

1 The respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease

2

3 Anneloes L. van Rijn¹, Sander van Boheemen¹, Ellen C. Carbo¹, Nikos Pappas², Igor Sidorov¹, Hailiang

4 Mei², Marianne Aanerud^{3,4}, Per Bakke⁴, Eric C.J. Claas¹, Tomas M. Eagan^{3,4}, Pieter S. Hiemstra⁵, Aloys

5 C.M. Kroes¹, Jutte J.C. de Vries^{1*}

6

7 ¹Department of Medical Microbiology

8 Leiden University Medical Center

9 Leiden, the Netherlands

10

11 ²Sequencing Analysis Support Core, Department of Medical Data Sciences

12 Leiden University Medical Center

13 Leiden, the Netherlands

14

15 ³Department of Thoracic Medicine

16 Haukeland University Hospital

17 Bergen, Norway

18

19 ⁴Department of Clinical Science

20 University of Bergen

21 Bergen, Norway

22

23 ⁵Department of Pulmonology

24 Leiden University Medical Center

25 Leiden, the Netherlands

26

27 *corresponding author

28 Jutte J.C. de Vries, MD, PhD

29 Leiden University Medical Center

30 Department of Medical Microbiology

31 PO box 9600

32 2300 RC Leiden

33 The Netherlands

34 Telephone: +31 (0)71-5265242

35 Fax: +31 (0)71-5266981

36 jjcdevries@lumc.nl

37

38 **Take home message**

39 The respiratory virome in patients with chronic obstructive pulmonary disease was analysed using a

40 protocol that was shown to be highly sensitive and specific. Viral infections were associated with

41 reduced abundance of bacteriophages, implicating skewing of the virome.

42 **Abstract**

43 **Introduction:** Exacerbations are major contributors to morbidity and mortality in patients with
44 chronic obstructive pulmonary disease (COPD), and respiratory bacterial and viral infections are an
45 important trigger for the occurrence of such exacerbations. However, using conventional diagnostic
46 techniques, a causative agent is not always found. Metagenomic next-generation sequencing
47 (mNGS) allows analysis of the complete virome, but has not yet been applied in COPD exacerbations.

48 **Objectives:** To study the respiratory virome in nasopharyngeal samples during COPD exacerbations
49 using mNGS.

50 **Study design:** 88 nasopharyngeal swabs from 63 patients from the Bergen COPD Exacerbation Study
51 (2006-2010) were analysed by mNGS and in-house qPCR for respiratory viruses. Both DNA and RNA
52 were sequenced simultaneously using an Illumina library preparation protocol with in-house
53 adaptations.

54 **Results:** By mNGS, 23/88 samples tested positive. Sensitivity and specificity were both 96% for
55 diagnostic targets (23/24 and 1067/1120, respectively). Viral pathogens only detected by mNGS
56 were herpes simplex virus type 1 and coronavirus OC43. A positive correlation was found between
57 Cq value and mNGS viral species reads ($p=0.008$). Patients with viral pathogens had lower
58 percentages of bacteriophages ($p<0.000$). No correlation was found between viral reads (species and
59 genus level) and clinical markers.

60 **Conclusions:** The mNGS protocol used was highly sensitive and specific for semi-quantitative
61 detection of respiratory viruses. Excellent negative predictive value implicates the power of mNGS to
62 exclude any infectious cause in one test, with consequences for clinical decision making. Reduced
63 abundance of bacteriophages in COPD patients with viral pathogens implicates skewing of the
64 virome, and speculatively the bacterial population, during infection.

65

66 **Keywords:** metagenomic NGS; COPD; respiratory viruses; virome

67 **Introduction**

68 Chronic obstructive pulmonary disease (COPD) is characterized by exacerbations with high morbidity
69 and mortality, with over 65 million patients worldwide¹. A COPD exacerbation is an acute event
70 leading to worsening of the respiratory symptoms and is associated with a deterioration of lung
71 function². Exacerbations are mainly associated with infections, of which a large part is caused by
72 viruses (22-64%)³⁻⁶. However, in part of the exacerbations an etiologic agent is not detected.

73 Current routine virus diagnostics is based on polymerase chain reactions (PCR) and inherently the
74 number of detectable pathogens is restricted to the ones included in the assay. Rare, mutated and
75 pathogens with an uncommon clinical presentation will be missed, along with any new and currently
76 unknown ones. Over the last decades, several previously unidentified viruses have been discovered
77 as respiratory pathogen, including metapneumovirus⁷, middle-east respiratory syndrome
78 coronavirus⁸ and human bocavirus⁹.

79 Metagenomic next generation sequencing (mNGS) is an innovative method which enables the
80 detection of all genomes in a given sample. Proof of principle studies have shown that mNGS on
81 respiratory samples can confirm and extend PCR results and deliver typing and resistance data at the
82 same time¹⁰⁻¹⁴. The performance of mNGS in the clinical diagnostic setting, especially the positive
83 and negative predictive value, has not yet been elucidated and is likely to differ per clinical
84 syndrome and sample.

85 Previous data from reports on 16S rRNA analysis from the respiratory tract have led to increased
86 insight in the microbiome in patients with COPD¹⁵. Changes in bacterial populations have been
87 associated with exacerbation events and clinical phenotypes.¹⁵ However, these studies are
88 intrinsically limited to analysis of the bacterial part of the microbiome.

89 So far only a few studies using shotgun metagenomics focus on the respiratory virome in children
90 with acute respiratory infections^{16,17}. In this study, we analyse the composition of the virome in adult
91 patients with exacerbations of COPD.

92

93 **Objectives**

94 The aim of this study was to correlate the respiratory virome in COPD patients as found by mNGS
95 with qPCR and clinical data.

