in the winter compared to other seasons (Fig. 2d–f). This observation could be due to the fact that RNA viruses are generally less stable in the environment compared to DNA viruses, and RNA viruses have been shown to persist longer at colder temperatures [45, 46], although this would apply mainly to viruses of outdoor origin, since indoor temperatures do not vary much by season. We also saw a reduction in RNA virus richness when the daycare center was closed, but its evenness was comparable to that of the winter season. When the daycare center was open, there was a higher frequency of low-abundance viruses compared to when the facility was closed. A possible reason for this is that when the daycare center was open, the children and staff contributed a constant influx of viruses from themselves and the outside environment, thereby

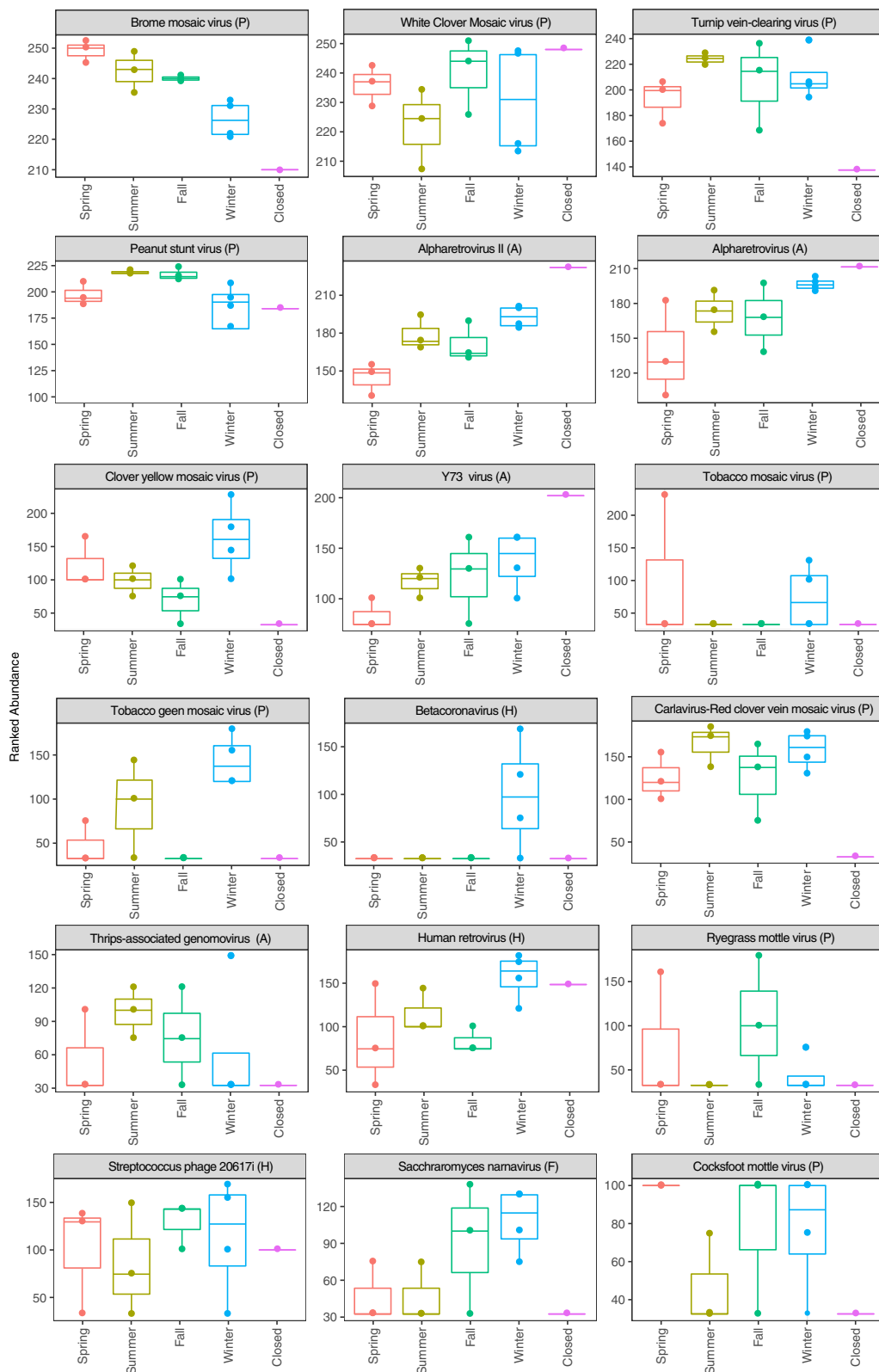


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Ranked abundances of RNA viral species that had the strongest seasonality. Random forest modeling was used to identify viruses that best distinguish between the different seasons: spring, summer, fall, and winter and when the daycare center was closed. The ranked abundances of the top 18 most discriminant RNA viruses are shown identified down to the strain level, if possible. *P* = plant-associated virus, *H* = human-associated virus, *A* = animal-associated virus, *I* = insect-associated virus, and *F* = fungal-associated virus

