

# 1 **Untargeted viral metagenomics of indoor air as a novel surveillance** 2 **tool for respiratory, enteric and skin viruses**

3

4 Mustafa Karatas<sup>1\*</sup>, Caspar Geenen<sup>2</sup>, Els Keyaerts<sup>3</sup>, Lore Budts<sup>3</sup>, Joren Raymenants<sup>2</sup>, Charlotte  
5 Eggers<sup>1</sup>, Bastiaan Craessaerts<sup>3</sup>, Emmanuel André<sup>2,3</sup> \*† and Jelle Matthijnsens<sup>1\*†</sup>

6

7 \* Corresponding authors

8 † These authors contributed equally.

9 <sup>1</sup> Laboratory of Viral Metagenomics, Clinical and Epidemiological Virology, Rega Institute,  
10 Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium

11 <sup>2</sup> Laboratory of Clinical Microbiology, KU Leuven, Leuven, Belgium

12 <sup>3</sup> Department of Laboratory Medicine, National Reference Center of Respiratory Pathogens,  
13 University Hospitals Leuven, Leuven, Belgium

## 14 **Summary**

15

16 Conventional hospital-based infectious disease surveillance struggles to detect mild or  
17 asymptomatic infections and incurs high costs for large-scale testing during outbreaks. In  
18 contrast, environmental surveillance can effectively monitor viral circulation and capture  
19 asymptomatic or mildly symptomatic infections that may otherwise go unnoticed. This  
20 longitudinal study explores the use of indoor air, in combination with targeted qPCR panels  
21 and untargeted viral metagenomics, as a novel virus surveillance tool. Indoor air samples were  
22 collected weekly from a daycare center in Leuven, Belgium, over a 12-month period using  
23 active indoor air sampling, followed by screening using respiratory and enteric qPCR panels,  
24 as well as untargeted viral metagenomics.

25

26 Human-associated viruses were detected in 95.2% (40/42) of samples, with MW  
27 polyomavirus being the most prevalent at 80.9%. Several other respiratory viruses (e.g.,  
28 rhinoviruses, RSV-B) and enteric viruses (e.g., rotavirus, astrovirus, adenovirus) were  
29 identified, correlating with their known epidemiological circulation patterns. Metagenomics  
30 resulted in multiple complete viral genome reconstructions, allowing discrimination of viral  
31 subtypes and the identification of coinfections with closely related virus variants. Finally, a  
32 plethora of animal, insect, fungal and plant-infecting viruses could be detected, representing  
33 the indoor and outdoor environment. Indoor air surveillance can be a robust complementary  
34 tool for virus surveillance.

35

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## 36 Introduction

37 Conventionally, infectious disease surveillance has revolved around patient-centered  
38 approaches, focusing on specific diseases or syndromes (e.g., respiratory, enteric symptoms)  
39 using methods such as pathogen culturing and molecular techniques (PCR or qPCR panels) on  
40 individual patient samples<sup>1</sup>. While valuable for immediate patient care, these methods often  
41 fail to detect novel viruses, can be invasive, and become costly when extensive testing is  
42 required<sup>2</sup>. In addition, consent is required, which could be challenging, especially for  
43 infections that cause many mild or even asymptomatic cases but can also lead to severe  
44 presentations. The COVID-19 pandemic underscored these challenges and highlighted the  
45 need for scalable surveillance methods. In response, environmental surveillance has emerged  
46 as a crucial alternative, allowing for the monitoring of viral circulation in communities without  
47 relying on direct patient sampling.

48  
49 Over the past decade, environmental surveillance for pathogens, using media such as  
50 wastewater, soil, rivers, and indoor air, has attracted growing interest, with wastewater being  
51 the most widely utilized method<sup>3-5</sup>. Wastewater analysis has effectively complemented other  
52 surveillance systems by tracking SARS-CoV-2, its variants of concern, and other respiratory and  
53 enteric pathogens using targeted (quantitative) PCR methods, thereby reinforcing its  
54 importance alongside conventional patient-centered surveillance<sup>6</sup>. This approach offers  
55 insights into prevalence and sometimes genetic diversity of pathogens across broader  
56 populations and enables the early detection of emerging strains and trends<sup>7,8</sup>. However,  
57 wastewater surveillance has limitations, including reduced precision in pinpointing specific  
58 locations or population subgroups, potential delays or signal dilution due to rainfall, and  
59 dependency on water usage<sup>9</sup>. Recently, our teams and others have demonstrated the  
60 potential of active indoor air sampling for monitoring viruses<sup>10-13</sup>. This method has  
61 successfully identified a variety of viruses, including respiratory and enteric viruses (e.g.,  
62 respiratory syncytial virus, influenza C, human parainfluenza virus) in community settings.

63  
64 The methodologies employed to detect pathogens in environmental surveillance are usually  
65 similar to those used in clinical diagnostics<sup>1,14</sup>, often using molecular based single- or multi-  
66 targeted (qPCR or multiplex qPCR) tests. Targeted test methods tend to have high sensitivity  
67 and specificity. Previously, a number of studies has focused on the detection of viruses and  
68 bacteria with targeted methods from indoor air of various community settings<sup>10,12,13</sup>, revealing  
69 temporal trends of virus circulation among specific populations and highlighting the early  
70 warning potential of indoor air sampling in real-world situations.

71  
72 However, for more precise genetic identification or subtyping, additional techniques such as  
73 Sanger sequencing or Whole Genome Sequencing (WGS) are often necessary. Characterizing  
74 novel or variant viruses can be challenging with both methods, because they require specific  
75 sequence knowledge of viruses of interest. However, Next Generation Sequencing (NGS)-  
76 based 'untargeted' shotgun metagenomics can overcome this limitation. These methods often

77 use a combination of viral particle purification, followed by deep sequencing using various  
78 NGS technologies and appropriate bio-informatics tools<sup>15</sup>. An established method, Novel  
79 Enrichment Techniques of Viromes (NetoVIR)<sup>16</sup>, has proven effective for viral enrichment  
80 across various sample types, including fecal<sup>17,18</sup>, insect<sup>19,20</sup>, and plant samples<sup>21</sup>, facilitating  
81 virome analysis with minimal bias and the discovery of novel viruses. This approach can  
82 identify the pathogen and provide genomic insights as a 'pathogen-agnostic' tool<sup>22</sup>. For indoor  
83 air surveillance, NGS faces challenges such as low biomass and contaminating nucleic acid,  
84 despite its relatively cleaner nature compared to soil, wastewater or stool<sup>23</sup>. Nevertheless,  
85 *Minor et al.* used RNA virus-targeted next-generation sequencing (NGS) to analyze 20 samples  
86 from various indoor community settings. They detected respiratory viruses such as influenza  
87 A and C, but also rotaviruses and astrovirus, which are transmitted through the fecal-oral  
88 route<sup>24</sup>.

89

90 In this study, we employed weekly active air sampling at a daycare center for children aged 0-  
91 3 years, spanning a full year, and analyzed these samples using untargeted viral  
92 metagenomics. Our objective was to investigate the potential of air sampling for surveillance,  
93 by employing both reference-guided and de-novo bioinformatics approaches to achieve a  
94 better understanding of the indoor air virome and its dynamics.

## 95 **Results**

96

### 97 **1. Study design**

98 Air samples were collected longitudinally from a daycare center unit in Leuven, Belgium,  
99 between January and December 2022. Particles suspended in indoor air were collected using  
100 an AerosolSense sampler (Thermo Fisher Scientific, Waltham, MA) for two hours. Ninety  
101 samples were collected from the same location, near the ventilation system's air extraction  
102 point in the changing room of the daycare unit. All ninety samples were processed with a  
103 respiratory qPCR panel, while forty-two available samples were further processed using an  
104 enteric qPCR panel and viral metagenomics (*see Methods*). For the latter, samples were  
105 enriched for viral-like particles (VLPs) using the NetoVIR protocol and sequenced on the  
106 Illumina NovaSeq 6000 platform. On average, 25 million reads were produced per sample.  
107 Produced reads were then further analyzed using two bioinformatics approaches: *EsVirtu*,  
108 referred to as "reference guided assembly", and *ViPER*, referred to as "*de-novo* assembly" (*see*  
109 *methods*). Findings from these two approaches are explained and discussed together, unless  
110 specified otherwise.

111

### 112 **2. Untargeted shotgun metagenomics revealed a spectrum of enteric, 113 respiratory and skin viruses in indoor air**

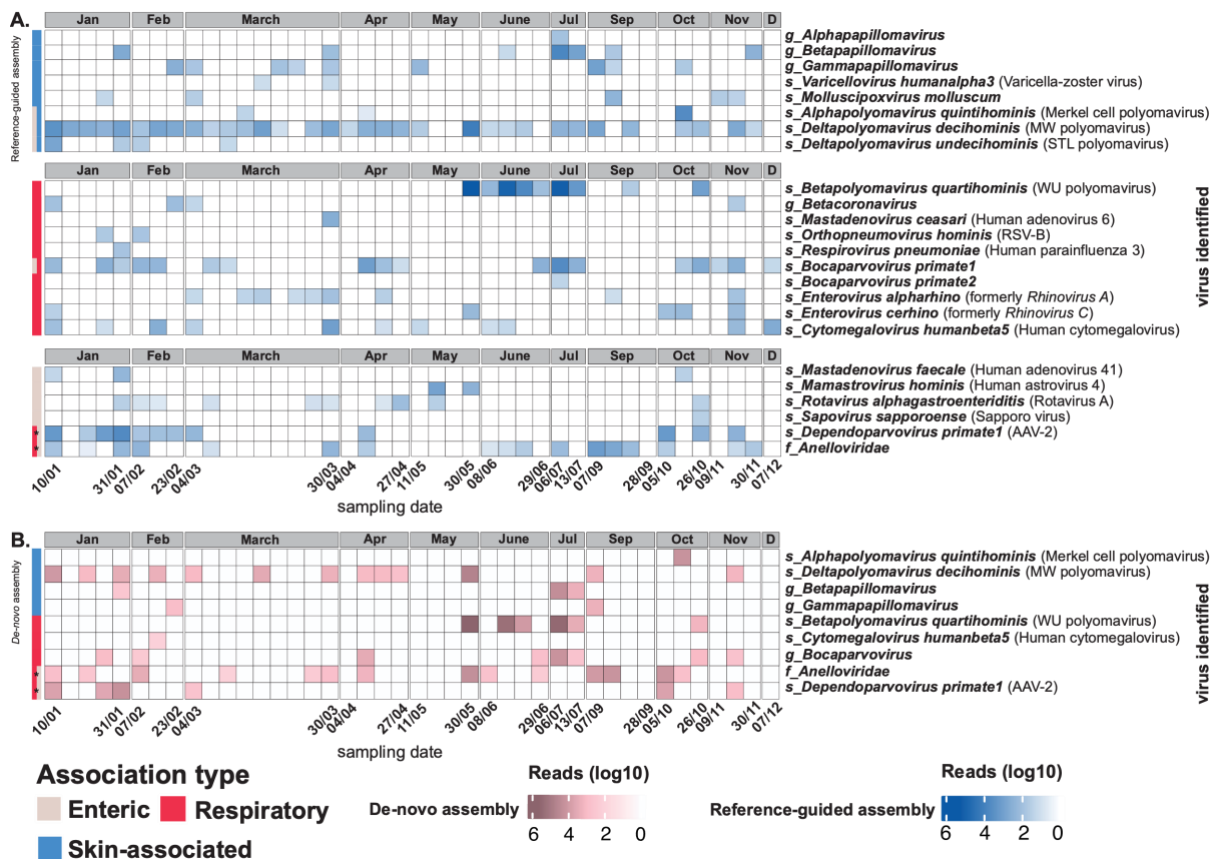
114

115 Using reference-guided assembly, we identified human-associated viruses in 40 out of 42  
116 samples (95%, list of viruses identified in each sample with read counts can be found in

117 **Supplementary information 1).** *De novo* assembly identified human-associated viruses in 29  
118 out of 42 samples (69%) (**Figure 1. A. & B.**). Several respiratory viruses were detected  
119 throughout the year, including human bocavirus 1, 2 and 3 (n=18, species *Bocaparvovirus*  
120 *primate1-2*), *Enterovirus alpharhino* and *Enterovirus cerhino* (formerly *Rhinovirus A* (n=9) and  
121 *C* (n=5)), human adenovirus 6 (n=1), viruses belonging to the genera *Betacoronavirus* (Human  
122 coronavirus HKU1, n=3 and OC43, n=1) and respiratory syncytial virus B / RSV-B, n=2). WU  
123 polyomavirus was consistently detected from the end of May until the summer holiday closure  
124 in July, and it accounted for the highest total number of human-infecting virus reads in our  
125 study. Among the identified respiratory viruses, we obtained complete genomes of WU  
126 polyomavirus and human bocavirus 1.