96 **Study design**

97 *Patients*

98 Patients with COPD were included in the Bergen COPD exacerbation study (BCES) between 2006 and
99 2010 in Bergen, Norway¹⁸. All patients lived in the Haukeland University Hospital district. Baseline
100 data included amongst others exacerbation history, comorbidities, spirometry and Global Initiative
101 for Chronic Obstructive Lung Disease (GOLD 2007) categorisation. Patients with an exacerbation
102 were scheduled for an appointment with a study physician the next working day. During
103 exacerbations, nasopharynx swabs were sampled and two different markers for the severity of the
104 exacerbation were scored. After an exacerbation a control visit was scheduled. During the study
105 period 154 patients had at least one exacerbation and in total 325 exacerbations were included in
106 the Bergen COPD study, of which 88 exacerbation samples were tested in the current study.

107

108 *Sample selection*

109 Nasopharyngeal samples were frozen and stored at -80°C. In total 88 nasopharyngeal samples of
110 patients at the time of exacerbation were selected based on the availability of other samples
111 (outside the current focus) and sent to the Leiden University Medical Center for further testing.

112

113 *Lab developed real-time PCR testing (qPCR)*

114 The viral respiratory panel covered by the multiplex real-time PCR (qPCR) developed in our
115 laboratory consists of coronavirus 229E, coronavirus HKU1, coronavirus NL63, coronavirus OC43,
116 influenza A, influenza B, human metapneumovirus, parainfluenza 1-4 (differentiation with probes),
117 respiratory syncytial virus, and rhinovirus¹⁹.

118 Total nucleic acids (NA) were extracted directly from 200 µl clinical sample, using the Total Nucleic
119 Acid extraction kit on the MagnaPure LC system (Roche Diagnostics, Almere, the Netherlands) with
120 100 µL output eluate. Nucleic acid amplification and detection by real-time PCR was performed on a
121 BioRad CFX96 thermocycler, using primers, probes and conditions as described previously¹⁹. Cq
122 values were normalized using a fixed baseline fluorescence threshold.

123

124 *Metagenomic next generation sequencing (mNGS)*

125 The metagenomics protocol used has been described previously¹⁴. In short, internal controls, Equine
126 Arteritisvirus (EAV) for RNA and Phocid Herpesvirus-1 (PhHV) for DNA (kindly provided by prof. dr.
127 H.G.M. Niesters, the Netherlands), were spiked in 200 µl of the virus transport medium in which the
128 nasopharyngeal swab was stored. Nucleic acids were extracted directly from 200 µl clinical sample
129 using the Magnapure 96 DNA and Viral NA Small volume extraction kit on the MagnaPure 96 system
130 (Roche Diagnostics, Almere, The Netherlands) with 100 µL output eluate. For library preparation, 7
131 µl of nucleic acids were used, using the NEBNext[®] Ultra™ Directional RNA Library Prep Kit for
132 Illumina[®], with several in-house adaptations to the manufacturers protocol in order to enable
133 simultaneous detection of both DNA and RNA. The following steps were omitted: poly A mRNA
134 capture isolation, rRNA depletion and DNase treatment step. This resulted in a single tube per
135 sample throughout library preparation containing both DNA and RNA. Metagenomic sequencing was
136 performed on an Illumina NextSeq 500 sequencing system (Illumina, San Diego, CA, USA), and
137 approximately 10 million 150 bp paired-end reads per sample were obtained.

138 After quality pre-processing, sequencing reads were taxonomically classified with Centrifuge²⁰ using
139 an index constructed from NCBI's RefSeq and taxonomy databases (accessed November 2017) with
140 reference nucleotide sequences for the domains of viruses, bacteria, archaea, fungi, parasites,
141 protozoa. Reads with multiple best matches were uniquely assigned to the lowest common ancestor

142 (k=1 setting; previously validated¹⁴). Horizontal coverage (%) was determined using
143 www.genomedetective.com²¹ version 1.111 (accessed 2018, December 30th).

144

145 *Assembly of the betacoronavirus*

146 For samples with dubious or inconclusive classification results a *de novo* assembly was performed.
147 Pre-processed short reads assigned to a higher taxonomic level of a suspected viral target were
148 extracted and *de novo* assembled with SPAdes version 3.11.1²² into longer stretches of contiguous
149 sequences (contigs). The resulting contigs were then run against the blast NCBI's nucleotide (nt)
150 database (accessed 2017) using blastn 2.7.1²³. After identification of a putative target sequence, all
151 the reads from the original sample were mapped against the identified best BLAST hit for further
152 confirmation using BWA 0.7.17 software package²⁴.

153

154 *Statistical analysis*

155 Sensitivity, specificity, positive and negative predictive values were calculated based on 24 PCR
156 positive and 1120 PCR negative target results of 88 samples.

157 Correlation between qPCR Cq value and logarithm of numbers of mNGS viral reads was tested with
158 population Pearson correlation coefficient.

159 Potential correlations of mNGS data with clinical variables were tested as follows. Cq value/ viral
160 reads and clinical parameters (exacerbation severity, duration of exacerbation or decrease/increase
161 in Forced Expiratory Volume in 1 second (FEV₁, control visit compared to baseline) were tested with
162 one-way ANOVA and Kruskal-Wallis test when appropriate (depending on distribution). Comparison
163 of the percentage of phages of all viral reads (after subtraction of the internal control EAV reads)
164 between mNGS virus positive samples and negative samples was tested with Mann-Whitney U test,
165 comparison with clinical parameters with Kruskal-Wallis test. Diversity of the virome in different

166 patient groups was characterized by Shannon Diversity Index (H) and tested with Welch two sample
167 t-test. Statistical analyses were performed using IBM SPSS Statistics version 23 software for
168 Windows and R version 3.3.0. Differences at a p-value <0.05 were considered statistically significant.