increasing viral richness. The somewhat counterintuitive increase in evenness during closure appears to result from the fact that there were relatively similar abundances of the persisting viruses.

One of the most striking observations was a noticeable clustering in the virome samples based on the season. The significant effect of seasonality on the virome was notable, as previous work has suggested that although seasonal dynamics help shape the bacterial microbiome in this and other buildings, seasonality is not as significant as other factors (e.g., human occupancy, building location, etc.) [34, 43]. However, for fungi in indoor air, researchers have shown that the community is largely shaped by outdoor air [35, 47]. In sum, our results suggest that a combination of seasonality and human occupancy affects the airborne virome in this daycare center.

A biplot was also used to identify viruses that were the main drivers of seasonal differences in the community. For DNA viruses, we found that the relative abundance of Cytomegalovirus (Human betaherpes virus 5) increased during the winter compared to other seasons, while the relative abundance of Lactococcus phage decreased. Cytomegalovirus infection and transmission occur frequently in daycare centers [48, 49]. This is due to saliva and urine being the primary transmission routes, coupled with the fact that nearly one third of children are infected by age 5 [50]. Lactococci are lactic acid bacteria isolated from numerous sources, including

plant surfaces, milk, and animal gastrointestinal tracts [51, 52]. For RNA viruses, the mean relative abundance of plant-associated viruses such as Turnip vein-clearing virus and Carlavirus-Red clover mosaic virus increased in the summer while the abundance of Brome Mosaic Virus and Peanut stunt virus increased in the spring. The blue biplot arrows for the tobacco mild green mosaic virus did not point towards one specific season cluster; however, they did seem to point more towards the summer than the winter season suggesting that their abundance was highest during the summer season. Griffin et al. [53] previously showed that viruses can be transported long distances in the atmosphere, so it is plausible that nearby farms could have been a contributing source. Our results could potentially correlate to the local growing and harvesting season of crops affected by these viruses, although a future source tracking study would be needed to confirm this hypothesis.

Phages were important in driving seasonal variability. Phages with human-associated bacteria as hosts tended to have a higher ranked abundance in the winter compared to the summer. For example, the crAssphage virus, a human gut-associated bacteriophage that is highly abundant in human feces [54], had a higher normalized abundance in the winter than in the summer, consistent with our other results showing greater abundance of human-associated viruses during the winter. A lower air-exchange rate, in combination with frequent changing of diapers, could explain the increased presence of crAssphage in the winter.

The seasonal dynamics in the airborne viral community of daycare center are likely attributable to a combination of human occupancy, building ventilation, and outdoor factors. Human occupancy appears to shape the virome in the wintertime when the building is “tight” and has less air-exchange with outdoors. The large abundance of human skin-associated bacteriophages is not surprising, as humans shed an estimated 14×10^6 bacterial cells per person per day [55]. The increase in abundance of plant-associated viruses in the summer could be explained by an increase in natural ventilation (i.e., windows and doors open more frequently) and possibly their greater abundance in outdoor air. In this geographic region, vegetation is mostly dormant during the winter.

Our results showing clear seasonal dynamics in the indoor air virome contrast with our finding of no seasonality in the bacterial microbiome in this same daycare

Table 1 Sampling dates for viral analysis

Sample number	Start date	End date	Season
1	20 January 2014	17 February 2014	Winter
2	17 February 2014	17 March 2014	Winter
3	17 March 2014	14 April 2014	Spring
4	14 April 2014	12 May 2014	Spring
5	12 May 2014	09 June 2014	Spring
6	09 June 2014	07 July 2014	Summer
7	07 July 2014	04 August 2014	Summer
8	04 August 2014	02 September 2014	Summer
9	02 September 2014	29 September 2014	Fall
10	29 September 2014	27 October 2014	Fall
11	27 October 2014	24 November 2014	Fall
12	24 November 2014	23 December 2014	Winter
13	05 January 2015	02 February 2015	Winter
14 ^a	23 December 2014	05 January 2015	Closed