127  
128 Enteric virus reads detected in the air included rotavirus A (n=10), human astrovirus 4 (n=2),  
129 sapporo virus (n=1), and human adenovirus 41 (n=1). The highest horizontal genome coverage  
130 from strictly enteric pathogens was reached for Human astrovirus 4 with more than 70%  
131 coverage. While these pathogens are known to cause significant illness in young children,  
132 other viruses such as MW polyomavirus, STL polyomavirus, human bocavirus 1-3,  
133 anelloviruses, and Adeno-associated virus-2 (AAV-2) are described in stool without being  
134 associated with enteric diseases. AAV-2 was identified in 12 samples, and anelloviruses were  
135 detected in 21 samples. In four of these 21 samples, anelloviruses were only identified  
136 through *de novo* assembly, highlighting their high genetic variability, which may have  
137 complicated detection by reference-based approaches.

138  
139 Notably, the most frequently identified virus was MW polyomavirus, present in 80.9% (34/42)  
140 of the samples throughout the year. This virus showed 98.8% nucleotide similarity to a strain  
141 reported to be shedding from the skin of healthy individuals in the USA<sup>25</sup>. Several skin-  
142 associated polyomaviruses were detected, including STL polyomavirus and Merkel Cell  
143 polyomavirus. Additionally, Varicella-Zoster Virus (VZV) and variety of human  
144 papillomaviruses (18 different types, including HPV-151, HPV-120, HPV-129, HPV-9, and HPV-  
145 76, detected in 14 samples) were identified.



146  
147 **Figure 1.** Human-associated viruses. On the x-axis, the first and last sample collection in each  
148 month is labelled. **A.** Human-associated viruses, identified by reference-based assembly. **B.**  
149 Human-associated viruses identified by de-novo assembly. Different taxonomical levels of  
150 identifications are abbreviated as follows: f: family g: genus, s: species. \*AAV-2 is known to  
151 infect a wide range of human cells and anelloviruses do not have a clear host cell, however  
152 both are reported to shed from the enteric and respiratory tracts.

153

### 154 **3. Seasonality of respiratory viruses was reflected in the air**

155

156 Examination of the seasonality of respiratory viruses with shotgun metagenomics revealed  
157 distinct patterns. Rhinoviruses were detected throughout the year, showing no clear seasonal  
158 pattern. Bocaparvoviruses were identified frequently in blocks of several consecutive positive  
159 samples (mid-January, April, end-June, mid-October), presumably representing outbreaks. In  
160 contrast, viruses with well-established seasonal patterns, such as coronaviruses and RSV,  
161 typically peak during the winter months<sup>26,27</sup> and were only detected in the months of January,  
162 February, and November in our study (**Figure 1A**).

163

164 To increase the resolution for seasonality analyses, we used qPCR respiratory panel results  
165 from 90 samples (the same 42 samples as above, plus 48 additional indoor air samples from  
166 the same year and location, *see methods*). RSV was detected only from January to April (n=12,  
167 CT range: 32.0-39.1) and November to December (n=7, CT range: 32.9-35.3). Similarly, human  
168 parainfluenza virus 3 was found in January to March (n=12, CT range: 33.5-38.7) and once in

169 May (n=1, CT: 35.3). Human coronaviruses (HKU-1 and OC43) were detected 42 times, mostly  
170 between January and March (n=31, CT range: 27.9-37.8) and in November to December (n=6,  
171 CT range: 27.9-37.8). Thus, both qPCR panels and shotgun metagenomics revealed distinct  
172 seasonality patterns for respiratory viruses in indoor air, particularly for RSV and human  
173 coronaviruses. Interestingly, Human parainfluenza virus 3 was detected in winter by indoor air  
174 sampling at daycare center, which is different than its defined peak in the literature, late spring  
175 and early summer. Furthermore, the respiratory qPCR panel and shotgun metagenomics did  
176 not detect any SARS-CoV-2, whereas the more sensitive TaqPath assay (*see methods*) detected  
177 SARS-CoV-2 in 45 out of 90 samples (**Supplementary information 2. S1.**).

178

#### 179 **4. Indoor air surveillance as a non-invasive way to follow temporal trends of** 180 ***rhinovirus and rotavirus A***

181

182 A recent study revealed that human rhinoviruses cause up to two-thirds as many hospital  
183 admissions in infants as RSV<sup>28</sup>, while rotavirus remains the most common gastroenteritis cause  
184 in children younger than five-year-old worldwide. In our data RSV was only identified in two  
185 indoor samples using shotgun metagenomics, potentially because infants with more severe  
186 disease were kept away from the daycare center. On the other hand, rotavirus and  
187 rhinoviruses were consistently detected (**Figure 1.A**), and therefore, they were selected for a  
188 more thorough comparison to other openly available surveillance data .

189

190 For rhinovirus epidemiology, from patients with respiratory symptoms from the Leuven  
191 university hospital, which tested positive for enteroviruses (which includes rhinoviruses; *see*  
192 *methods*) using the respiratory qPCR panel, were used as a proxy (**Figure 2.A**, brown bars).  
193 The same qPCR panel was applied to all indoor air samples (n=90), revealing a 95.5% positivity  
194 rate (87/90) (indicated with "PCR #" **Figure 2.A**). Shotgun metagenomics identified  
195 rhinoviruses in 30.9% (13/42) of the samples and not every month, which is lower compared  
196 to qPCR panel tests. Average qPCR values per month showed similar trend to clinical cases  
197 between January-March, May- July and September-November. Despite low read counts (read  
198 counts can be found in **Supplementary Information 1**), our shotgun metagenomics data  
199 allowed identification and subtyping of rhinoviruses (**Figure 2.A**, heatmap).

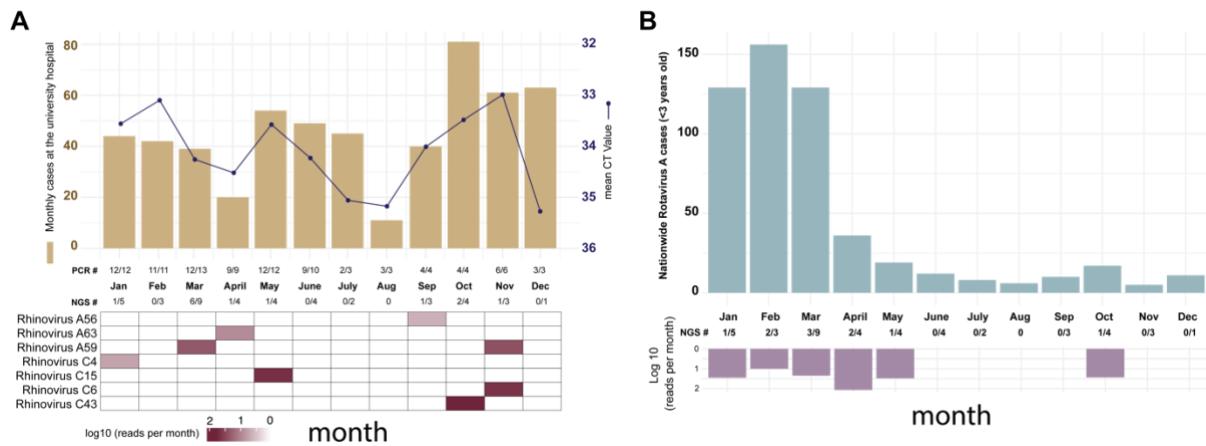
200

201 To compare indoor air surveillance with actual rotavirus A circulation, nationwide data from  
202 the National Reference Center (NRC) for rotavirus A (*see methods*) among children younger  
203 than three *years old* were used (**Figure 2.B**). Unlike rhinoviruses, rotavirus A was not identified  
204 using the qPCR enteric panel (0/42), but it was detected in 10 samples (January to May and  
205 October) using untargeted shotgun metagenomics. These detections corresponded well with  
206 nationwide data and the seasonality reported by the NRC for rotaviruses in Belgium (**Figure**  
207 **2.B**).

208

209





210  
 211 **Figure 2. A.** Monthly enterovirus positive (including rhinoviruses) cases at the university  
 212 hospital laboratory (brown bars) and average qPCR CT values per month (black line) measured  
 213 from indoor air samples collected at the daycare center coupled with heatmap (below)  
 214 created using data from shotgun metagenomics. **B.** Monthly rotavirus cases nationwide  
 215 coupled with inverted bar plot of aligned reads (log10 scale) per month.  
 216

217 **5. Untargeted shotgun metagenomics allows detailed phylogenetic analyses,**  
 218 **and discrimination of distinct viral variants**

219  
 220 In addition to the detection (absence or presence) of a particular pathogen, further molecular  
 221 characterization of viral genomes can be important for public health purposes, such as  
 222 subtyping or identifying mutations responsible for resistance against therapies. Therefore, we  
 223 aimed to further leverage the obtained NGS data for the molecular characteristics of viral  
 224 genomes found in indoor air samples. The data allowed us to reconstruct several partial and  
 225 (near) complete genomes of several skin-associated (**Figure 3.A-C**) respiratory (**Figure 3.D-F**)  
 226 and enteric viruses (**Figure 3. G-H**) as well as multiple near-complete genomes of AAV-2  
 227 (**Figure 3.I**). Moreover, various animal (bird and cat) viruses (**Figure 3.J-M**) and a complete  
 228 genome of a novel densovirus were identified (see below). These obtained genomic data can  
 229 further be used for phylogenetic analysis, as exemplified for the partial RSV genome data and  
 230 the near-complete genomes of bocaviruses to reliably study their taxonomy and evolutionary  
 231 relationships (**Figure 3.N-O**). The partial genome of RSV clustered closely with RSV-B  
 232 sequences, whereas bocavirus genomes showed over 99% similarity to Human bocavirus 1  
 233 (OP255990) and human bocavirus 3 (NC\_012564), reference sequences. Further examples of  
 234 phylogenetic analyses for AAV-2, WU and MW polyomaviruses, HPV-151, and the partial  
 235 genomes of felis domesticus papillomavirus, a novel densovirus, a divergent canary  
 236 polyomavirus, and human astrovirus 4, are provided in **Supplementary information 2 S2**.

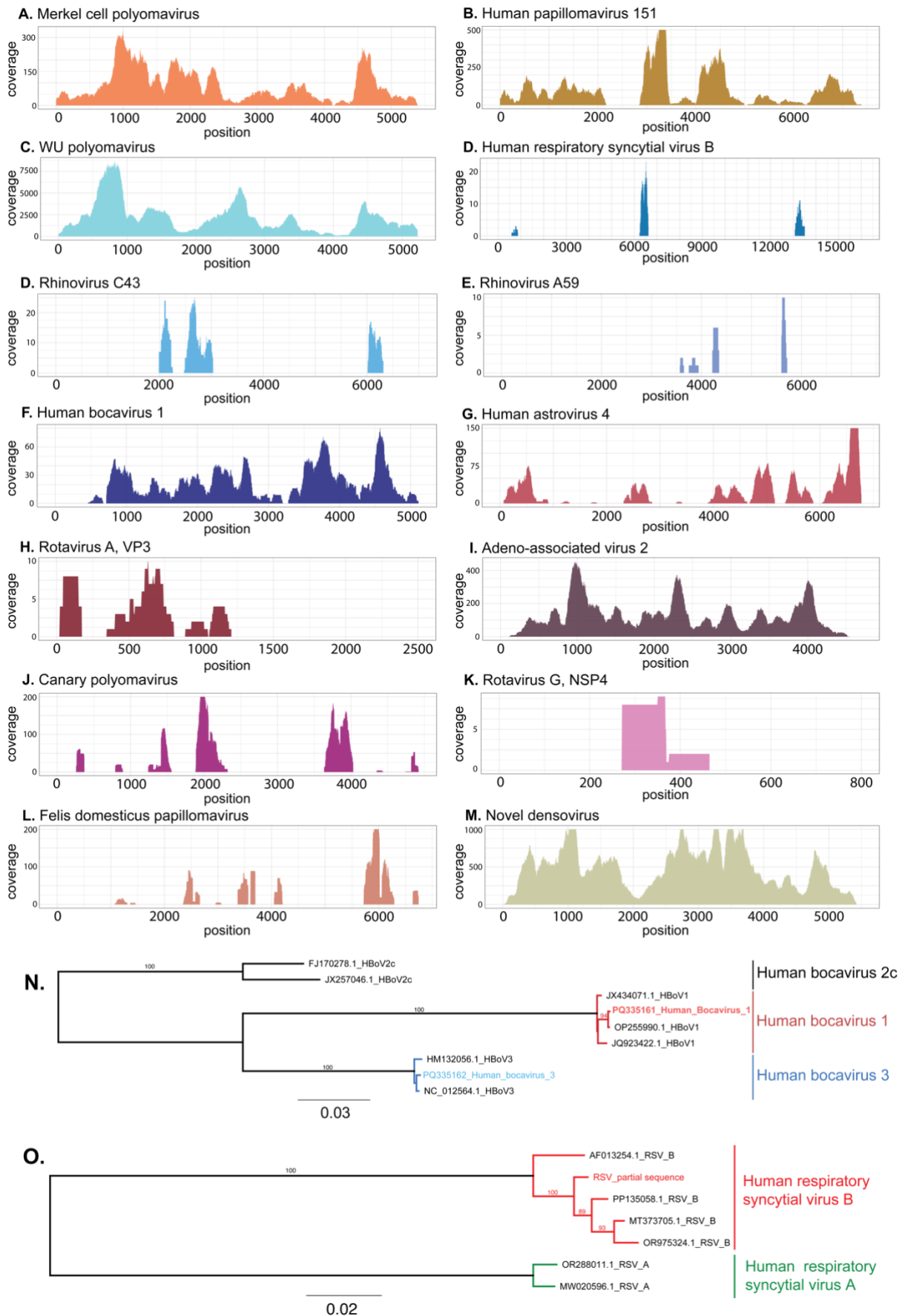
237  
 238 Interestingly, we detected WU polyomavirus in the daycare center for the first time at the end  
 239 of May, and it was consistently present for nearly two months with high abundance, until the  
 240 start of the summer break (**Figure 1.A**). This prompted us to investigate whether this  
 241 prolonged detection was due to viral shedding of a single strain by one or more individuals or  
 242 if there were multiple introductions of distinct viruses. Therefore, we reconstructed four near-

243 complete genomes of WU polyomavirus from samples collected between the end of May and  
244 July (May 30th, June 15th, June 22<sup>nd</sup> and July 6th). We show that the viral genomes collected  
245 at the first two time points are identical, whereas the third sample showed 12 mutations,  
246 suggesting the introduction of a second WU polyomavirus variant into the daycare center  
247 (mutations in different variants and read level analysis can be found in **Supplementary**  
248 **information 2. S3**). Finally, our in-depth read-level analysis suggests that both of these  
249 variants were present in the 4<sup>th</sup> sample.