169

170 *Ethical approval*

171 The BCES study was approved by the ethical committee REK-Vest in Norway (REK number 165.08).

172 The performance of additional testing, including mNGS, was approved by the medical ethics review

173 committee of the Leiden University Medical Center (CME number B16.004).

174 **Results**

175 *Patients and samples*

176 In total 63 patients with 88 exacerbations were included with a median of one exacerbation per
177 patient (range 1-5). Baseline patient characteristics and exacerbation characteristics are shown in
178 tables 1 and 2 respectively.

179

180 *Lab developed real-time PCR*

181 Of the 88 samples, 23 (26%) tested positive with in-house PCR: 14 (61%) were rhinovirus positive,
182 three influenza A, two coronavirus NL63, one coronavirus OC43, two parainfluenza 3 and one
183 parainfluenza 4. Cq values ranged from 19-38 (Table 3).

184

185 *Metagenomic next generation sequencing*

186 A median of 11 million (7,522,643-20,906,019) sequence reads per sample were obtained. Of the 11
187 million reads, approximately 94% were *Homo sapiens* reads, 2 % were bacterial and 0.1% viral (Table
188 4). No fungal reads were detected. A median of 2% of the reads could not be assigned to sequences
189 in the Centrifuge index database (NCBI RefSeq).

190

191 *Comparison of mNGS to qPCR*

192 Of the 23 qPCR positive samples, 22 tested positive with mNGS, resulting in a sensitivity of mNGS of
193 96%. Only one sample, that was rhinovirus positive by qPCR (Cq 38), could not be detected by mNGS
194 (Table 3). Coverage of reference genomes was high (93-100%) with the exception of two samples:
195 30% coverage of rhinovirus C (1,401,120 mapped reads, 88,353-fold depth), and 3% coverage of
196 influenza A virus (single genome segment, 8 mapped reads). Bowtie alignment confirmed the

197 rhinovirus C mapping, but not the influenza A mapping. Additional viral pathogens detected by
198 mNGS were herpes simplex virus type 1 (32,159 reads, 82% coverage, 36-fold depth) which is not in
199 qPCR viral respiratory panel, in the sample with the 8 influenza virus reads, and a betacoronavirus.
200 Of these 83,252 betacoronavirus reads, *de novo* assembly resulted in 3 contigs (size 30743, 274 and
201 232 bp respectively) with best BLAST hit coronavirus OC43 (reference genome accession AY391777).
202 A coverage plot of all reads against this reference strain (Figure 1) showed good horizontal and
203 vertical coverage (read coverage depth 428). The original OC43 qPCR amplification appeared to have
204 been inhibited, and repeated OC43 qPCR confirmed the positive mNGS result (Cq 25).

205

206 *Sensitivity, specificity and predictive value*

207 The sensitivity, specificity and predictive values of mNGS were calculated based on 24 PCR positive
208 and 1120 PCR negative target results of 88 samples (Table 5). Calculations were made using different
209 cut-off values of respectively ≥ 0 , >15 and >50 mapped sequence reads. With a cut-off of >15 reads,
210 the sensitivity was 92% and specificity 100%. With increasing cut-off levels, the positive predictive
211 value (PPV) increased to 87%. The negative predictive value (NPV) was 100% for all cut-off levels.

212

213 *Typing*

214 mNGS provides additional typing data, as compared to qPCR. Of the 13 rhinoviruses detected with
215 mNGS, 6 (46.2%) were rhinovirus A, 2 (15.4%) rhinovirus B and 5 (38.5%) rhinovirus C. The three
216 influenza viruses were assigned to be H3N2 strains by mNGS.

217

218 *Semi-quantification by means of mNGS read count*

219 In order to analyse the semi-quantitative quality of the mNGS assay, the number of the sequence
220 reads (log) mapping to qPCR target viruses (species level) as obtained with mNGS were compared to
221 the Cq values of qPCR. A significant negative correlation was found (Figure 2; Pearson correlation
222 coefficient $\rho=-0.5$, $p=0.008$).

223

224 *Clinical parameters and mNGS pathogen read count*

225 The following markers were tested for potential associations with clinical severity of exacerbation
226 (exacerbation severity, self-reported exacerbation severity), length of exacerbation and a
227 decrease/increase in FEV₁ (control visit compared to baseline): mNGS pathogen positive versus
228 negative exacerbation (qPCR targets), the number of species reads (log) for the different target
229 viruses (species and family level), the number of target virus genus reads (%) of all virus reads. No
230 correlation was found between these markers and the different disease severity parameters (results
231 not shown).

232

233 *The respiratory virome*

234 Overall viral families detected by mNGS and abundance of mNGS reads for these families are shown
235 in Figure 3a and b (bacteriophages). Patients with viral pathogens (PCR target viruses) had
236 significantly reduced amounts of bacteriophages when compared to patients without viral pathogen:
237 17% and 54% bacteriophages respectively ($P<0.000$, bacteriophage reads vs. all viral reads, excluding
238 EAV control reads). Furthermore, Shannon diversity scores were significantly higher for COPD
239 exacerbations of viral etiology ($p<0.000$, viral reads in PCR positive versus negative patients, Figure
240 4).

241 No significant association was found between the diversity scores, nor the percentage of
242 bacteriophages, and the following parameters: disease severity, length of exacerbation, number of

243 exacerbations during the study period, difference in FEV₁, GOLD stage, smoking, CRP level, and the

244 virus species (results not known).