^aDaycare center was closed during this entire period

center [43]. In terms of total number, viruses are approximately equally abundant as bacteria in the air [21], although the seasonal dynamics of numbers and community composition are not known for viruses. The influence of outdoor air on the indoor microbiome could be stronger for viruses than bacteria, as virus-sized particles have a higher penetration efficiency through the building envelope than do larger bacteria-sized particles [56]. Differences in the relative source strength of humans for bacteria vs. viruses could also contribute to differences in seasonal patterns. For fungi, a strong seasonal signal appears to be due to the fact that there are few sources of fungi indoors [47], unless a building suffers from mold. A future source-tracking study, similar to what has been done for bacteria in outdoor air [57], would be valuable for viruses in the BE. Future work is also needed to determine if seasonality also plays an important role in shaping the airborne virome in other buildings and outdoor air.

There are some limitations to this study. Compared to bacteria and fungi, there is a limited representation of viruses in reference databases [22, 58]. Although metagenomics is a very powerful tool to overcome the fact that the vast majority of microbes is not culturable [59], using a sequence-based approach to identify communities does not provide any insight into the viability of the collected viruses. Viability is critical, of course, in understanding these viruses' role in microbial ecology [60]. Indeed viability is a knowledge gap in many studies examining the microbiome of the BE [61]. Further, all of the numbers reported here are relative abundances, making it impossible to assess whether a higher relative abundance between seasons is due to an increase in certain types of viruses or a decrease in others. Future work would be enhanced by including both viability and absolute abundance in analyses. There is no perfect sample processing protocol for metagenomics, and each step can introduce bias (e.g., filter processing, nucleic acid extraction, sample preparation for sequencing, bioinformatics) [62]. To minimize any potential bias and/or contamination, we included two negative controls (unexposed filter and molecular biology grade water). Both of our controls showed an extremely low number of BLAST hits of DNA and RNA viruses (Additional file 1: Figure S1). RNA degrades more rapidly than DNA, and it is possible some of the RNA degraded before analysis. To minimize any bias and limit degradation, we stored collected samples at -80°C , and we analyzed the DNA and RNA viruses independently from each other. Finally, we were able to collect only one sample when the daycare center was closed and unoccupied for an extended period of time. Future studies are needed to examine the effect of occupancy on the airborne virome.

Airborne bacteria and fungi have been relatively well studied in the BE, while viruses have been the “forgotten siblings” [10]. A similar study to that of Kembel et al. [5] examining how building design and operation influences bacteria in the BE should be undertaken for viruses. Researchers should begin to examine how different building parameters (e.g., relative humidity, temperature, and moisture) can affect the virome. For example, relative humidity has been shown to play an important role in the transmission of influenza [63, 64], and it is possible relative humidity could also alter the virome. Additionally, understanding the differences in the airborne virome of the BE between rural and urban settings would address fundamental questions regarding human and plant sources of viruses in the BE.

Conclusions

Viruses play an important role in human health and the microbial ecology of the built environment, but relatively little is known about the indoor viral microbiome even as we have learned much about the bacterial and fungal communities. We have shown that both human occupancy and season are important in driving the community composition of airborne viruses in a daycare center. Human-associated viruses were much more diverse and dominant in the winter, while the summertime virome contained a high relative proportion and diversity of plant-associated viruses. Armed with a more complete understanding of the airborne microbiome and sources of bioaerosols, biologists, engineers, and architects can work together to optimize the microbiome of the BE for improved health and well-being [5, 65].

Methods

Sample collection

We collected filters from the heating, ventilation, and air conditioning (HVAC) system in a daycare center in Blacksburg, VA, USA, between January 2014 and February 2015, with permission from the center's director and staff as previously described [43]. Previous work has shown that microbial communities collected on HVAC filters were not different from those in air samples collected using an impinger [66]. Blacksburg, Virginia, has a moist continental mid-latitude climate and exhibits four distinct seasons. The center is open from 7:15 am to 5:45 pm Monday through Friday, and a typical day includes organized indoor activities, outdoor play, snack time, lunch time, and nap time. The rooms are cleaned daily, including removal of garbage, vacuuming and mopping of floors, and cleaning of kitchen and bathroom surfaces. The center has a total floor area of 1187 m^2 ($12,800\text{ ft}^2$) split between two buildings, each of which is served by a 4-ton split-system heat pump rated at $2000\text{ ft}^3\text{ min}^{-1}$ (carrier) [43]. The HVAC system