250

251





252  
 253  
 254

**Figure 3.A-M.** Coverage plots of viruses found in indoor air samples. The y-axis for each panel is different and shows the depth of coverage, while the x-axis shows nucleotide positions. **N-**

255 **O.** Phylogenetic trees of 2 Human bocaviruses and RSV found in indoor air samples. Trees are  
 256 constructed by the maximum likelihood method and numbers above the branch points denote  
 257 the confidence level of the relationship of the paired sequences determined by bootstraps  
 258 (n=100).

259

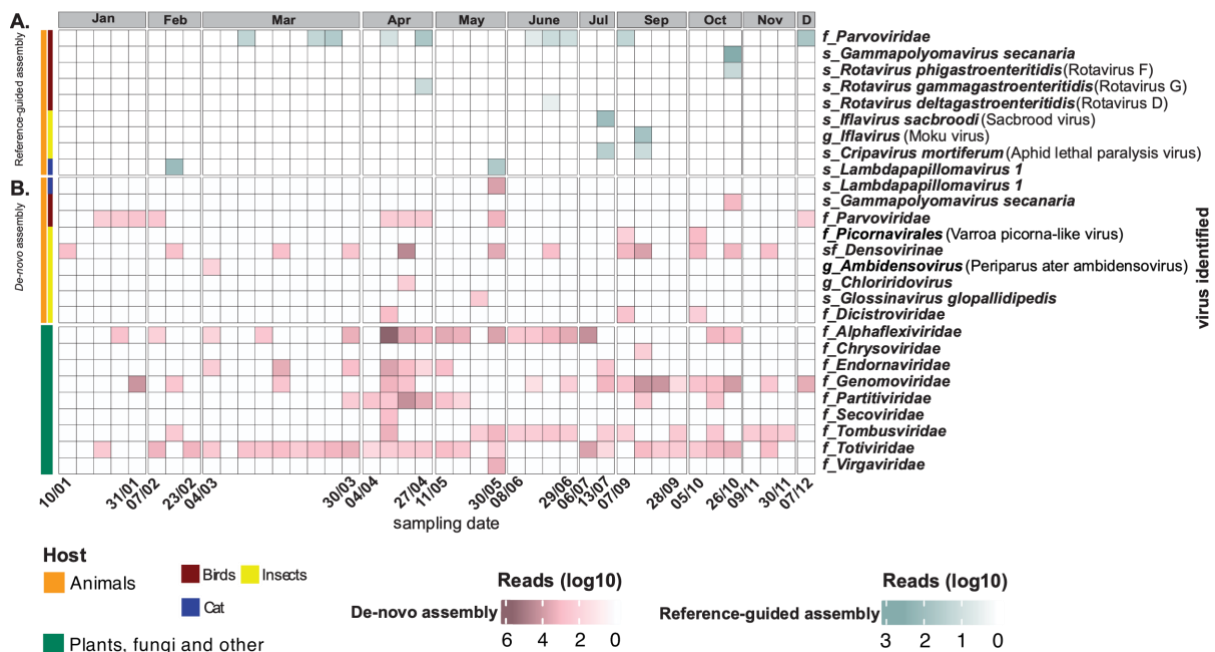
260 **6. Untargeted shotgun metagenomics on indoor air samples can reflect**  
 261 **temporal dynamics of both the indoor and outdoor environment**  
 262

263

263 Next, we focused on the air virome linked to other entities such as plants, animals (including  
 264 insects), and fungi. Although contigs identified as bacteriophages were also abundantly  
 265 identified, due to their very short average contig length (~1kb), we did not further explore  
 266 them.

267

268 Untargeted deep sequencing, combined with reference-based assembly and *de novo*  
 269 assembly (**Figure 5.A**), allowed us to detect insect and animal-associated viruses in 29 out of  
 270 42 samples (69%) (**Figure 4**). Avian species-infecting viruses, such as avian rotaviruses  
 271 (detected in April, June, and October), and *Gammampolyomavirus Secanaria* (October) were  
 272 detected. Additionally, honeybee-related viruses, such as Sacbrood virus and Moku virus  
 273 (**Figure 4.A**), were identified in July and September, along with novel and divergent insect  
 274 infecting densovirus throughout the year (in 12 out of 42 samples). As expected, divergent  
 275 (and novel) densovirus were only identified by *de novo* assembly, while viruses with low  
 276 read counts but high similarity to known references, such as avian rotaviruses, were only  
 277 identified by reference-based assembly.



278

279 **Figure 4.** Animal (except humans) and plant viruses. On the x-axis, the first and last sample  
 280 collection in each month is visible and the y-axis represents taxonomical level of viruses. **A.**  
 281 Animal viruses identified by reference guided assembly. **B.** Animal and plant viruses identified

282 by *de novo* assembly. Different taxonomical level of identifications abbreviated as follows: f:  
283 family g: genus, s: species, sf: subfamily.

284

## 285 **7. Non-human infecting viruses were more prevalent when the windows were** 286 **open**

287 A diverse array of plant and fungi infecting viruses were identified in 39/42 samples collected  
288 throughout the year. The identified viruses belonged to the families *Totiviridae*  
289 (predominantly fungi-infecting, n=27), *Tombusviridae* (plant-infecting, n=15), and 3 families of  
290 viruses known to infect both plants and fungi: *Genomoviridae* (n=17), *Partitiviridae* (n=9) and  
291 *Alphaflexiviridae* (n=18) (**Figure 4.B**, green panel and **Figure 5.B**). The majority of animal-  
292 associated viruses were detected between April and October (**Figure 4.B**, orange panel),  
293 during the period when windows were open during the air collection periods (metadata can  
294 be found in **Supplementary information 2. S4**), suggesting a link to the outside environment.  
295 Next, we investigated the seasonality of human-infecting and non-human infecting viruses.  
296 Given that the daycare center typically kept its windows open from April to October, we  
297 categorized detections into two periods: April-October and the rest of the year. This  
298 categorization revealed a significantly higher prevalence of human-infecting viruses during the  
299 closed-window periods, while non-human-infecting viruses were more prevalent during the  
300 open-window period ( $p < 0.01$ , **Figure 5.C-D**).

301

302 With respect to fungal viruses, we noted that some of the more abundantly identified viral  
303 genome sequences from the family *Totiviridae* were most closely related (~80% amino acid  
304 identity) to viruses known to infect fungi associated with humans: “*Geotrichum candidum*  
305 infecting totivirus” and “*Malassezia-restricta* infecting totivirus” (**Figure 5.B**). Overall,  
306 totiviruses were consistently detected in February-March (10 out of 12 samples) and July-  
307 October (7 out of 9 samples) (**Figure 5.B**). Although our NGS protocol is aimed at enriching  
308 for virus sequences, the *de novo* approach also identified contigs classified as *Malasseziaecea*,  
309 and reads counts mapping to these contigs correlated strongly with the relative abundance of  
310 totiviruses (**Figure 5.B-E**).

311

312

313

314

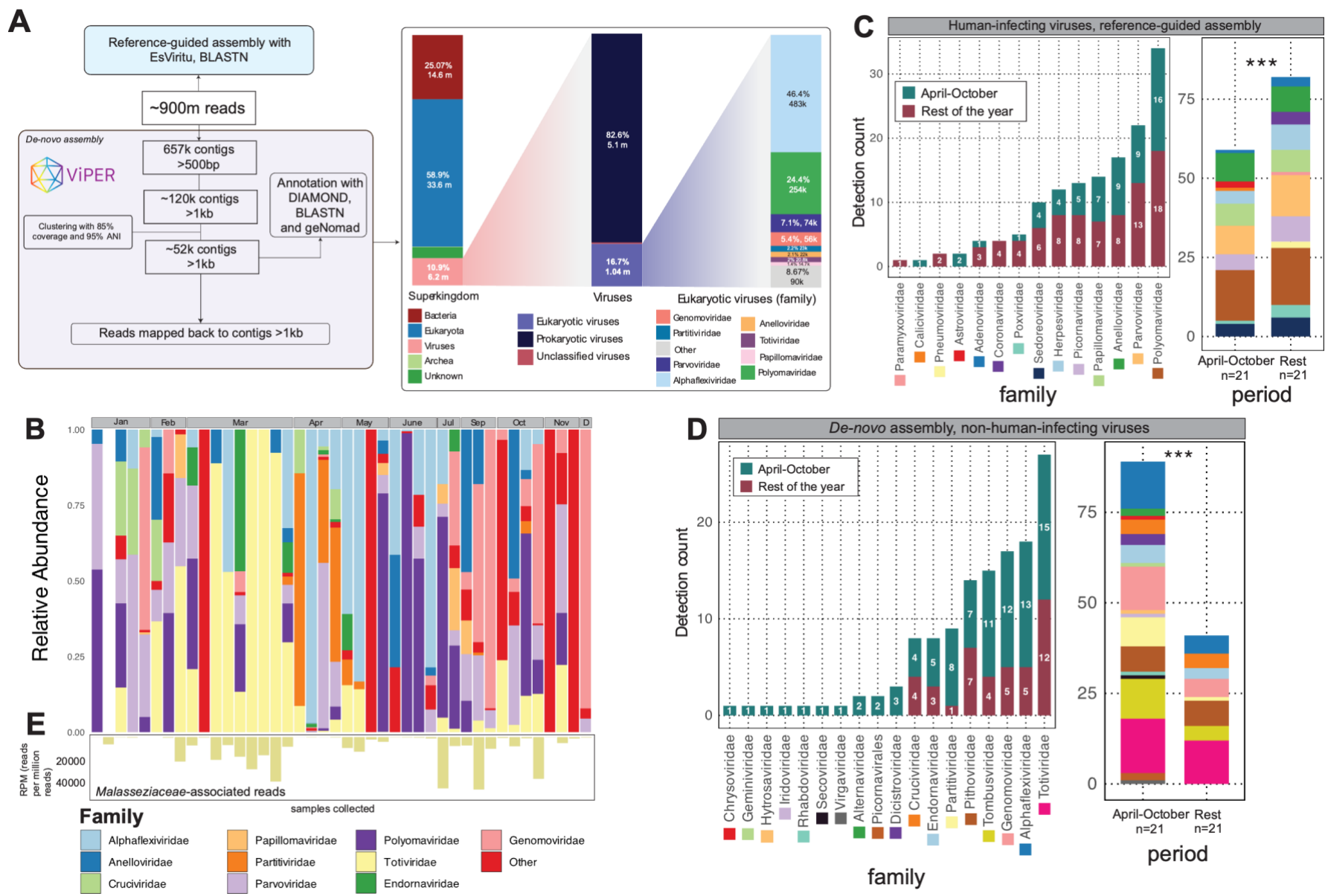
315

316

317

318

319



**Figure 5.A.** Bioinformatics pipeline used for reference-based and *de-novo* assembly-based virus identification. **B.** Relative abundance of different viral families detected in each sample. **C.** Number of detections of human-infecting viruses for each family (left panel) during period with open or closed windows. Aggregated detection counts of human-infecting viruses for the periods between April-October and rest of the year (right panel),  $\chi^2=6.86$ ,  $df=1$ ,  $p<0.01$  ( $p=0.0087$ ). **D.** Number of detections of non-human-infecting viruses for each family (left panel). Aggregated detection counts of human-infecting viruses for the periods between April-October and rest of the year (right panel),  $\chi^2=33.985$ ,  $df=1$ ,  $p<0.001$ . **E.** Reads mapping back to contigs of family *Malasseziaceae*, as RPM (reads per million).