245

246 **Discussion**

247 In this study, the respiratory virome in patients with COPD exacerbations was analysed with both
248 mNGS and qPCR, and combined with clinical data. The incidence of viral pathogens was 26% with
249 both mNGS and qPCR: mNGS failed to detect one Rhinovirus with low load (Cq 38) and PCR failed
250 once due to one of the limitations of PCR, *i.e.* inhibition of amplification. One additional viral
251 pathogen was detected: herpes simplex virus 1, found by others to be associated with COPD²⁵.

252 The incidence of viral pathogens was comparable to that in previous publications (22-64%^{3,5,6}). The
253 viral pathogen with the highest incidence was rhinovirus, followed by influenza, coronaviruses and
254 para-influenza viruses. Interestingly, subtyping data was readily available by mNGS, accentuating the
255 high resolution of mNGS, with rhinovirus RV species A and C being most frequent, followed by RV-B.
256 RV-C was first identified in 2006 and associated with high symptom burdens in children and
257 asthmatics^{26,27}. Recently, an asthma-related cadherin-related family member 3 (CDHR3) gene
258 variant²⁸ was associated with greater RV-C receptor display on pulmonary cell surfaces of children
259 and adults, and associated with higher susceptibility to severe virus-triggered asthma episodes^{29,30}.

260 In line, Romero-Espinoza et al detected predominantly RV-C in children with acute asthma
261 exacerbations by mNGS³¹. The significance of RV-C infection in the adult population is less well
262 studied. Although RV-C has been previously associated with exacerbations of COPD^{32,33}, to our
263 knowledge, to date, CDHR3 polymorphisms have not yet found to be associated with COPD.

264 Furthermore, the complete respiratory virome showed a high phage abundance that could be linked
265 to the absence of viral pathogens. Lower phage abundance may be the result of viral expansion.
266 Hypothetically, a healthy virome size and diversity fits a certain size and diversity of bacteriophages,
267 while during viral infection, pathogens predominate the virome. Alternatively, others have
268 hypothesized that viral and microbial diversity may play a role in infection susceptibility and the
269 development of acute and chronic respiratory diseases³¹. Others have found a higher phage
270 abundance in patients with severe COPD when compared with moderate COPD and healthy controls,

271 in line with the hypothesis of a state of dysbiosis that increases with disease progression²⁵. In COPD
272 patients, viral infections have been suggested to trigger bacterial overgrowth and infections^{34,35},
273 demonstrating the significance of viral-bacterial interactions. Moreover, hypothetically,
274 bacteriophages play a role in the horizontal gene transfer of bacterial virulence factors. Study of the
275 lower airways by means of e.g. protected brushes during bronchoscopy are needed for further
276 analysis of bacterial and viral (sub)populations.

277 The sensitivity, specificity and positive and negative predictive values of mNGS were high: 92%,
278 100%, 82% and 100%, respectively, when encountering a cut-off of >15 sequence reads, with a
279 detection limit of approximately Cq 38. The high negative predictive value implicates the power of
280 mNGS to exclude any viral infectious cause in one test. The potential to exclude any infectious cause,
281 both viral and bacterial, would have significant consequences for starting and/or continuation of
282 antimicrobial or, at the other end of the spectrum, immune-modulating treatment. The viral species
283 sequence read count might give an indication of the viral burden and the clinical relevancy of the
284 detected virus. Although in our dataset we could not find any correlation with disease severity,
285 several paediatric studies demonstrated a correlation between virus load and disease severity in
286 respiratory infections³⁶⁻³⁹. Further analysis with a larger number of infected patients and/or a
287 different spectrum of exacerbation severity will be needed to demonstrate or exclude such an
288 association in COPD patients.

289 Though mNGS renders the possibility to detect all viruses in direct respiratory material, this
290 revolutionary method is not yet used as routine accredited diagnostic procedure for pathogen
291 detection. Before mNGS can be implemented as a routine diagnostics, the optimal protocol must be
292 defined and analysis and interpretation of the metagenomic data must be standardized, followed by
293 external quality assessment. This study demonstrates good performance of our mNGS protocol, in
294 line with other studies^{40,41} and seems to overcome some of the current thresholds for
295 implementation in clinical diagnostics.

296

297 **Conclusions**

298 The mNGS protocol used was highly sensitive and specific for semi-quantitative detection of
299 respiratory viruses. Excellent negative predictive value implicates the power of mNGS to exclude any
300 infectious cause in one test, with consequences for clinical decision making. Reduced abundance of
301 bacteriophages in COPD patients with viral pathogens implicates skewing of the virome, and
302 speculatively the bacterial population, during infection.

303 **Acknowledgements**

304 We thank our GENERADE project partners Floyd Wittink, Wouter Suring (Hogeschool Leiden), Danny
305 Duijsings (BaseClear) and Christiaan Henkel (Leiden University). We also like to thank Tom Vreeswijk,
306 Lopje Höcker and Mario van Bussel (KML, LUMC) for help with the pre-sequencing experiments and
307 Jeroen Laros (Human Genetics, LUMC) for help with the bioinformatics.

308

309 **Disclosures**

310 Funding: Pilot experiments were partially funded by GENERADE Centre of Expertise Genomics in
311 Leiden.

312 Competing interests: none declared.

313 No part of the manuscript has been previously published.

314

315 **Author's contributions**

316 Initial patient inclusion and sample selection was performed by TME.

317 ALR, SB, ECJC, TME, PB, MA, PSH, ACMK and JJC contributed to the study design. MA examined all
318 patients. ALR performed qPCR testing, ALR, SB, ECC and IS analysed and interpreted the mNGS data.