was operating 10–60% of the time when the daycare center was occupied and 38% of the time when the daycare center was closed/unoccupied [43]. Further, the average temperature inside the building ranged between 17 and 21 °C and relative humidity ranged between 26 and 66% throughout the sampling campaign [43]. Every 2 weeks, we removed an exposed filter (Nordic Pure, Tulsa, OK) and installed a new one in an HVAC return duct that was located in a hallway connecting the lobby to the kitchen and four children's rooms. The filter had a Minimum Efficiency Reporting Value (MERV) rating of 14, meaning that its average particle collection efficiency was > 98%, and its efficiency over the particle size range of 0.3 to 1.0 μm , the most difficult size to collect, was 75–85%. We transported the exposed filter to the laboratory immediately, cut it into $\sim 8\text{ cm} \times 8\text{ cm}$ squares under sterile conditions, placed the filter pieces in a sterile bag (Lansinoh, Alexandria, VA, USA), and froze them at $-80\text{ }^\circ\text{C}$ until further processing.

Sample processing and nucleic acid isolation

To obtain a sample that was representative of about 1 month and maximize the amount of biomass, we combined two 2-week samples into a single, composite sample (Table 1). One square ($\sim 8\text{ cm} \times 8\text{ cm}$) from each filter was cut into smaller pieces ($\sim 2\text{ cm}^2$) and placed into a 50-mL conical tube. We removed the virus particles by vigorously vortexing the filter pieces in $\sim 20\text{ mL}$ of 3% beef extract and 0.05 M glycine in molecular biology grade water and then shook them for $\sim 15\text{ min}$ at 200 rpm, as previously described [43, 67]. We extracted viral nucleic acid using the QIAamp UltraSens Virus Kit following the manufacturer's protocol (Qiagen, Calencia, CA, USA). Additionally, we processed two controls (an unexposed filter and a true negative control of molecular biology grade water) identically to the exposed filters. We stored the extracted nucleic acid at $-80\text{ }^\circ\text{C}$ until sequencing.

Sequencing

We fragmented DNA samples by Covaris S2 (intensity setting = 5, duty% = 10, burst cycles = 200, 50 s with frequency sweeping mode). Following the manufacturer recommended protocol, we took the fragmented DNA products into the NEBNext Ultra DNA Library Prep Kit for Illumina with 15 cycles of PCR with no size selection post ligation. We cleaned PCR amplified libraries using 1X AmpureXP Beads and quantified the indexed libraries individually and pooled at equal molar quantity. We gel-purified the pooled libraries by E-Gel EX Agarose Gels, 2% using E-Gel iBase Power System by Thermo Fisher Scientific. Finally, we excised a range of 350–500 bp and sequenced on a NextSeq500 using 2×150 paired-end read.

We prepared RNA samples using the Illumina Script-Seqv2 RNA-Seq library preparation kit following the manufacturer recommended protocol. The protocol follows a fragmentation at 85 °C for 5 min. Using ScriptSeq index PCR primers, we barcoded the libraries and PCR amplified 15 cycles. We quantified the indexed libraries individually and pooled at equal molar quantity. We gel-purified the pooled libraries by E-Gel EX Agarose Gels, 4% using E-Gel iBase Power System by Thermo Fisher Scientific. Finally, we excised a range of 350–500 bp and sequenced on a NextSeq500 using 2×150 paired-end read.

Sequence analysis

Illumina sequencing yielded an average of ~ 21 million and ~ 8.7 million sequences across all DNA samples and their controls, respectively. An average of ~ 11.3 million and ~ 0.5 million sequences was obtained from RNA samples and their controls, respectively. We trimmed raw paired-end reads using the Trimmomatic (v.0.36) [68] default settings. This was followed by stringent error filtering using PRINSEQ (v.0.20.4) [69] with the following parameters: minimum sequence length of 60 bp, minimum mean quality score of 25, sequences containing any "N's" were removed and low-complexity threshold of 50 (using Entropy). We used DeconSeq software (coverage > 90, identity > 90) (v.0.4.3) [70] to filter out any human contamination. Following DeconSeq, our paired-end files were rewritten to make sure all reads had a mate and separated out any singletons using FASTQ Pair, available at <https://github.com/linsalrob/fastq-pair>. We merged overlapping pairs of reads using the default parameters in BBMerge (v.37.36) [71]. We noted that forward reads showed very high-quality scores; therefore, those sequences that did not merge were extracted from the unmerged output file (https://github.com/pjtorres/viral_bioaerosol/xtract_forward.py) as to not discard useful data.