## 321 Discussion

322 Studies utilizing multi-pathogen qPCR and metagenomics on indoor air samples have  
323 highlighted the potential for tracking various viruses including respiratory, enteric and plant  
324 associated viruses<sup>10–13,24,29</sup>. Building on these studies, we employed a longitudinal approach  
325 over one year at a day care center using the NetoVIR protocol for VLP enrichment combined  
326 with short-read untargeted sequencing. Parents were advised to keep their children at home  
327 if they exhibited any symptoms of illness such as fever, runny nose, or diarrhea. This led to our  
328 study population consisting of "healthy" individuals. Throughout the 12-month period, we  
329 successfully detected a diverse array of human-associated viruses, including respiratory,  
330 enteric, and skin viruses, alongside non-human viruses linked to animals, plants, and fungi.

331  
332 *Rosario et al.* identified Merkel cell polyomavirus, MW polyomavirus and papillomaviruses in  
333 air filter samples<sup>29</sup>. Similarly, polyomaviruses were the most abundant human viruses  
334 detected in our study. While WU polyomavirus found in our study has been reported in  
335 children with respiratory symptoms or immunocompromised individuals, other viruses have  
336 been reported among seemingly healthy individuals, such as Merkel cell polyomavirus and  
337 MW polyomavirus<sup>30,31</sup>. Although our understanding of polyomavirus transmission, cell  
338 tropism and its potential role in disease is currently limited, overall high prevalence and  
339 abundance of polyomaviruses suggests circulation of polyomaviruses among young  
340 individuals, as previously suggested<sup>32</sup>. Interestingly, alongside the human skin viruses, the  
341 most abundant fungal viruses were related to commensal human skin fungi from the family  
342 *Malasseziaceae*, providing evidence for the high abundance of human skin-associated  
343 microorganisms in the environment.

344  
345 Recent studies have investigated the potential of indoor air surveillance to monitor specific  
346 viral pathogen dynamics and epidemiology in various settings, including hospitals, offices, and  
347 residential environments<sup>10,12,24,33–36</sup>. All these studies tracking epidemiology through air relied  
348 on targeted tests requiring specific viral genomic knowledge, while our study employed a non-  
349 targeted approach. The longitudinal design of our study, in combination with available  
350 epidemiological data from the Leuven university hospital, allowed us to use rhinovirus and  
351 rotavirus as examples to showcase the power of air sampling in epidemiological surveillance,  
352 especially for asymptomatic circulation in a setting with no clinically ill individuals. In addition,  
353 our NGS-based untargeted approach allowed us to subtype rhinoviruses, showing the co-  
354 circulation of various rhinoviruses throughout the year.

355  
356 *Minor et al.* employed viral metagenomics for indoor air surveillance, detecting Influenza C in  
357 a preschool and providing genomic data for this frequently untested virus<sup>24</sup>. In our study, we  
358 have corroborated their findings by identifying multiple complete or partial genomes of  
359 viruses, such as Varicella zoster virus, papillomaviruses, and WU polyomavirus, which may be  
360 associated with respiratory or skin diseases but are not currently included in the routine qPCR  
361 panels for respiratory diseases in Belgian hospitals. Additionally, our study identified and was

362 able to subtype bocavirus, rhinovirus, parainfluenza, and RSV-B, without requiring any  
363 additional molecular testing. In cases of high viral abundance (such as respiratory-associated  
364 WU polyomavirus), our methods are even able to distinguish the co-circulation of closely  
365 related variants. This underscores the potential of untargeted metagenomics in uncovering a  
366 broader range of viral diversity than targeted qPCR panels, highlighting its utility in monitoring  
367 both common and lesser-known viruses that may otherwise go undetected. Furthermore,  
368 utilizing *de novo* and reference-guided assembly in combination, we discovered novel and  
369 divergent viruses, demonstrating that we can use indoor air as a virus surveillance medium  
370 for pandemic preparedness.

371

372 Beyond surveillance of human viruses, indoor air and environmental surveillance with  
373 metagenomics could also be of interest for animal farms and greenhouses. *Kwok et al.*  
374 analyzed dust samples and pooled chicken feces from farms using shotgun metagenomics,  
375 identifying multiple viruses from the families *Parvoviridae*, *Picornaviridae*, *Caliciviridae*, and  
376 *Astroviridae* that matched those found in the feces from the same location<sup>37</sup>. In our study, we  
377 unexpectedly detected avian viruses (rotaviruses and a divergent polyomavirus), a cat  
378 papillomavirus, and several insect-infecting viruses, including a novel densovirus, in an animal-  
379 free environment. We propose that the animal viruses mentioned are highly resilient in the  
380 environment and may be transported indoors by insects, air currents, or even on human skin,  
381 where they can subsequently be shed. Additionally, the presence of multiple plant viruses  
382 demonstrates that indoor air can serve as an effective medium for detecting pathogens that  
383 typically require labor-intensive sampling from various crops. Therefore, we propose that  
384 indoor air surveillance in strategic animal and plant production locations can be invaluable for  
385 revealing epidemiological trends and potential outbreaks of viruses.

386

387 It was previously reported that indoor air shotgun metagenomics with other methods resulted  
388 in 1% to 3.2% of reads that could be assigned to viruses<sup>38</sup>, underscoring the low viral biomass  
389 of indoor air samples. However, using the NetoVIR protocol, 10.9% of reads mapped to viral  
390 contigs, primarily phages, aligning with earlier findings<sup>39</sup>. Interestingly, shotgun  
391 metagenomics identified a broader diversity of enteric viruses, such as rotavirus and  
392 sapovirus, than the enteric qPCR panel, whereas the respiratory qPCR panel detected  
393 rhinovirus, RSV, adenovirus, and parainfluenza more consistently than metagenomics. These  
394 differences likely reflect the distinct sensitivities of the qPCR panels, which are tailored for  
395 detecting specific viral loads in defined sample types. The viral load in indoor air is likely at or  
396 near detection thresholds, with commonly circulating human viruses comprising less than  
397 0.005% of total reads, as reported by *Prussin et al.*<sup>39</sup>. In our study, we confirm this low viral  
398 load of human-associated viruses in indoor air with the exception of skin- and respiratory-  
399 associated polyomaviruses. Overall, we suggest that our method is robust for indoor air  
400 surveillance and combining various bioinformatics approaches can enhance our  
401 understanding of the air virome.

402



403 Given the longitudinal nature of our sample collection, it is possible to discuss seasonality of  
404 viruses. *Prussin II et al.* employed passive indoor air sampling and shotgun metagenomics  
405 throughout the year in a daycare center. They found human-associated viruses to be much  
406 more diverse and dominant in the winter, while the summertime virome contained a higher  
407 proportion and diversity of plant-associated viruses<sup>39</sup>. Similarly, we also observed a higher  
408 prevalence of animal- or plant-associated viruses when windows were open, particularly  
409 during the warmer months from April to October and respiratory and enteric viruses mostly  
410 during the wintertime when windows were closed. Therefore, we suggest that active indoor  
411 air sampling can reflect dynamics of both the inside and outside environment and can be a  
412 way to track indoor air quality.

413  
414 Overall, untargeted viral metagenomics offers a strong and complementary tool for pathogen  
415 surveillance via indoor air, particularly in cases of suspected asymptomatic shedding. Future  
416 research should validate this approach at sentinel locations for skin, enteric, and respiratory  
417 viruses in targeted settings, ideally incorporating clinical infection data of individuals. Our  
418 findings underscore the ability of our method to investigate pathogenic animal and plant  
419 viruses, in relevant settings and natural environments, utilizing air sampling techniques and  
420 combining *de novo* and reference-based bioinformatics approaches.

421  
422 Our study is subject to a number of limitations. Primarily, it was conducted at a single location,  
423 without access to information regarding the health status or medical history of individuals in  
424 the vicinity. Nonetheless, it was assumed that individuals were unlikely to exhibit severe  
425 respiratory or gastroenteritis-related symptoms, as it was recommended that families keep  
426 children at home if they displayed such symptoms. Secondly, our inability to acquire complete  
427 genomes for numerous viruses present, which may stem from methodological constraints of  
428 sequencing and low viral load in the air. Additionally, our results are not confirmed by culture-  
429 based techniques to determine if the detected viral genomes are indicative for infectious  
430 viruses. This would be important to know, with respect to the potential of airborne  
431 transmission of skin-associated and enteric viruses. On the other hand, our study has several  
432 significant strengths. Firstly, we employed a longitudinal approach over a 12-month period,  
433 allowing for monitoring of seasonal dynamics. Secondly, we utilized advanced viral  
434 enrichment methods, which enhanced our ability to detect even low-abundance viral species  
435 with greater sensitivity. Additionally, we conducted in-depth bioinformatics analysis,  
436 combining *de novo* assembly and reference-guided assembly techniques. This approach  
437 resulted in the identification of not only known human-associated viruses with very low  
438 abundance but also novel and divergent viruses, which can translate as "pathogen X" in the  
439 future, indicating its potential significance for pandemic preparedness efforts.

440

441

## 442 **Conclusion**

443

444 Our study highlights indoor air sampling as a robust method for surveillance of both human-  
445 associated and non-human-associated viruses. We suggest that this method, implemented in  
446 sentinel locations and targeted populations, can be complementary to traditional patient-  
447 based surveillance methods.

448

449

## 450 **Methods**

451

### 452 **Sample collection**

453

454 Between January 2022 and December 2022, 90 indoor air samples were collected from a  
455 daycare center in Leuven, hosting children up to 3 years old (on average, 15 children and 3  
456 adults). We aimed to select at least 4 samples from each month (weekly), however, after the  
457 respiratory PCR panel, sufficient material of 42 samples was available for further analysis with  
458 metagenomics and an enteric PCR panel. Sampling was done between 9.00-11.00 (AM) for 2  
459 hours, which are active hours at the daycare center, with children being present in the main  
460 area. When the daycare center was closed, sample collection was not performed. For  
461 sampling, an AerosolSense active air sampler was used<sup>10</sup>. Ambient air was sampled with a rate  
462 of 200L/min through a vertical collection pipe and impacted onto the AerosolSense Capture  
463 Media (Thermo Fisher Scientific, Waltham, MA). CO<sub>2</sub> concentrations were measured using a  
464 remote climate sensor (Elsys ERC CO<sub>2</sub>, Umeå, Sweden; placed next to the air sampler). Natural  
465 ventilation was also estimated by a Likert scale (no natural ventilation, one window open, door  
466 open, multiple windows open, door and window open).

467

### 468 **Sample processing**

469

470 After removing the standard cartridges from the sampler, they were transported to the lab on  
471 the same day they were collected. Samples were dissolved in Universal Transport Medium  
472 (UTM), to be used for respiratory and enteric multiplex qPCR panels as well as metagenomics  
473 (see below). Samples were stored at 4 degrees until processing with the respiratory multiplex  
474 qPCR panel. If they required storage over the weekend, they were frozen at -80 degrees  
475 Celsius. Processing with the enteric multiplex qPCR panel and shotgun metagenomics was  
476 conducted in February 2023, on samples stored at -20 degrees Celsius, with 2 or less freeze  
477 thaw cycles. As negative controls, UTM (used to dissolve samples) was processed together  
478 and sequenced as deep as samples.

479

### 480 **Respiratory and enteric multiplex qPCR panels**

481

482 Samples dissolved in UTM were subjected to multiplex quantitative PCR (qPCR) testing. The  
483 detailed protocols for these procedures have been previously described<sup>10,40</sup>. The respiratory  
484 panel is designed to detect twenty-two viruses and seven bacterial pathogens and was applied  
485 on 90 samples when they were collected. In contrast, the enteric panel detects seventeen  
486 non-viral pathogens (four parasites and thirteen bacteria) and six viruses, specifically Human  
487 adenovirus F40/F41, Norovirus GI, Norovirus GII, Rotavirus A, Astrovirus, and Sapovirus  
488 (genogroups GI, GII, GIV, and GV). The enteric panel was applied to the same 42 samples used  
489 for shotgun metagenomics (see below).

490

### 491 **Shotgun metagenomics**

492

493 Samples were subjected to *Novel Enrichment Technique of Viromes* (NetoVIR)<sup>41</sup>. They were  
494 homogenized by vortexing for 10 seconds and homogenous samples were centrifuged at  
495 17,000 g for 3 minutes and filtered with a 0.8µm PES (Polyethersulfone) centrifugal filter, in  
496 order to eliminate eukaryotic cells and bacteria. Free floating nucleic acids were removed by  
497 benzonase and micrococcal nuclease while being incubated for 2 hours at 37°C. DNA/RNA  
498 Amplification was done using a modified Whole Transcriptome Amplification 2 (WTA2) Kit to  
499 amplify both DNA and RNA. To obtain DNA amplification, the initial denaturation step was  
500 performed at 95°C instead of 70°C. Library preparation was done using a Nextera XT DNA  
501 Sample Preparation kit and sequencing was performed on an NovaSeq™ platform (Illumina).

502

### 503 **Clinical data collection**

504

505 We collected the results of respiratory panel tests for enterovirus (including rhinovirus)  
506 performed at the clinical laboratory of University Hospitals Leuven during the same period as  
507 the air sample collection. The PCR target also included non-rhinovirus enteroviruses. However,  
508 we assumed that the majority of enteroviruses detected in individual respiratory samples  
509 were rhinoviruses. Clinical respiratory qRT-PCR panels are performed according to specific  
510 clinical indications. In immunocompetent individuals, these panels are utilized for respiratory  
511 infections requiring intensive care admission or those unresponsive to initial therapy for a  
512 respiratory infection. For immunocompromised patients, respiratory panels are more  
513 commonly used when lower respiratory infections are suspected.