319 Analyses and interpretation of combined data was performed by ALR, SB, ECC, ECJC, TME, PSH,

320 ACMK and JJC. ALR wrote the first version of the manuscript. SB, ECC, ECJC, TME, PSH, ACMK and

321 JJC contributed and revised the manuscript. All authors read and approved the final manuscript.

322

323

324

325

326 References

- 327 1. <http://www.who.int/respiratory/copd/burden/en/> accessed 2018, October 8th
- 328 2. Aaron SD. Management and prevention of exacerbations of COPD. *BMJ (Clinical research ed)* 2014;349:g5237.
- 329
- 330 3. Ko FW, Chan KP, Hui DS, et al. Acute exacerbation of COPD. *Respirology (Carlton, Vic)* 2016;21:1152-65.
- 331
- 332 4. Wedzicha JA, Seemungal TA. COPD exacerbations: defining their cause and prevention. *Lancet* 2007;370:786-96.
- 333
- 334 5. Sapey E, Stockley RA. COPD exacerbations . 2: aetiology. *Thorax* 2006;61:250-8.
- 335 6. Hewitt R, Farne H, Ritchie A, Luke E, Johnston SL, Mallia P. The role of viral infections in
- 336 exacerbations of chronic obstructive pulmonary disease and asthma. *Therapeutic advances*
- 337 *in respiratory disease* 2016;10:158-74.
- 338 7. van den Hoogen BG, de Jong JC, Groen J, et al. A newly discovered human pneumovirus
- 339 isolated from young children with respiratory tract disease. *Nat Med* 2001;7:719-24.
- 340 8. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel
- 341 coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012;367:1814-20.
- 342 9. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a
- 343 human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci U*
- 344 *S A* 2005;102:12891-6.
- 345 10. Prachayangprecha S, Schapendonk CM, Koopmans MP, et al. Exploring the potential of next-
- 346 generation sequencing in detection of respiratory viruses. *Journal of clinical microbiology*
- 347 2014;52:3722-30.
- 348 11. van Boheemen S, de Graaf M, Lauber C, et al. Genomic characterization of a newly
- 349 discovered coronavirus associated with acute respiratory distress syndrome in humans.
- 350 *mBio* 2012;3.
- 351 12. Yang J, Yang F, Ren L, et al. Unbiased parallel detection of viral pathogens in clinical samples
- 352 by use of a metagenomic approach. *Journal of clinical microbiology* 2011;49:3463-9.
- 353 13. Lewandowska DW, Schreiber PW, Schuurmans MM, et al. Metagenomic sequencing
- 354 complements routine diagnostics in identifying viral pathogens in lung transplant recipients
- 355 with unknown etiology of respiratory infection. *PloS one* 2017;12:e0177340.
- 356 14. van Boheemen S. Metagenomic sequencing for combined detection of RNA and DNA viruses
- 357 in respiratory samples from paediatric patients. *bioRxiv* 2018, Dec 10.
- 358 15. Wang Z, Bafadhel M, Haldar K, et al. Lung microbiome dynamics in COPD exacerbations. *Eur*
- 359 *Respir J* 2016;47:1082-92.
- 360 16. Wang Y, Zhu N, Li Y, et al. Metagenomic analysis of viral genetic diversity in respiratory
- 361 samples from children with severe acute respiratory infection in China. *Clin Microbiol Infect*
- 362 2016;22:458 e1-9.
- 363 17. Lysholm F, Wetterbom A, Lindau C, et al. Characterization of the viral microbiome in patients
- 364 with severe lower respiratory tract infections, using metagenomic sequencing. *PloS one*
- 365 2012;7:e30875.
- 366 18. Eagan TM, Ueland T, Wagner PD, et al. Systemic inflammatory markers in COPD: results from
- 367 the Bergen COPD Cohort Study. *Eur Respir J* 2010;35:540-8.
- 368 19. Loens K, van Loon AM, Coenjaerts F, et al. Performance of different mono- and multiplex
- 369 nucleic acid amplification tests on a multipathogen external quality assessment panel.
- 370 *Journal of clinical microbiology* 2012;50:977-87.
- 371 20. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of
- 372 metagenomic sequences. *Genome Res* 2016;26:1721-9.
- 373 21. Vिल्sker M, Moosa Y, Nooij S, et al. Genome Detective: An Automated System for Virus
- 374 Identification from High-throughput sequencing data. *Bioinformatics* 2018.