We determined the composition of the metagenome sequence library by using a BLASTn pipeline against a viral RefSeq database. BLASTn (v. 2.6.0) parameters were modified using stringent alignments with *E* value 10^{-5} , a sequence identity of 90%, and a minimum raw gapped score of 105. We obtained viral taxonomic lineage using the NCBI accession number for sequences in the reference database and pulling their taxonomic lineage from NCBI (https://github.com/pjtorres/viral_bioaerosol/get_taxlineage_fna.py). We built a custom database containing all viral reference genomes from the NCBI RefSeq database (<ftp://ftp.ncbi.nih.gov/refseq/release/viral/>). All the scripts used from processing the raw data to BLAST output can be found at https://github.com/pjtorres/viral_bioaerosol. There were 7 DNA viral species and 9 RNA viral species that were removed from further analysis due to their high number of

BLAST hits in their respective negative controls. In addition, there were a number of hits to dsDNA and ssDNA viruses in our RNA library. In order to focus solely on RNA viruses we removed viruses whose phylum indicated “dsDNA viruses no RNA stage,” “ssDNA viruses,” and “unclassified bacterial viruses.” In the end, the average total hits in the BLAST output were 3892 for DNA samples and 100 for their controls, and 2302 for the RNA samples and 2 for their negative controls.

We used Rstudio (v.1.0.153) to compute the alpha and beta diversity metrics using phyloseq [72] and vegan [73] packages at a rarefied sampling depth of 1300 for DNA and 850 for RNA viruses. Both of our controls (unexposed filter and molecular biology grade water processed the same as the samples) showed an extremely low relative abundance of DNA and RNA viruses (Additional file 1: Figure S1 a-b). We used three alpha diversity metrics. To estimate species, we used observed species to define the total number of unique species in a community. To estimate species evenness, we used the Pielou evenness index. To account for both abundance and evenness of the species present, we used the Shannon diversity index. We used the Bray-Curtis distance matrix to compare the similarity (beta diversity) among viral communities and generated non-metric multidimensional scaling (NMDS) plots. We were also interested in identifying the probably environmental sources of the viruses and their hosts. To do so, we first inferred the phage host from the viral name (e.g., *Staphylococcus* phage host would be *Staphylococcus* sp., *Lactococcus* phage ul36 host would be *Lactococcus lactis*), and then, we looked up the sources of the bacteria via literature search (e.g., *Staphylococcus* sp. normally reside on the human skin, *Lactococcus lactis* are mainly isolated from plant material).

Statistical analysis

We performed statistical calculations in the Rstudio. We constructed NMDS plots using the metaMDS function in the R vegan package. Bray-Curtis dissimilarity compares the community structures by taking into account the abundance distribution of viral species. We then used a variation of the BIO-ENV [74], routine dubbed BIO-BIO (<http://menugget.blogspot.co.uk/2011/06/clarke-and-ainsworths-bioenv-and-bvstep.html>), to identify the subset of viral species which best correlated to the overall biological pattern of the dissimilarity matrix. We overlaid the vectors of the best correlated biological variables on the NMDS plot. The length of the arrow is proportional to the correlation between the viral species and biological pattern of the dissimilarity matrix. Blue arrows indicate positive correlation (the highest viral relative abundance in that direction), and red arrows indicate negative correlation (the lowest viral relative abundance in that direction). We

assessed variability in community structure by analysis of beta dispersion, which is based on the distances of each sample to its respective group centroid. Permutational Multivariate Analysis of Variance (PERMANOVA), aka Adonis, used Bray-Curtis dissimilarity measures to assess viral community compositional differences and its relationship to the different Seasons (999 permutations “vegan” package). We implanted the random forest classifier [75] in R using the “randomForest” library to identify viral species (present at least 10 times) that discriminate between the different seasons. Random forest is a supervised machine learning algorithm able to discriminate between two or more groups with high accuracy even in the presence of noise, high dimensional, and undersampled data, typical of biological problems [76]. Furthermore, the random forest can output the individual features (in our case viral hits) that contributed the most to the accuracy in discriminating between the groups (seasons).

Additional file

Additional file 1: Figure S1. Heat map showing the abundance of the 500 most abundance DNA (A) and RNA (B) viruses. Both of our controls (unexposed filter and molecular biology grade water processed the same as the samples) showed extremely low raw BLASTN counts of DNA (mean = 100) and RNA (mean = 2) viruses. (PDF 637 kb)

Acknowledgements

We thank Rob A. Edwards for the use of the Anthill computational cluster. We also thank Rob A. Edwards, Adrian Cantu, and Daniel A. Cuevas for their technical expertise and assistance. We are grateful to the daycare center for allowing us to collect samples there.