514 For rotavirus A infections in the age group of 0 to 3 years old, we used data from the National  
515 Reference Center (NRC) for rotavirus in Belgium over the same period. Samples of the NRC  
516 were collected through nationwide laboratory system on a voluntary basis and under the  
517 regulations defined by the Royal Decree of 09/02/2011.

518

### 519 **Bioinformatics analyses**

520 We employed *de novo* assembly and reference-guided assembly on the reads obtained from  
521 sequencing. Analyses were conducted on VSC KU Leuven (The Flemish Supercomputer Center)  
522 servers.

523

#### 524 **Reference-guided assembly**

525 To identify human and animal-associated viruses, we utilized EsVirtu<sup>42</sup> (v0.2.3) and BLASTN.  
526 EsVirtu employs a database of all human and animal viruses in GenBank as of November 16th,  
527 2022 (EsVirtu DB v2.0.2), curated to include one representative genome per 95% average  
528 nucleotide identity cluster. Reads underwent quality control, adapter trimming, and mapping  
529 to reference sequences with 90% identity and 90% coverage. Viruses were considered present  
530 if the consensus sequence exceeded 200 base pairs or 10% of the reference genome and at  
531 least 2 reads mapped to it. Consensus sequences were obtained using Samtools consensus<sup>43</sup>  
532 and subsequently verified using BLASTN. Reference-guided assembly only showed HPV-21  
533 with 39 reads in one negative control, which resulted in a 236-nucleotide long consensus  
534 sequence, found to be exactly the same as in one other higher viral load sample with same  
535 virus, and was therefore classified as cross-contamination from this sample.

536

#### 537 **De-novo assembly**

538 For *de novo* assembly of viral genomes (or contigs), we utilized the Virome Paired-End Reads  
539 pipeline (ViPER)<sup>44</sup>. The pipeline processed quality-controlled and adapter-trimmed reads,  
540 performing *de novo* triple-assembly using metaSPAdes<sup>45</sup>. Contigs were then clustered at 95%  
541 identity and 85% coverage to reduce redundancy, first within individual samples and then  
542 among all samples. Reads were mapped back to contigs >1kbp, considering those with over  
543 50% horizontal coverage present in each sample. Contig identification was subsequently  
544 performed using DIAMOND<sup>46</sup> for eukaryotic viruses, geNomad<sup>47</sup> for bacteriophages, and  
545 BLASTN for identification of genetic material associated to other entities (i.e., fungi, plant,  
546 human). For *de novo* assembly, all contigs present in negative controls were treated as  
547 contaminants and removed before downstream analyses.

548

#### 549 **Statistical analyses**

550 Statistical analyses for the difference of detections in open and closed window times of non-  
551 human infecting and human-infecting viruses were conducted using the chi-square test and  
552 performed in RStudio. The analysis assessed the association between seasonality and the  
553 presence of these viruses. A p-value of less than 0.05 was considered statistically significant.

554

#### 555 **Data and code availability**

556 Raw data, excluding human reads, is available at the Sequence Read Archive (SRA) with  
557 Bioproject number PRJNA1158979. Codes used to conduct the analyses and produce figures  
558 are available on [https://github.com/Matthijnssenslab/Indoorair\\_daycare](https://github.com/Matthijnssenslab/Indoorair_daycare).

559

#### 560 **Acknowledgements**

561 Mustafa Karatas is supported by a Research Foundation Flanders (FWO) fundamental research  
562 scholarship (number: 11P7I24N). We would like to thank Pieter Wets and Steven Traets for  
563 their help in collection of samples. Additionally, we thank the Rotavirus National Reference

564 Center and National Reference Center for Respiratory Pathogens for their support throughout  
565 this research.

566

567 **Ethical approval**

568 The study was approved by the Ethics Committee Research UZ/KU Leuven (S66518,  
569 B3222022000873). No informed consent was required from occupants of the sampled  
570 environments.

571

572 **Competing interests**

573 The authors declare no competing interests.

574

575 **Author contributions**

576 Conceptualization was carried out by JM, EA, EK and MK. The project was managed by EK, LB,  
577 JR and EA. CE, BC, MK, and EK performed sample collection and processing. Data analysis,  
578 visualization, and drafting of the initial manuscript were conducted by MK. Supervision was  
579 provided by JM, EA, and EK. All authors reviewed and approved the final manuscript.

580

581 **References**

- 582 1. Abat, C., Chaudet, H., Rolain, J.-M., Colson, P. & Raoult, D. Traditional and syndromic  
583 surveillance of infectious diseases and pathogens. *Int. J. Infect. Dis.* **48**, 22–28 (2016).
- 584 2. Leifels, M. *et al.* The one health perspective to improve environmental surveillance of  
585 zoonotic viruses: lessons from COVID-19 and outlook beyond. *ISME Commun.* **2**, 107  
586 (2022).
- 587 3. Vieira, C. B. *et al.* Viruses Surveillance Under Different Season Scenarios of the Negro  
588 River Basin, Amazonia, Brazil. *Food Environ. Virol.* **8**, 57–69 (2016).
- 589 4. Iaconelli, M. *et al.* One-year Surveillance of Human Enteric Viruses in Raw and Treated  
590 Wastewaters, Downstream River Waters, and Drinking Waters. *Food Environ. Virol.* **9**,  
591 79–88 (2017).
- 592 5. Moresco, V., Oliver, D. M., Weidmann, M., Matallana-Surget, S. & Quilliam, R. S. Survival  
593 of human enteric and respiratory viruses on plastics in soil, freshwater, and marine  
594 environments. *Environ. Res.* **199**, 111367 (2021).
- 595 6. Kilaru, P. *et al.* Wastewater Surveillance for Infectious Disease: A Systematic Review. *Am.*  
596 *J. Epidemiol.* **192**, 305–322 (2023).
- 597 7. Amman, F. *et al.* Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national  
598 scale. *Nat. Biotechnol.* **40**, 1814–1822 (2022).
- 599 8. Jahn, K. *et al.* Early detection and surveillance of SARS-CoV-2 genomic variants in  
600 wastewater using COJAC. *Nat. Microbiol.* **7**, 1151–1160 (2022).
- 601 9. Wastewater monitoring comes of age. *Nat. Microbiol.* **7**, 1101–1102 (2022).
- 602 10. Raymenants, J. *et al.* Indoor air surveillance and factors associated with respiratory  
603 pathogen detection in community settings in Belgium. *Nat. Commun.* **14**, 1332 (2023).

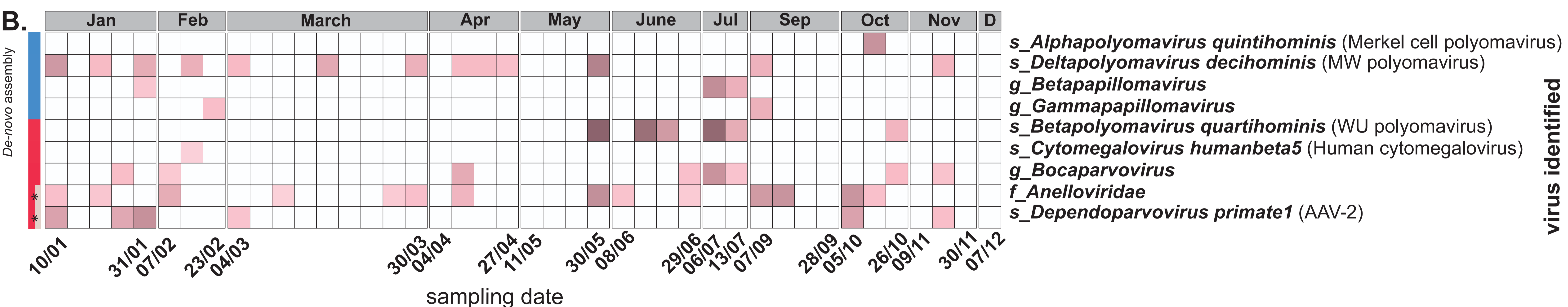
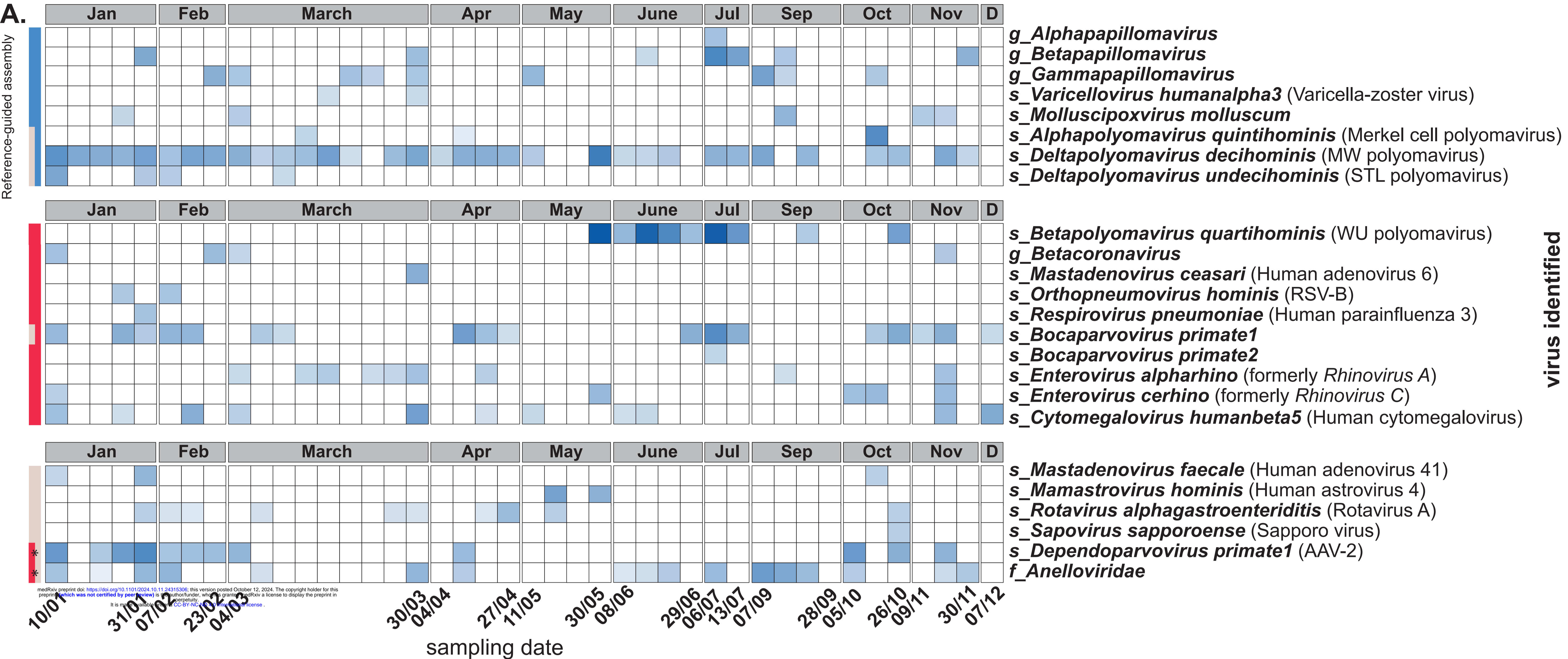


- 604 11. Ramuta, M. D. *et al.* SARS-CoV-2 and other respiratory pathogens are detected in  
605 continuous air samples from congregate settings. *Nat. Commun.* **13**, 4717 (2022).
- 606 12. Raymenants, J. *et al.* Detection of mpox virus in ambient air in a sexual health clinic.  
607 *Arch. Virol.* **168**, 210 (2023).
- 608 13. Happaerts, M. *et al.* *Centralised Air Sampling from a Ventilation System for the*  
609 *Surveillance of Respiratory Pathogens.* <https://www.ssrn.com/abstract=4596460> (2023)  
610 doi:10.2139/ssrn.4596460.
- 611 14. Sinclair, R. G., Choi, C. Y., Riley, M. R. & Gerba, C. P. Pathogen Surveillance Through  
612 Monitoring of Sewer Systems. in *Advances in Applied Microbiology* vol. 65 249–269  
613 (Elsevier, 2008).
- 614 15. Wang, G. *et al.* Optimization and evaluation of viral metagenomic amplification and  
615 sequencing procedures toward a genome-level resolution of the human fecal DNA  
616 virome. *J. Adv. Res.* **48**, 75–86 (2023).
- 617 16. Conceição-Neto, N., Yinda, K. C., Van Ranst, M. & Matthijssens, J. NetoVIR: Modular  
618 Approach to Customize Sample Preparation Procedures for Viral Metagenomics. in *The*  
619 *Human Virome* (eds. Moya, A. & Pérez Brocal, V.) vol. 1838 85–95 (Springer New York,  
620 New York, NY, 2018).
- 621 17. Van Espen, L. *et al.* A Previously Undescribed Highly Prevalent Phage Identified in a  
622 Danish Enteric Virome Catalog. *mSystems* **6**, (2021).
- 623 18. Yinda, C. K. *et al.* A single bat species in Cameroon harbors multiple highly divergent  
624 papillomaviruses in stool identified by metagenomics analysis. *Virol. Rep.* **6**, 74–80  
625 (2016).
- 626 19. Shi, C. *et al.* Stable distinct core eukaryotic viromes in different mosquito species from  
627 Guadeloupe, using single mosquito viral metagenomics. *Microbiome* **7**, (2019).