- 375 22. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its
376 applications to single-cell sequencing. *J Comput Biol* 2012;19:455-77.
- 377 23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol*
378 *Biol* 1990;215:403-10.
- 379 24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
380 *Bioinformatics* 2009;25:1754-60.
- 381 25. Elbeheery AHA, Feichtmayer J, Singh D, Griebler C, Deng L. The Human Virome Protein Cluster
382 Database (HVPC): A Human Viral Metagenomic Database for Diversity and Function
383 Annotation. *Front Microbiol* 2018;9:1110.
- 384 26. Arden KE, McErlean P, Nissen MD, Sloots TP, Mackay IM. Frequent detection of human
385 rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract
386 infections. *Journal of medical virology* 2006;78:1232-40.
- 387 27. Lamson D, Renwick N, Kapoor V, et al. MassTag polymerase-chain-reaction detection of
388 respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like
389 illness in New York State during 2004-2005. *The Journal of infectious diseases*
390 2006;194:1398-402.
- 391 28. Bonnelykke K, Sleiman P, Nielsen K, et al. A genome-wide association study identifies CDHR3
392 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet*
393 2014;46:51-5.
- 394 29. Bochkov YA, Watters K, Ashraf S, et al. Cadherin-related family member 3, a childhood
395 asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc Natl*
396 *Acad Sci U S A* 2015;112:5485-90.
- 397 30. Kanazawa J, Masuko H, Yatagai Y, et al. Genetic association of the functional CDHR3
398 genotype with early-onset adult asthma in Japanese populations. *Allergol Int* 2017;66:563-7.
- 399 31. Romero-Espinoza JA, Moreno-Valencia Y, Coronel-Tellez RH, et al. Virome and bacteriome
400 characterization of children with pneumonia and asthma in Mexico City during winter
401 seasons 2014 and 2015. *PloS one* 2018;13:e0192878.
- 402 32. Lau SK, Yip CC, Lin AW, et al. Clinical and molecular epidemiology of human rhinovirus C in
403 children and adults in Hong Kong reveals a possible distinct human rhinovirus C subgroup.
404 *The Journal of infectious diseases* 2009;200:1096-103.
- 405 33. McErlean P, Shackelton LA, Lambert SB, Nissen MD, Sloots TP, Mackay IM. Characterisation
406 of a newly identified human rhinovirus, HRV-QPM, discovered in infants with bronchiolitis.
407 *Journal of clinical virology : the official publication of the Pan American Society for Clinical*
408 *Virology* 2007;39:67-75.
- 409 34. D'Anna SE, Balbi B, Cappello F, Carone M, Di Stefano A. Bacterial-viral load and the immune
410 response in stable and exacerbated COPD: significance and therapeutic prospects. *Int J*
411 *Chron Obstruct Pulmon Dis* 2016;11:445-53.
- 412 35. Molyneaux PL, Mallia P, Cox MJ, et al. Outgrowth of the bacterial airway microbiome after
413 rhinovirus exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*
414 2013;188:1224-31.
- 415 36. El Saleeby CM, Bush AJ, Harrison LM, Aitken JA, Devincenzo JP. Respiratory syncytial virus
416 load, viral dynamics, and disease severity in previously healthy naturally infected children.
417 *The Journal of infectious diseases* 2011;204:996-1002.
- 418 37. Houben ML, Coenjaerts FE, Rossen JW, et al. Disease severity and viral load are correlated in
419 infants with primary respiratory syncytial virus infection in the community. *Journal of*
420 *medical virology* 2010;82:1266-71.
- 421 38. Zhao B, Yu X, Wang C, et al. High human bocavirus viral load is associated with disease
422 severity in children under five years of age. *PloS one* 2013;8:e62318.
- 423 39. Christensen A, Nordbo SA, Krokstad S, Rognlien AG, Dollner H. Human bocavirus in children:
424 mono-detection, high viral load and viraemia are associated with respiratory tract infection.

- 425 Journal of clinical virology : the official publication of the Pan American Society for Clinical
426 Virology 2010;49:158-62.
- 427 40. Thorburn F, Bennett S, Modha S, Murdoch D, Gunson R, Murcia PR. The use of next
428 generation sequencing in the diagnosis and typing of respiratory infections. Journal of
429 clinical virology : the official publication of the Pan American Society for Clinical Virology
430 2015;69:96-100.
- 431 41. Graf EH, Simmon KE, Tardif KD, et al. Unbiased Detection of Respiratory Viruses by Use of
432 RNA Sequencing-Based Metagenomics: a Systematic Comparison to a Commercial PCR
433 Panel. Journal of clinical microbiology 2016;54:1000-7.
- 434

Table 1. Baseline patient characteristics

	Patients (n=63)
Age median yrs (range)	63.5 (46.6-74.5)
Male sex	40 (64%)
BMI median (kg/m ²)	25 (15-39)
Body composition	
Cachectic	7 (11%)
Normal	24 (38%)
Overweight	22 (35%)
Obese	10 (16)
Smoking	
Never	0 (0%)
Sometimes	37 (59%)
Daily	26 (41%)
GOLD stage	
II (FEV ₁ 50-80%)	29 (46%)
III (FEV ₁ 30-50%)	27 (43%)
IV (FEV ₁ <30%)	7 (11%)
FEV ₁ in % median (range)	0.49 (0.23-0.74)
>1 exacerbation past 12 months	16 (25%)
Inhalation steroids	50 (79%)

Table 2. COPD patient and exacerbation characteristics among patients having a viral or non-viral exacerbation.

	Virus* detected n=23	Virus* not detected n=65	p**
Patient characteristics			
<i>Sex, %</i>			0.21
Women	34.5	65.5	
Men	22.0	78.0	
<i>smoking status, %</i>			0.53
Ex-smoker	23.4	76.6	
Current-smoker	29.3	70.7	
<i>GOLD stage (2007), %</i>			0.35
II (FEV ₁ 50-80%)	26.3	73.7	
III (FEV ₁ 30-50%)	30.8	69.2	
IV (FEV ₁ < 30%)	9.1	90.9	
<i>Frequent exacerbator, %</i>			0.72
No	25.0	75.0	
Yes	28.6	71.4	
<i>Using inhalation steroids, %</i>			0.55
No	20.0	80.0	
Yes	27.4	72.6	
<i>Age, mean yrs</i>	63.7	64.9	0.10
<i>BMI, mean kg/m²</i>	27.0	25.9	0.92
<i>FEV₁ in % predicted</i>	49.3	47.5	0.48
Exacerbation characteristics			
<i>Exacerbation severity for entire exacerbation</i>			0.75
Mild (not requiring AB or oral steroids or hospitalization)	14.3	85.7	
Moderate (requiring AB or oral steroids)	26.9	73.1	
Severe (Emergency room or hospital admission)	28.6	71.4	
<i>Self-reported exacerbation severity at time of study sampling</i>			0.64
Dyspnea unchanged or increased on errands outside home	36.4	63.6	
Increased dyspnea doing housework	26.5	73.5	
Increased dyspnea at rest	28.6	71.4	
Must sit up at night due to dyspnea	14.3	85.7	
<i>CRP (ng/mL) at time of study sampling†</i>	32.5	34.2	0.27