Funding

This work was supported by an Alfred P. Sloan Foundation Postdoctoral Fellowship Award to AJP, a National Institutes of Health New Innovator Award (1-DP2-A1112243), and the National Science Foundation (CBET-1438103 and ECCS-1542100). Additional support was provided by the Virginia Tech Institute for Critical Technology and Applied Science.

Availability of data and materials

The sequences were submitted in NCBI Sequence Read Archive (SRA) under the Bioproject ID PRJNA525405 with the following BioSample numbers: SAMN11050300 - SAMN11050331.

Authors' contributions

AJP, KJB, and LCM conceived the study. STK and LCM jointly supervised this work. AJP collected the samples and extracted the nucleic acid. SRH and JS processed the samples for sequencing. SRH and JS sequenced the samples. PJT built the bioinformatics pipeline and performed the analysis. AJP, PJT, KJB, STK, and LCM designed the experiment and analyzed the results. AJP, PJT, KJB, STK, and LCM wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA. ²Department of Biology, San Diego State University, San Diego, CA 92182, USA. ³Next Generation Sequencing and Microarray Core Facility, The Scripps Research Institute, La Jolla, CA 92037, USA. ⁴Department of Civil and Environmental Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.

Received: 24 August 2018 Accepted: 22 March 2019

Published online: 01 April 2019

References

- National Academies of Sciences, Engineering, and Medicine: Microbiomes of the built environment: a research agenda for indoor microbiology, human health, and buildings. National Academies Press; 2017.
- Kelley ST, Gilbert JA. Studying the microbiology of the indoor environment. *Genome Biol.* 2013;14:202.
- Adams RI, Bhargar S, Dannemiller KC, Eisen JA, Fierer N, Gilbert JA, Green JL, Marr LC, Miller SL, Siegel JA. Ten questions concerning the microbiomes of buildings. *Build Environ.* 2016;109:224–34.
- Gilbert JA, Stephens B. Microbiology of the built environment. *Nat Rev Microbiol.* 2018;1.
- Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, Bohannon BJ, Brown G, Green JL. Architectural design influences the diversity and structure of the built environment microbiome. *The ISME Journal.* 2012;6:1469–79.
- Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science.* 2014;345:1048–52.
- Hoisington AJ, Brenner LA, Kinney KA, Postolache TT, Lowry CA. The microbiome of the built environment and mental health. *Microbiome.* 2015;3:60.
- Konya T, Scott JA. Recent advances in the microbiology of the built environment. *Current Sustainable/Renewable Energy Reports.* 2014;1:35–42.
- Fouquier J, Schwartz T, Kelley S. Rapid assemblage of diverse environmental fungal communities on public restroom floors. *Indoor Air.* 2016;26:869–79.
- Williams SC. The other microbiome. *Proc Natl Acad Sci.* 2013;110:2682–4.
- Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ Microbiol.* 2004;6:1–11.
- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol.* 2007;5:801–12.
- Samson JE, Magadán AH, Sabri M, Moineau S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol.* 2013;11:675.
- Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol.* 2010;8:317.
- Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. *EMBO Rep.* 2001;2:376–81.
- Tellier R. Review of aerosol transmission of influenza a virus. *Emerg Infect Dis.* 2006;12:1657.
- Dick EC, Jennings LC, Mink KA, Wartgow CD, Inborn SL. Aerosol transmission of rhinovirus colds. *J Infect Dis.* 1987;156:442–8.
- Prince DS, Astry C, Vonderfecht S, Jakab G, F-m S, Yolken RH. Aerosol transmission of experimental rotavirus infection. *Pediatr Infect Dis J.* 1986;5:218–22.
- Chang L-Y, King C-C, Hsu K-H, Ning H-C, Tsao K-C, Li C-C, Huang Y-C, Shih S-R, Chiou S-T, Chen P-Y. Risk factors of enterovirus 71 infection and associated hand, foot, and mouth disease/herpangina in children during an epidemic in Taiwan. *Pediatrics* 2002, 109:e88–e88.
- Bonifait L, Charlebois R, Vimont A, Turgeon N, Veillette M, Longtin Y, Jean J, Duchaine C. Detection and quantification of airborne norovirus during outbreaks in healthcare facilities. *Clin Infect Dis.* 2015;61:299–304.
- Prussin AJ, Garcia EB, Marr LC. Total concentrations of virus and bacteria in indoor and outdoor air. *Environmental Science & Technology Letters.* 2015;2:84–8.