- 628 20. Deboutte, W. *et al.* Honey-bee–associated prokaryotic viral communities reveal wide  
629 viral diversity and a profound metabolic coding potential. *Proc. Natl. Acad. Sci.* **117**,  
630 10511–10519 (2020).
- 631 21. Ortega Parra, N., Zisi, Z., Vogel, E., Vos, C. & Hanssen, I. First report of infections with  
632 isolates of the Asian genotype of Cucumber green mottle mosaic virus in cucumber  
633 crops in Belgium. *Plant Dis.* PDIS-11-23-2532-PDN (2024) doi:10.1094/PDIS-11-23-2532-  
634 PDN.
- 635 22. Gauthier, N. P. G., Chorlton, S. D., Krajden, M. & Manges, A. R. Agnostic Sequencing for  
636 Detection of Viral Pathogens. *Clin. Microbiol. Rev.* **36**, e00119-22 (2023).
- 637 23. Prussin, A. J., Marr, L. C. & Bibby, K. J. Challenges of studying viral aerosol metagenomics  
638 and communities in comparison with bacterial and fungal aerosols. *FEMS Microbiol. Lett.*  
639 **357**, 1–9 (2014).
- 640 24. Minor, N. R. *et al.* Metagenomic sequencing detects human respiratory and enteric  
641 viruses in air samples collected from congregate settings. *Sci. Rep.* **13**, 21398 (2023).
- 642 25. Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L. & Buck, C. B. Merkel Cell  
643 Polyomavirus and Two Previously Unknown Polyomaviruses Are Chronically Shed from  
644 Human Skin. *Cell Host Microbe* **7**, 509–515 (2010).
- 645 26. Aldridge, R. W. *et al.* Seasonality and immunity to laboratory-confirmed seasonal  
646 coronaviruses (HCoV-NL63, HCoV-OC43, and HCoV-229E): results from the Flu Watch  
647 cohort study. *Wellcome Open Res.* **5**, 52 (2020).
- 648 27. Ramaekers, K. *et al.* Prevalence and seasonality of six respiratory viruses during five  
649 consecutive epidemic seasons in Belgium. *J. Clin. Virol.* **94**, 72–78 (2017).
- 650 28. Horvat, C. *et al.* Contribution of Infant Rhinovirus Bronchiolitis to Hospital Bed and  
651 Ventilation Use. *JAMA Netw. Open* **7**, e2355033 (2024).

- 652 29. Rosario, K., Fierer, N., Miller, S., Luongo, J. & Breitbart, M. Diversity of DNA and RNA  
653 Viruses in Indoor Air As Assessed via Metagenomic Sequencing. *Environ. Sci. Technol.* **52**,  
654 1014–1027 (2018).
- 655 30. Abed, Y., Wang, D. & Boivin, G. WU Polyomavirus in Children, Canada. *Emerg. Infect. Dis.*  
656 **13**, 1939–1941 (2007).
- 657 31. Rockett, R. J. *et al.* Acquisition of Human Polyomaviruses in the First 18 Months of Life.  
658 *Emerg. Infect. Dis.* **21**, 365–367 (2015).
- 659 32. Cook, L. Polyomaviruses. *Microbiol. Spectr.* **4**, 4.4.24 (2016).
- 660 33. Borges, J. T., Nakada, L. Y. K., Maniero, M. G. & Guimarães, J. R. SARS-CoV-2: a systematic  
661 review of indoor air sampling for virus detection. *Environ. Sci. Pollut. Res.* **28**, 40460–  
662 40473 (2021).
- 663 34. Dietz, L. *et al.* Exploring Integrated Environmental Viral Surveillance of Indoor  
664 Environments: A comparison of surface and bioaerosol environmental sampling in  
665 hospital rooms with COVID-19 patients. Preprint at  
666 <https://doi.org/10.1101/2021.03.26.21254416> (2021).  
667

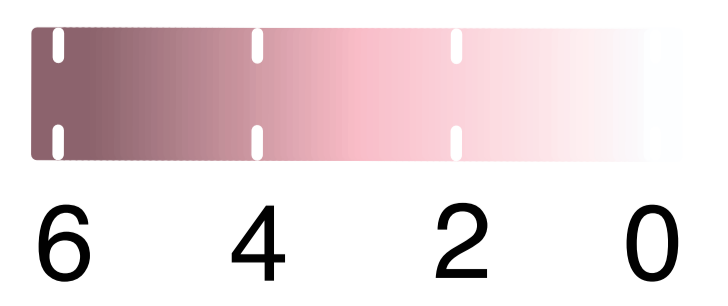




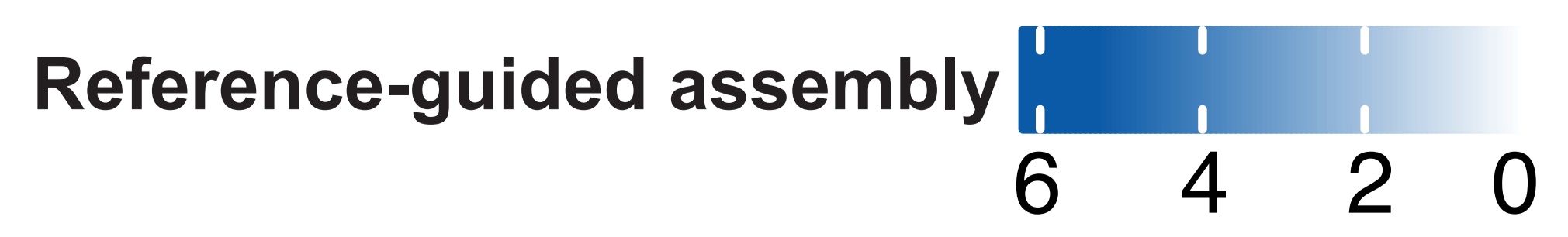
**Association type**

- Enteric
- Respiratory
- Skin-associated

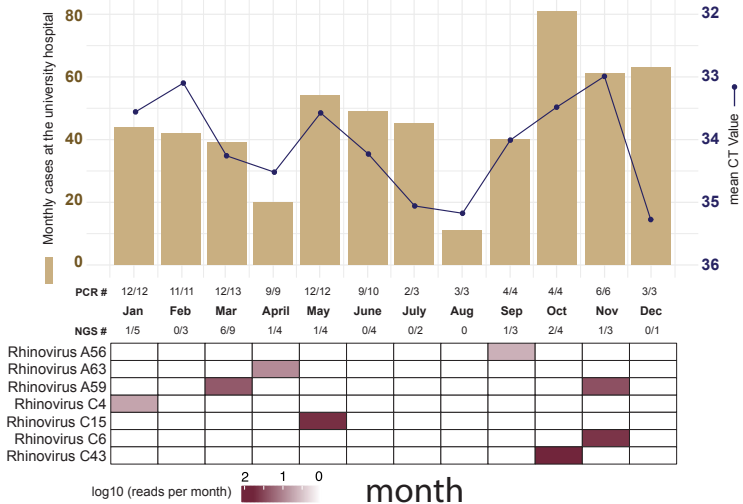
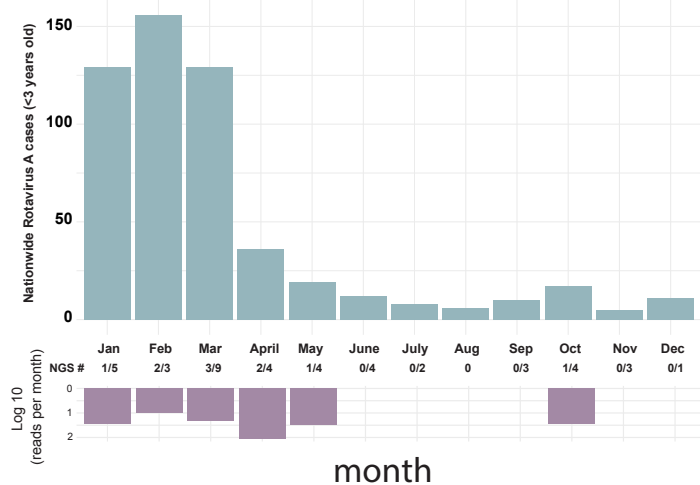
**Reads (log10)**



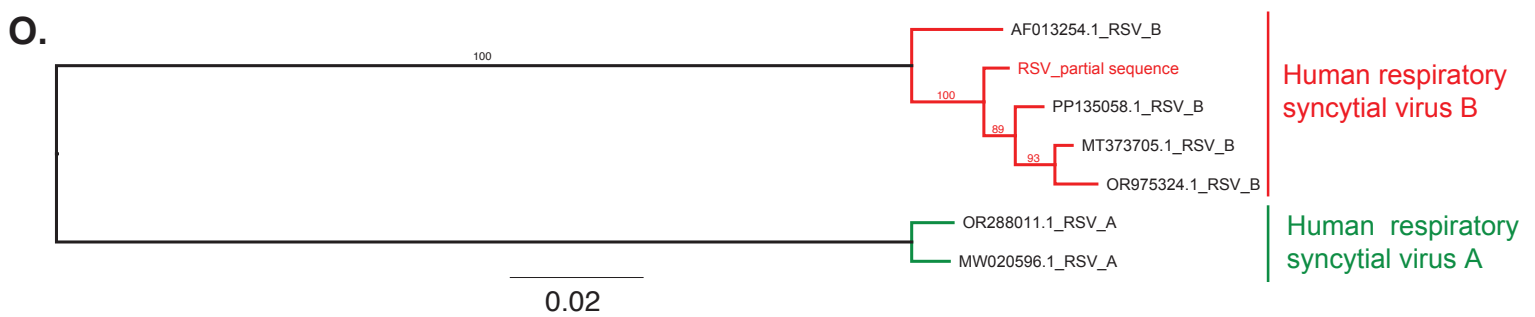
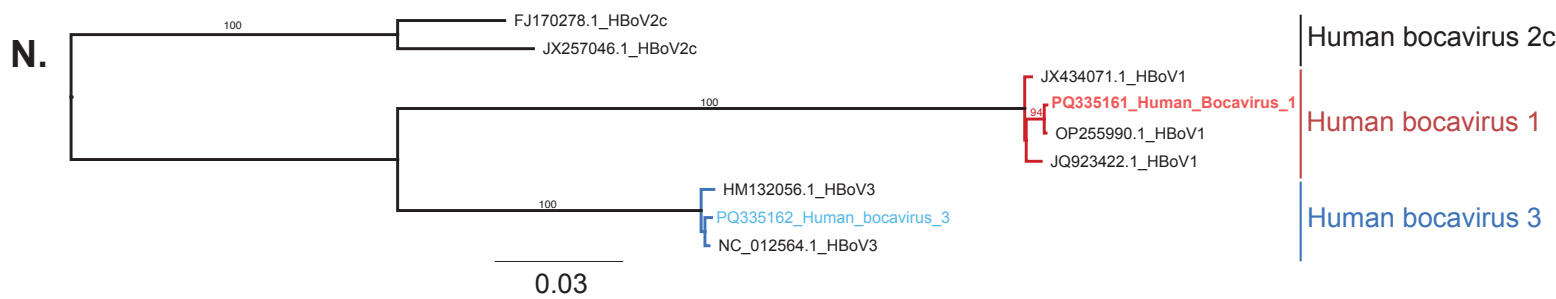
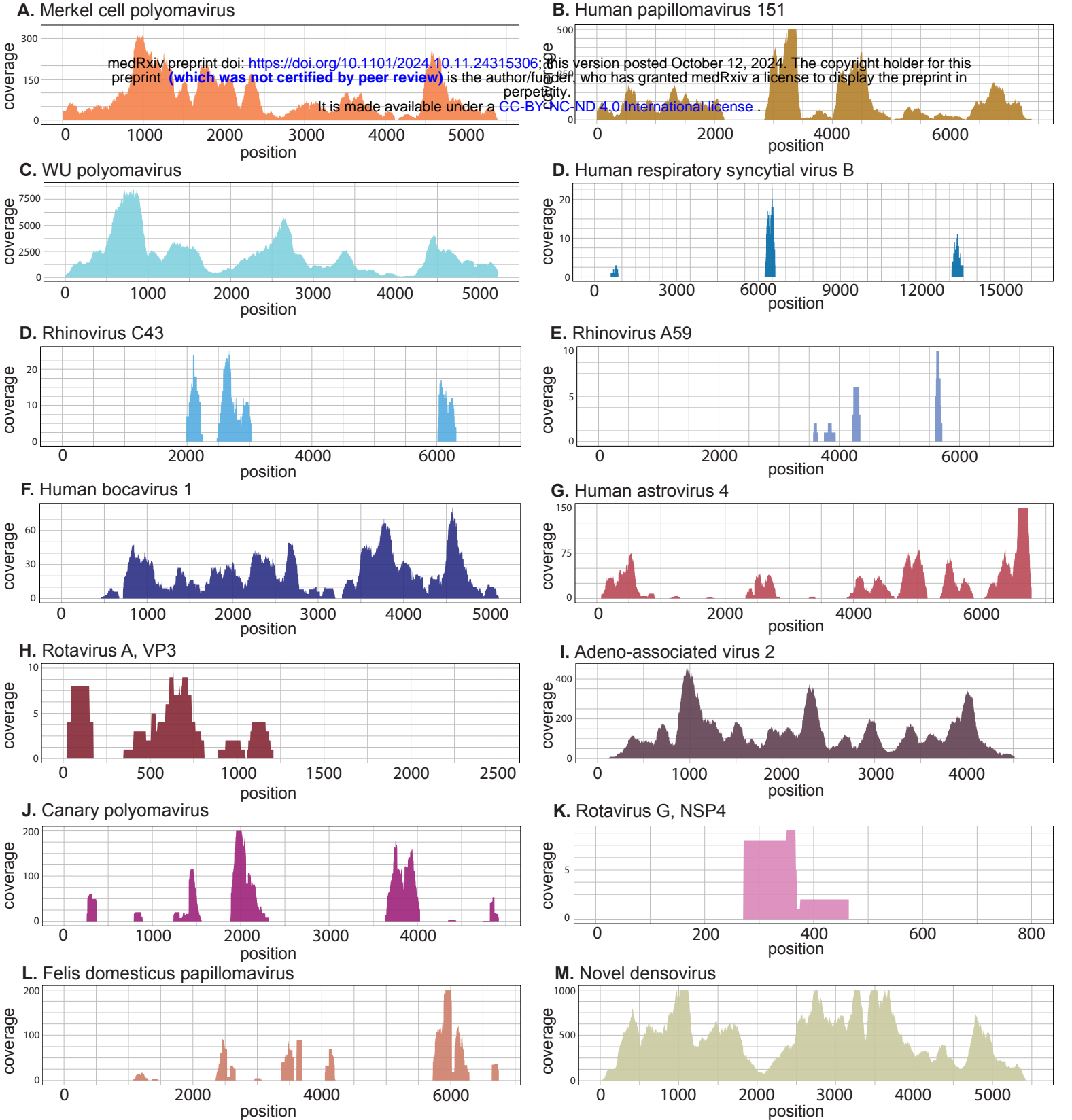
**Reads (log10)**