*qPCR target virus

** Pearson chi-square for categorical variables and t-test for continuous variables

† missing data for 4 (1 virus positive, 3 virus negative) exacerbations

Table 3. qPCR positive samples with respective mNGS results

Samples	qPCR positive (%)	Cq values range	mNGS species positive (%)	mNGS species reads (range)	Coverage (% range)
All targets	23/88 (26)	19-38	23/88 (26)	0-1,319,849	3-100
Influenza A	3/23 (13)	29-36	3/23 (13)	9-557	3-98
Cov NL63	2/23 (9)	32	2/23 (9)	1,518-139,814	93-100
Cov OC43	1/23 (4)	27	2*/23 (4)	1,319,849	99-99
PIV3	2/23 (9)	26-36	2/23 (9)	58-275,644	-**
PIV4	1/23 (4)	24	1/23 (4)	185,121	100-100
Rhinovirus	14/23 (61)	19-38	13***/23 (57)	0-40,409	
			RV-A: 6/13	632-420,551	94-100
			RV-B: 2/13	38,421-409,480	100-100
			RV-C: 5/13	9,398-4,992,575	30-100

*Retesting by qPCR confirmed the OC43 finding of mNGS

** No coverage using GenomeDetective, Bowtie alignment confirmed Centrifuge mapping

*** Rhinovirus not detected with mNGS had PCR Cq value 38

Table 4. mNGS read counts

	Median	Min	Max
Root reads	10,764,981	7,522,643	20,906,019
% unassigned reads	2	0.6	22
Homo sapiens reads(% root)	9,495,259 (94)	2,510,133	18,672,027
Bacterial reads (%root)	233,472 (2)	5,086	10,532,753
Viral reads (% root)	12,139 (0.1)	500	1,544,795

Table 5. Sensitivity and specificity of mNGS for PCR target viruses. PPV: positive predictive value,

NPV, Negative predictive value.

	Cut-off, reads		
	0	15	50
Sensitivity	96% (23/24*)	92% (22/24)	83% (20/24)
Specificity	96% (1076/1120)	100% (1115/1120)	100% (1117/1120)
PPV	34%	82%	87%
NPV	100%	100%	100%

*The sample with positive confirmatory OC43 PCR included.

FIGURES

Figure 1. Coverage plot of betacoronavirus reads to coronavirus OC43 reference genome AY391777.1 (depth of coverage: 428).

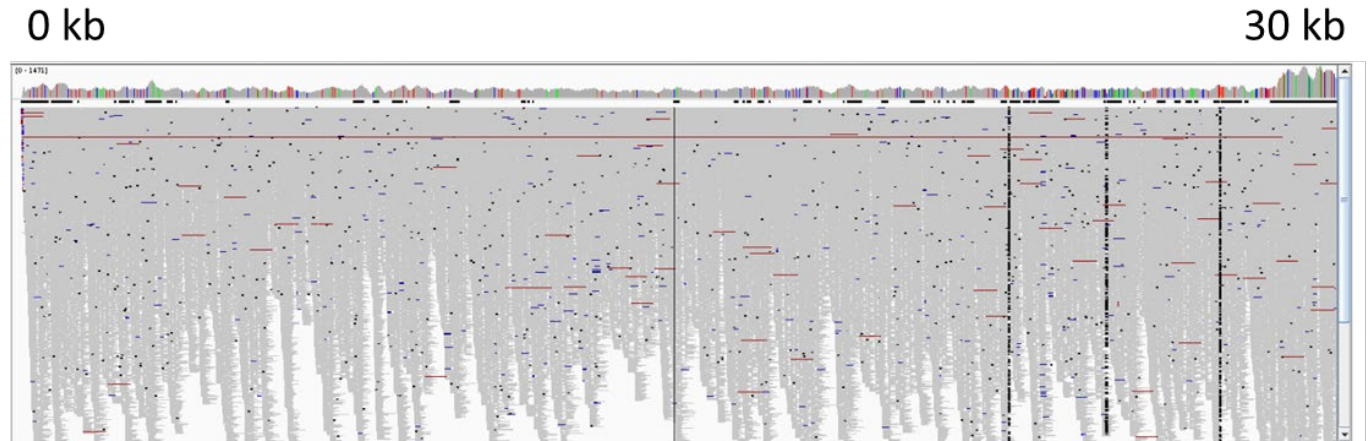


Figure 2: Correlation between mNGS viral species reads (log) and Cq value ($\rho=-0.5$, $p=0.008$)