- Prussin AJ II, Marr LC, Bibby KJ. Challenges of studying viral aerosol metagenomics and communities in comparison with bacterial and fungal aerosols. *FEMS Microbiol Lett.* 2014.
- Wu HM, Fornek M, Schwab KJ, Chapin AR, Gibson K, Schwab E, Spencer C, Henning K. A norovirus outbreak at a long-term-care facility: the role of environmental surface contamination. *Infection Control & Hospital Epidemiology.* 2005;26:802–10.
- Morter S, Bennet G, Fish J, Richards J, Allen D, Nawaz S, Iturriza-Gómara M, Broly S, Gray J. Norovirus in the hospital setting: virus introduction and spread within the hospital environment. *J Hosp Infect.* 2011;77:106–12.
- Keswick B, Pickering L, DuPont H, Woodward W. Survival and detection of rotaviruses on environmental surfaces in day care centers. *Appl Environ Microbiol.* 1983;46:813–6.
- Gibbons SM, Schwartz T, Fouquier J, Mitchell M, Sangwan N, Gilbert JA, Kelley ST. Ecological succession and viability of human-associated microbiota on restroom surfaces. *Appl Environ Microbiol.* 2015;81:765–73.
- Be NA, Thissen JB, Fofanov VY, Allen JE, Rojas M, Golovko G, Fofanov Y, Koshinsky H, Jaing CJ. Metagenomic analysis of the airborne environment in urban spaces. *Microb Ecol.* 2015;69:346–55.
- Hall RJ, Leblanc-Maridor M, Wang J, Ren X, Moore NE, Brooks CR, Peacey M, Douwes J, McLean DJ. Metagenomic detection of viruses in aerosol samples from workers in animal slaughterhouses. *PLoS One.* 2013;8:e72226.
- Whon TW, Kim M-S, Roh SW, Shin N-R, Lee H-W, Bae J-W. Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. *J Virol.* 2012;86:8221–31.
- Brisebois E, Veillette M, Dion-Dupont V, Lavoie J, Corbeil J, Culley A, Duchaine C. Human viral pathogens are pervasive in wastewater treatment center aerosols. *J Environ Sci.* 2018;67:45–53.
- Rosario K, Fierer N, Miller SL, Luongo J, Breitbart M. Diversity of DNA and RNA viruses in indoor air as assessed via metagenomic sequencing. *Environmental Science & Technology.* 2018;52(3):1014–27.
- Bowers RM, McCubbin IB, Hallar AG, Fierer N. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmos Environ.* 2012;50:41–9.
- Nazaroff WW. Indoor bioaerosol dynamics. *Indoor Air.* 2014;26(1).
- Rintala H, Pitkäranta M, Toivola M, Paulin L, Nevalainen A. Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiol.* 2008;8:56.
- Adams RI, Miletto M, Lindow SE, Taylor JW, Bruns TD. Airborne bacterial communities in residences: similarities and differences with fungi. *PLoS One.* 2014;9:e91283.
- Leung MH, Lee PK. The roles of the outdoors and occupants in contributing to a potential pan-microbiome of the built environment: a review. *Microbiome.* 2016;4:21.
- Loda FA, Glezen WP, Clyde WA. Respiratory disease in group day care. *Pediatrics.* 1972;49:428–37.
- Ansari SA, Springthorpe VS, Sattar SA. Survival and vehicular spread of human rotaviruses: possible relation to seasonality of outbreaks. *Review of Infectious Diseases.* 1991;13:448–61.
- Grassly NC, Fraser C. Seasonal infectious disease epidemiology. *Proc R Soc Lond B Biol Sci.* 2006;273:2541–50.
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature.* 1999;399:541–8.
- Bartlett AV, Moore M, Gary GW, Starko KM, Erben JJ, Meredith BA. Diarrheal illness among infants and toddlers in day care centers. I. Epidemiology and pathogens. *J Pediatr.* 1985;107:495–502.
- Hutchinson MK: Infectious diseases and infection control in infant-toddler daycare centers. *Child and Youth Care Forum Springer;* 1992, 21: 183–193.
- Prussin AJ II, Vikram A, Bibby KJ, Marr LC. Seasonal dynamics of the airborne bacterial community and selected viruses in a children's daycare center. *PLoS One.* 2016;11:e0151004.
- Hospodsky D, Qian J, Nazaroff WW, Yamamoto N, Bibby K, Rismani-Yazdi H, Peccia J. Human occupancy as a source of indoor airborne bacteria. *PLoS One.* 2012;7:e34867.
- Wood JP, Choi YW, Chappie DJ, Rogers JV, Kaye JZ. Environmental persistence of a highly pathogenic avian influenza (H5N1) virus. *Environmental Science & Technology.* 2010;44:7515–20.
- Tamerius JD, Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK, Comrie A, Viboud C. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog.* 2013;9:e1003194.