medRxiv preprint doi: <https://doi.org/10.1101/2024.10.11.24315306>; this version posted October 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

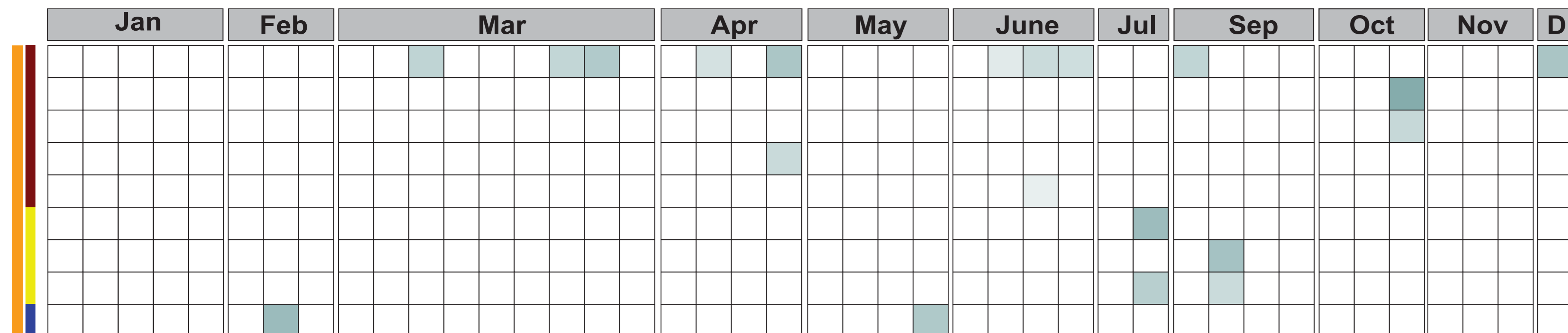
**A****B**





**A.**

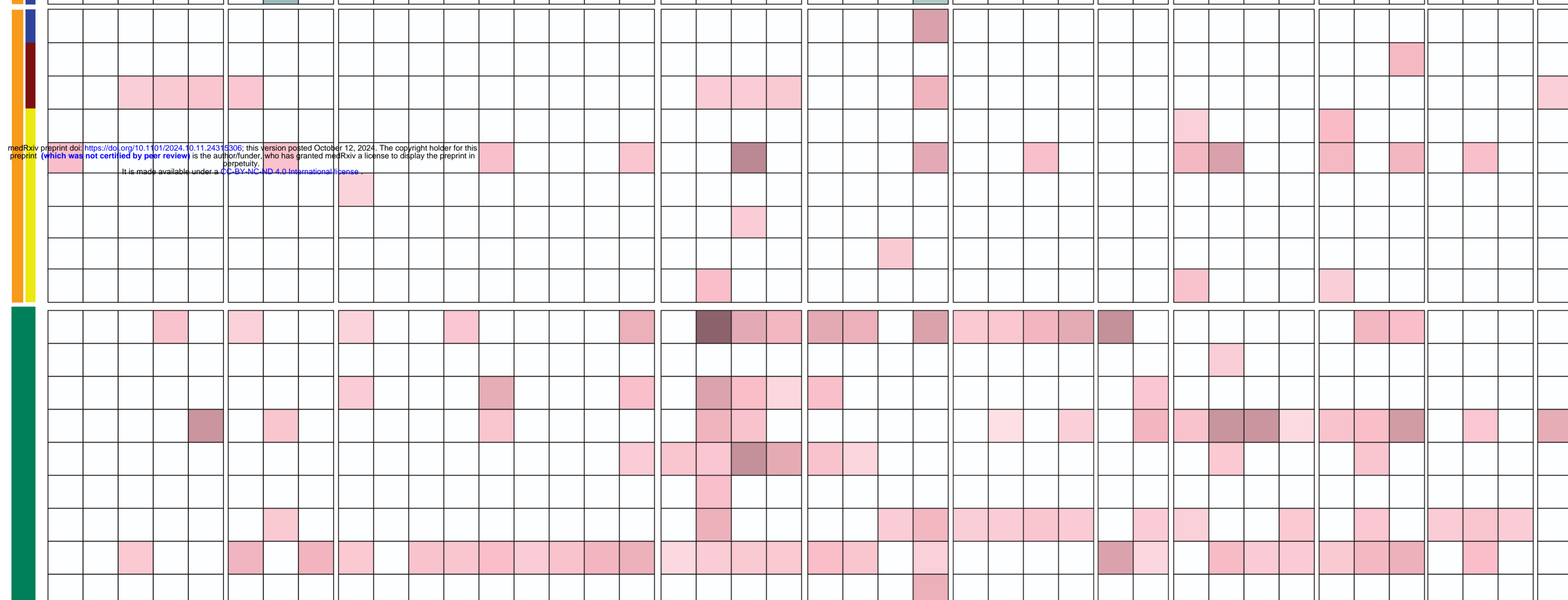
Reference-guided assembly



*f\_Parvoviridae*  
*s\_Gammapolyomavirus secanaria*  
*s\_Rotavirus phigastroenteritidis* (Rotavirus F)  
*s\_Rotavirus gammagastroenteritidis* (Rotavirus G)  
*s\_Rotavirus deltagastroenteritidis* (Rotavirus D)  
*s\_Iflavirus sacbroodi* (Sacbrood virus)  
*g\_Iflavirus* (Moku virus)  
*s\_Cripavirus mortiferum* (Aphid lethal paralysis virus)  
*s\_Labdapapillomavirus 1*

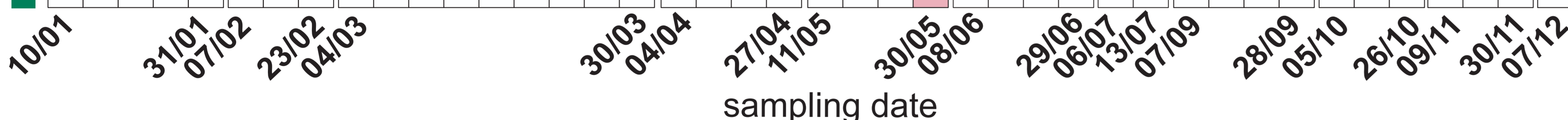
**B.**

De-novo assembly

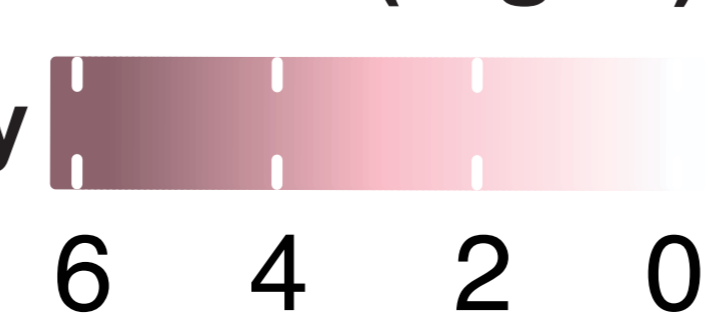


*s\_Labdapapillomavirus 1*  
*s\_Gammapolyomavirus secanaria*  
*f\_Parvoviridae*  
*f\_Picornavirales* (Varroa picorna-like virus)  
*sf\_Densovirinae*  
*g\_Ambidensovirus* (Periparus ater ambidensovirus)  
*g\_Chloriridovirus*  
*s\_Glossinavirus glopallidipedis*  
*f\_Dicistroviridae*  
*f\_Alphaflexiviridae*  
*f\_Chrysoviridae*  
*f\_Endornaviridae*  
*f\_Genomoviridae*  
*f\_Partitiviridae*  
*f\_Secoviridae*  
*f\_Tombusviridae*  
*f\_Totiviridae*  
*f\_Virgaviridae*