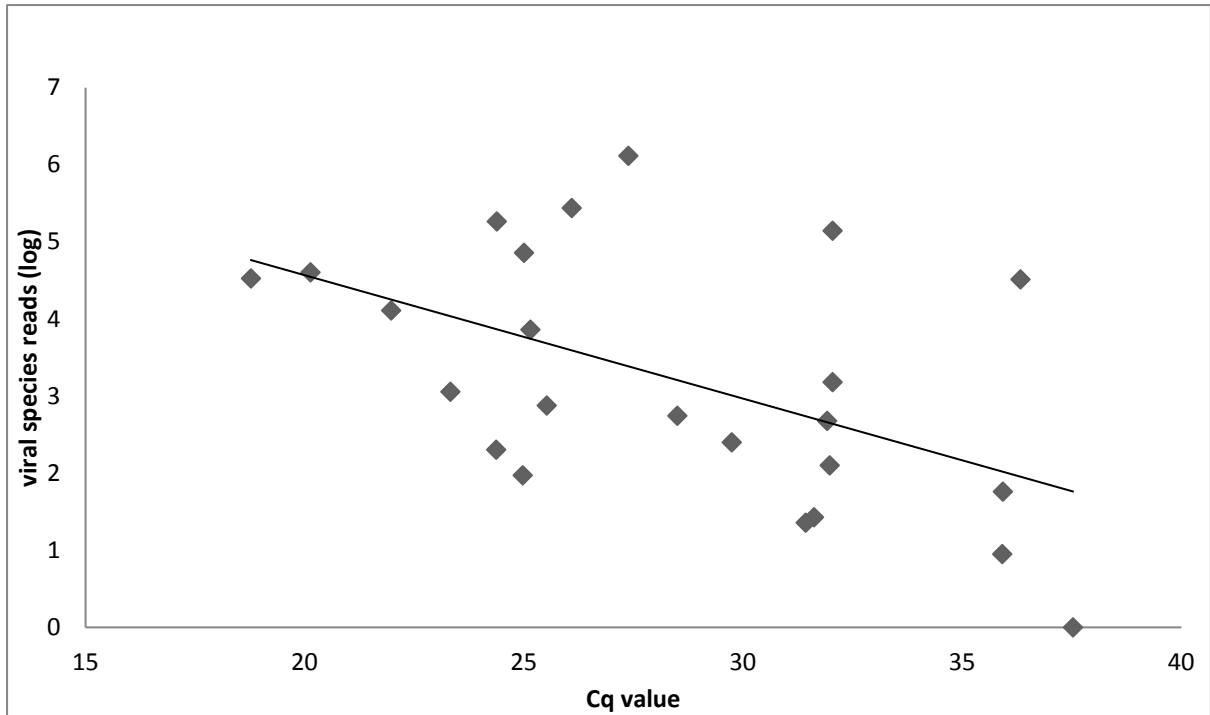


Figure 3. The respiratory virome: abundance of (a) viral families (bacteriophages excluded) and (b)

bacteriophages. Mean sequence read counts per family. Arteriviridae and Herpesviridae include internal control reads (EAV and PhHV-1 respectively). Patients with viral PCR pathogens had lower amounts of bacteriophages ($p < 0.000$).

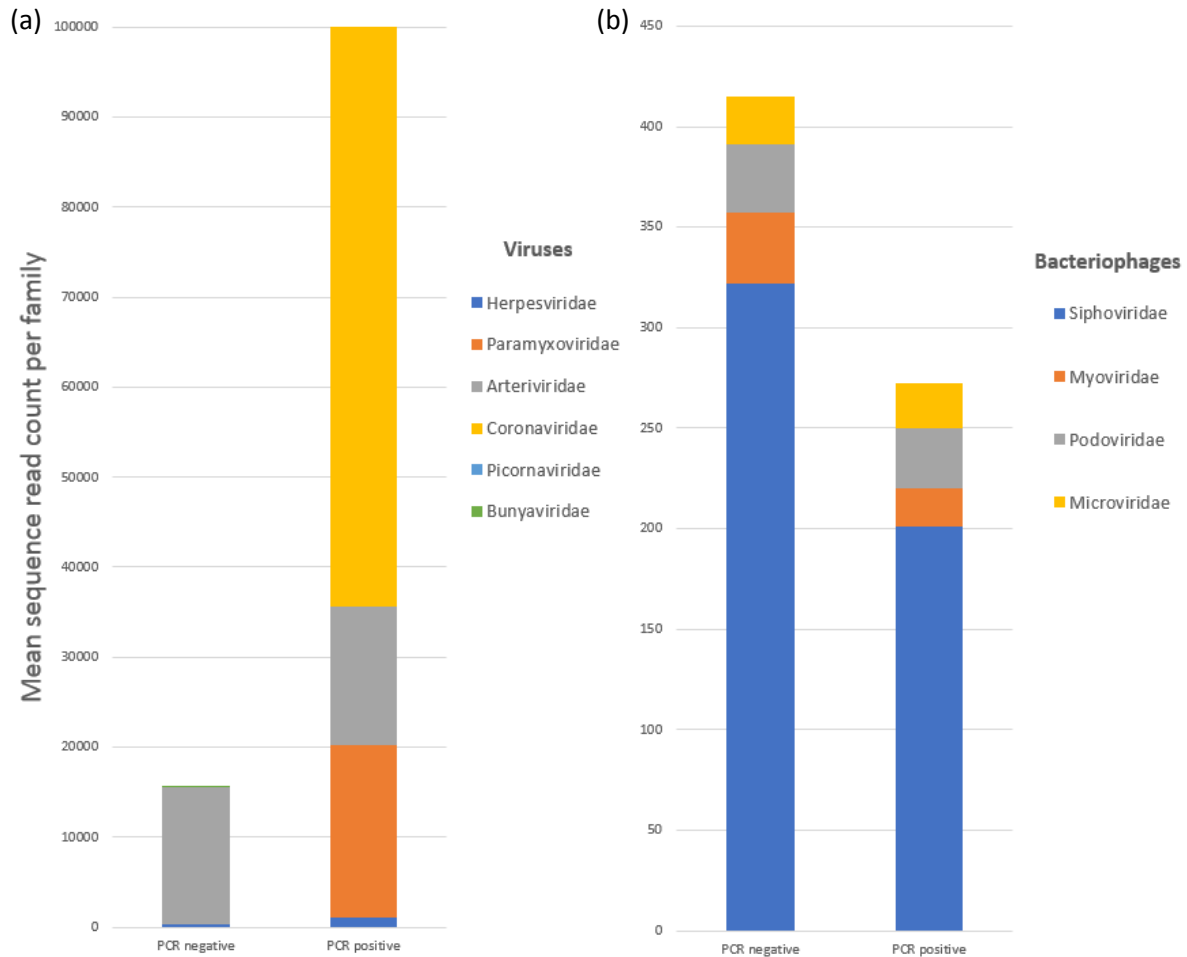


Figure 4. Shannon diversity scores for: (a) viruses, (b) viral PCR targets, (c) bacteriophages. COPD exacerbations of viral etiology had significant lower diversity (b). Boxes span IQR.