47. Adams RI, Miletto M, Taylor JW, Bruns TD. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The ISME Journal*. 2013;7:1262–73.
48. Murph JR, Baron JC, Brown CK, Ebelhack CL, Bale JF. The occupational risk of cytomegalovirus infection among day-care providers. *JAMA*. 1991;265:603–8.
49. Murph JR, Bale JF Jr, Murray JC, Stinski MF, Perlman S. Cytomegalovirus transmission in a Midwest day care center: possible relationship to child care practices. *J Pediatr*. 1986;109:35–9.
50. Pass RF. Epidemiology and transmission of cytomegalovirus. *J Infect Dis*. 1985;152:243–8.
51. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarne K, Weissenbach J, Ehrlich SD, Sorokin A. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res*. 2001;11:731–53.
52. Deveau H, Labrie SJ, Chopin M-C, Moineau S. Biodiversity and classification of lactococcal phages. *Appl Environ Microbiol*. 2006;72:4338–46.
53. Griffin DW, Garrison VH, Herman JR, Shinn EA. African desert dust in the Caribbean atmosphere: microbiology and public health. *Aerobiologia*. 2001;17:203–13.
54. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun*. 2014;5:ncomms5498.
55. Hospodsky D, Yamamoto N, Nazaroff W, Miller D, Gorthala S, Peccia J. Characterizing airborne fungal and bacterial concentrations and emission rates in six occupied children's classrooms. *Indoor Air*. 2015;25:641–52.
56. Nazaroff WW. Indoor particle dynamics. *Indoor Air*. 2004;14:175–83.
57. Bowers RM, Sullivan AP, Costello EK, Collett JL, Knight R, Fierer N. Sources of bacteria in outdoor air across cities in the midwestern United States. *Appl Environ Microbiol*. 2011;77:6350–6.
58. Bibby K. Improved bacteriophage genome data is necessary for integrating viral and bacterial ecology. *Microb Ecol*. 2014;67:242–4.
59. Streit WR, Schmitz RA. Metagenomics—the key to the uncultured microbes. *Curr Opin Microbiol*. 2004;7:492–8.
60. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol*. 2005;13:278–84.
61. Stephens B. What have we learned about the microbiomes of indoor environments? *MSystems*. 2016;1:e00083-00016.
62. Maestre JP, Jennings W, Wylie D, Horner SD, Siegel J, Kinney KA. Filter forensics: microbiota recovery from residential HVAC filters. *Microbiome*. 2018;6:22.
63. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog*. 2007;3:e151.
64. Yang W, Marr LC. Dynamics of airborne influenza viruses indoors and dependence on humidity. *PLoS One*. 2011;6:e21481.
65. Dai D, Prussin AJ, Marr LC, Vikesland PJ, Edwards MA, Pruden A. Factors shaping the human exposome in the built environment: opportunities for engineering control. *Environmental Science & Technology*. 2017;51:7759–74.
66. Noris F, Siegel JA, Kinney KA. Evaluation of HVAC filters as a sampling mechanism for indoor microbial communities. *Atmos Environ*. 2011;45:338–46.
67. Farnsworth JE, Goyal SM, Kim SW, Kuehn TH, Raynor PC, Ramakrishnan M, Anantharaman S, Tang W. Development of a method for bacteria and virus recovery from heating, ventilation, and air conditioning (HVAC) filters. *J Environ Monit*. 2006;8:1006–13.
68. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
69. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863–4.
70. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One*. 2011;6:e17288.
71. Bushnell B, Rood J, Singer E. BBMerge—accurate paired shotgun read merging via overlap. *PLoS One*. 2017;12:e0185056.
72. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
73. Oksanen J. *Vegan: community ecology package*. R package version 2.0-2. <http://CRAN.R-project.org/package=vegan> 2011.
74. Clarke K, Ainsworth M. A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser*. 1993;205–19.
75. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
76. de Ruiter J, Knijnenburg T, de Ridder J. Mining the forest: uncovering biological mechanisms by interpreting random forests. *BioRxiv*. 2017:217695.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