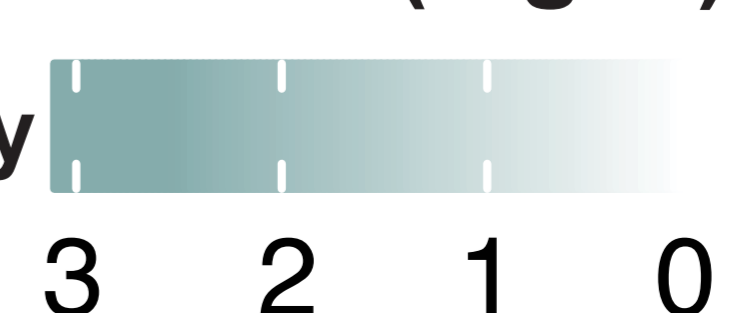
virus identified



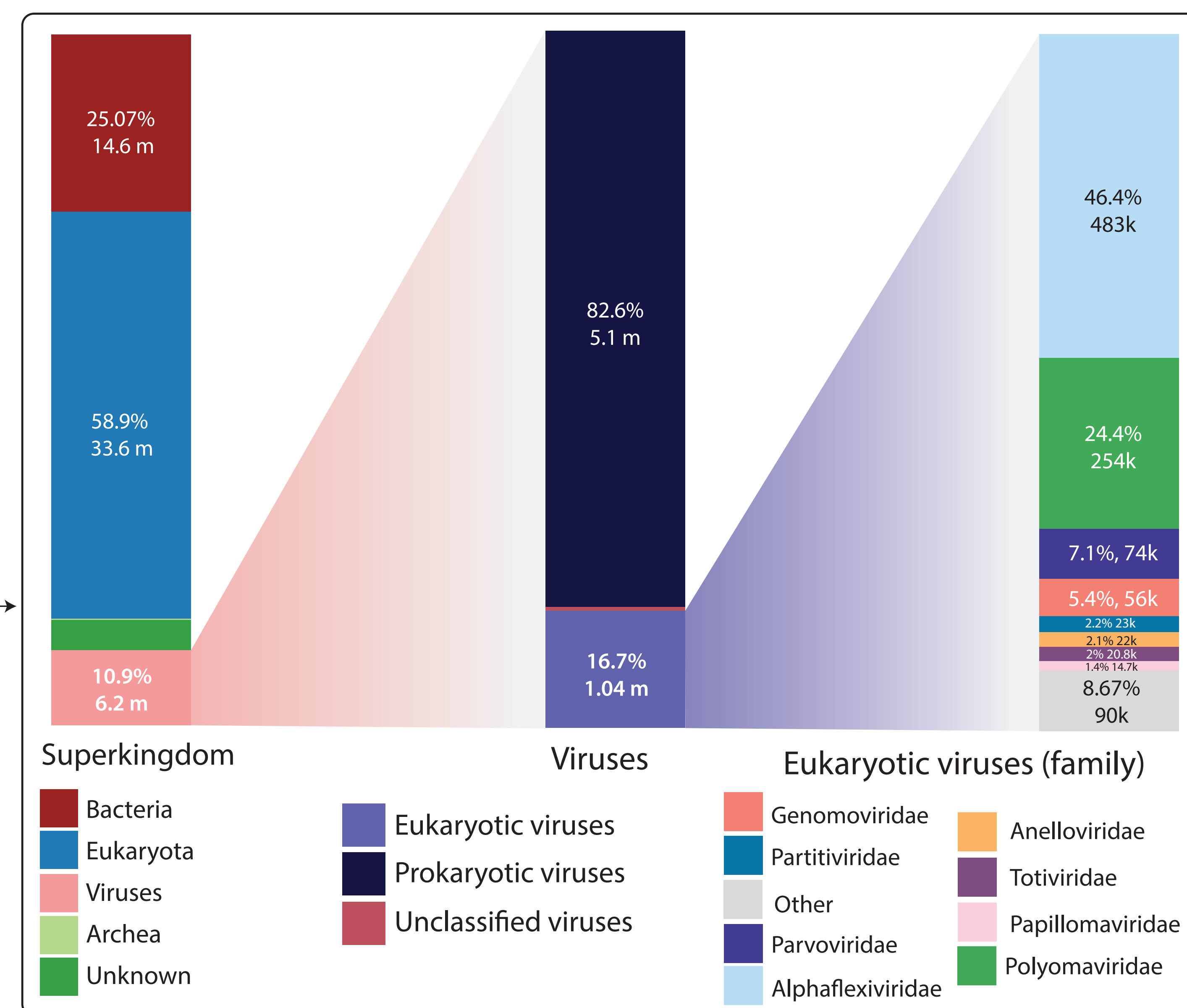
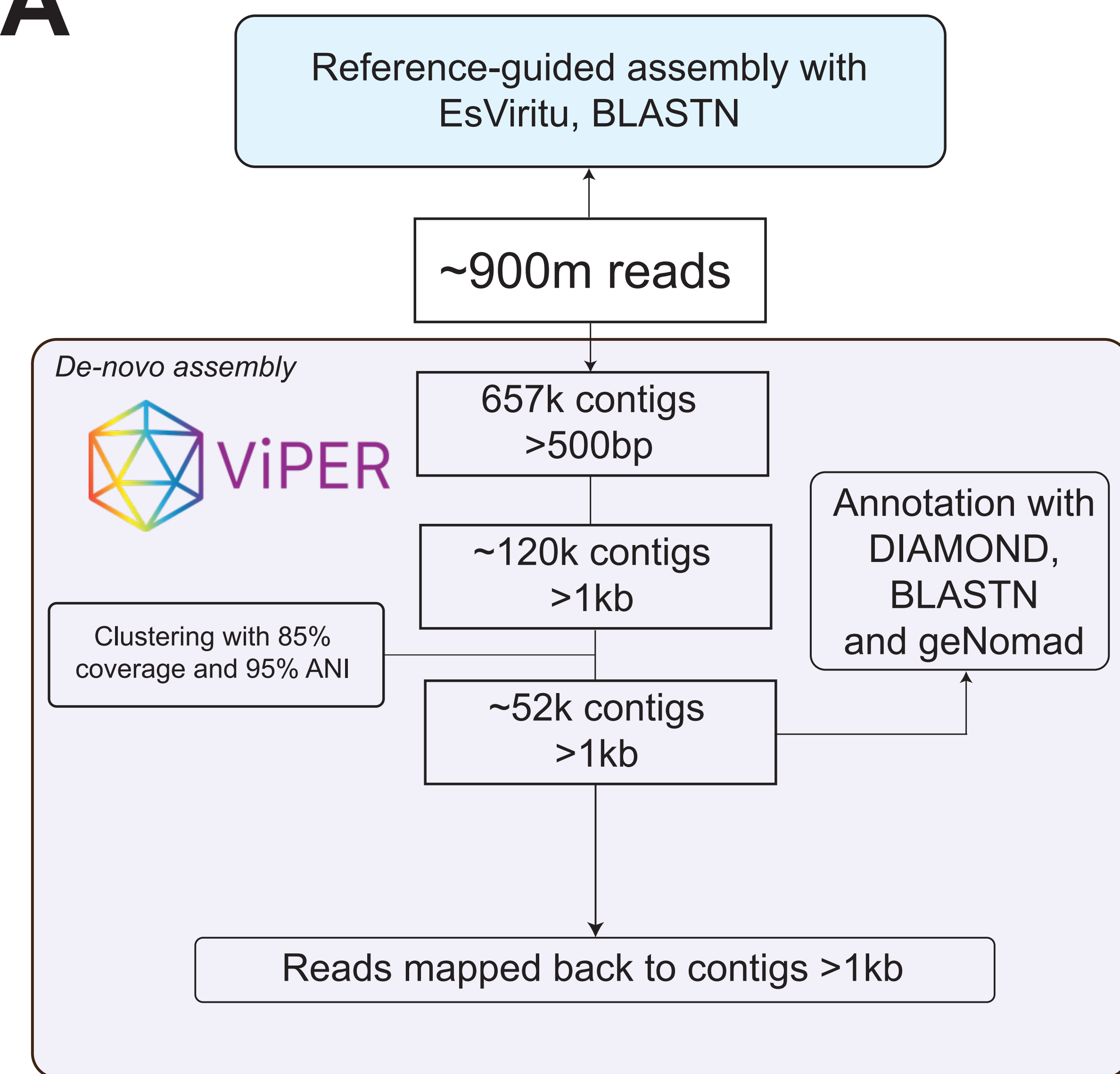
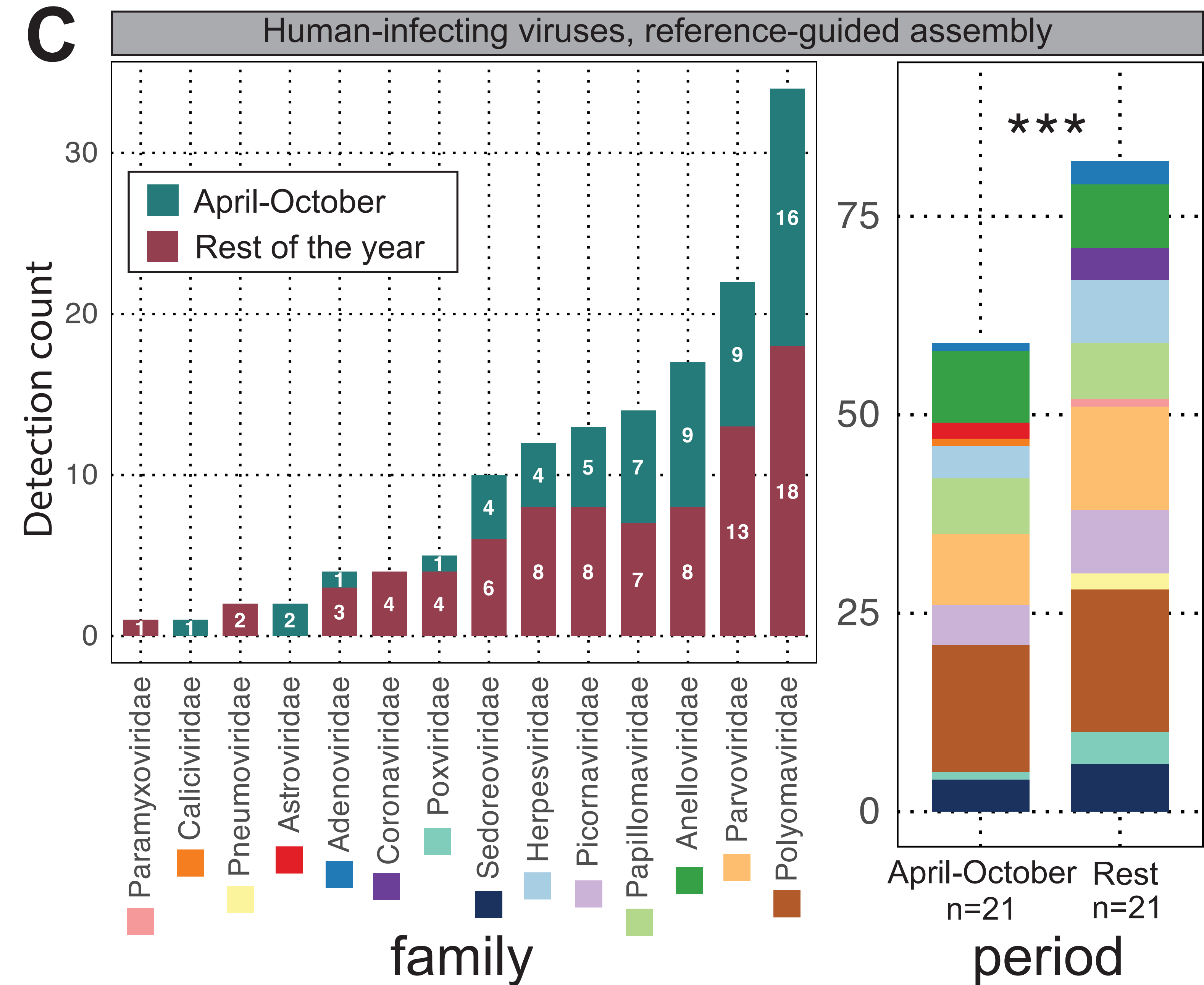
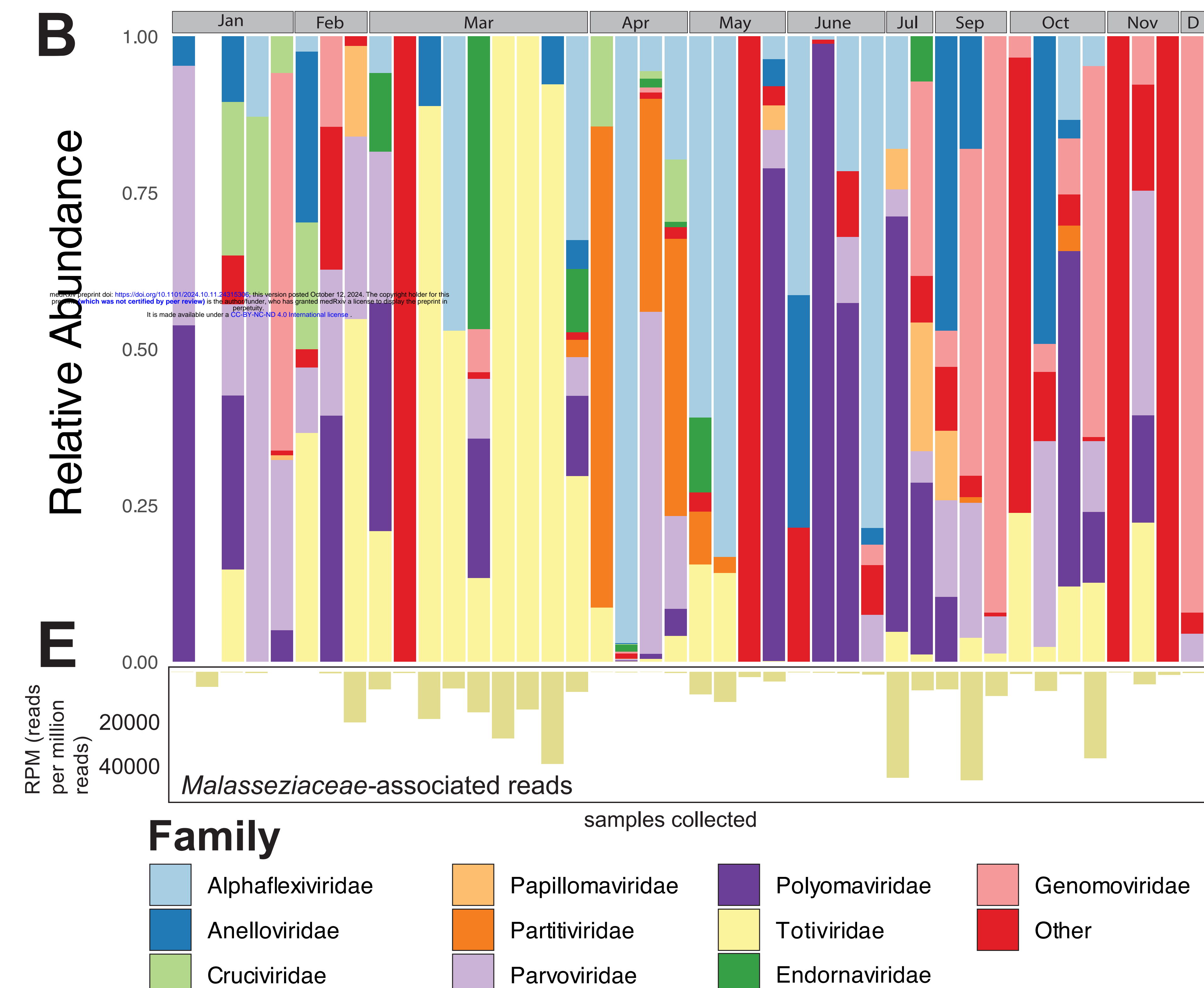
sampling date

**Host****Reads (log10)**

De-novo assembly

**Reads (log10)**

Reference-guided assembly

**A****C****B****D**