

1 **Performance of five metagenomic classifiers for virus pathogen detection using respiratory**
2 **samples from a clinical cohort**

3

4 Ellen C. Carbo^{a,*}, Igor A. Sidorov^a, Anneloes L. van Rijn-Klink^a, Nikos Pappas^{b,1}, Sander van
5 Boheemen^{a,2}, Hailiang Mei^b, Pieter S. Hiemstra^c, Tomas M. Eagan^d, Eric C.J. Claas^a, Aloys C.M. Kroes^a,
6 Jutte J.C. de Vries^a

7

8 ^a Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

9 ^b Sequencing Analysis Support Core, Department of Biomedical Data Sciences, Leiden University

10 Medical Center, Leiden, The Netherlands

11 ^c Department of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands

12 ^d Department of Thoracic Medicine, Haukeland University Hospital, Bergen, Norway

13

14 ¹ Present affiliation: Theoretical Biology and Bioinformatics, Department of Biology, Science for Life,
15 Utrecht University, Utrecht, The Netherlands

16 ² Present affiliation: Department of Viroscience, Erasmus Medical Center, Rotterdam, The
17 Netherlands

18 *Corresponding author. E-mail address: E.C.Carbo@lumc.nl

19

20 Author contributions:

21 Patient inclusion and sample collection: TME, PH, sample pre-treatment and sequencing: ALRK, SB,
22 data analysis: ECC, IS, HM, data visualisation: IS, drafting the manuscript: ECC and JJC, review and
23 editing: all authors, study design: ECC, IS, JJC. All authors have read and agreed to the final version
24 of the manuscript.

25

26 Word count: main text: 3109, abstract: 229

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

27 **Highlights**

- 28 • The performance of five metagenomic classifiers was assessed using datasets obtained from
29 respiratory samples from a clinical cohort of patients
- 30 • 88 samples were characterized by means of 1,144 respiratory virus PCR results
- 31 • Using PCR as gold standard, sensitivity and specificity ranged from 83-100% and 90-99%
32 respectively, with the overall highest scores resulting from amino-acid based classification
33 by Kaiju classifier. Performance was dependent on classification level and exclusion of
34 human reads prior to classification.
- 35 • Normalization of assigned read counts for corresponding genome lengths generally had
36 minor effect on performance, but negatively affected the detection of target viruses with
37 read counts around detection level.
- 38 • Correlation between sequence read counts and PCR Ct-values varied per classifier (12.1-
39 62.7% at species level), per data pre-processing, and per virus. Outliers were detected of up
40 to 3 \log_{10} reads the predicted read counts for viruses with high sequence diversity.
- 41 • Sensitivity and specificity of the classifiers were within the range of use for diagnostic
42 practice when combined with a determined cut-off for defining a positive result.

43 **Abstract**

44 Viral metagenomics is increasingly being applied in clinical diagnostic settings for detection of
45 pathogenic viruses. While a number of benchmarking studies have been published on the use of
46 metagenomic classifiers for abundance and diversity profiling of bacterial populations, studies on
47 the comparative performance of the classifiers for virus pathogen detection are scarce.
48 In this study, metagenomic data sets (N=88) from a clinical cohort of patients with respiratory
49 complaints were used for comparison of the performance of five taxonomic classifiers: Centrifuge,
50 Clark, Kaiju, Kraken2, and Genome Detective. A total of 1,144 positive and negative PCR results for a
51 total of 13 respiratory viruses were used as gold standard. Sensitivity and specificity of these
52 classifiers ranged from 83-100% and 90-99% respectively, and was dependent on the classification
53 level and data pre-processing. Exclusion of human reads generally resulted in increased specificity.
54 Normalization of read counts for genome length resulted in minor overall performance, however
55 negatively affected the detection of targets with read counts around detection level. Correlation of
56 sequence read counts with PCR Ct-values varied per classifier, data pre-processing (R^2 range 15.1-
57 63.4%), and per virus, with outliers up to 3 \log_{10} reads magnitude beyond the predicted read count
58 for viruses with high sequence diversity.
59 In this benchmarking study, sensitivity and specificity were within the ranges of use for diagnostic
60 practice when the cut-off for defining a positive result was considered per classifier.

61 **Introduction**

62 In the era of next-generation sequencing (NGS), clinical metagenomics, analysis of all microbial
63 genetic material in clinical samples, is being introduced in diagnostic laboratories, revolutionizing
64 diagnostics of infectious diseases [1]–[4]. As opposed to running a series of pathogen targeted
65 diagnostic PCR assays to identify suspected pathogens, one single metagenomic run enables the
66 detection of all potential pathogens in a clinical sample [5][6]. The use of this method, also known as
67 shotgun high-throughput sequencing, has resulted in the detection of several pathogens missed by
68 current routine diagnostic procedures [1][7]. The clinical application of metagenomic sequencing for
69 pathogen detection has for a large part focused on patients with encephalitis [1], [8]–[12]. However,
70 patients with clinical syndromes suspected for an infectious disease but negative conventional test
71 results are increasingly considered as candidates for metagenomic testing. With sequencing costs
72 decreasing and the significance of detection of unexpected, novel viruses being underscored by the
73 currently pandemic SARS-CoV-2 [13], metagenomics is increasingly moving towards implementation
74 in diagnostic laboratories.

75 Performance testing is typically part of the implementation procedure in diagnostic laboratories to
76 ensure the quality of diagnostic test results. Accurate bioinformatic identification of viral pathogens
77 depends on both the classification algorithm and the database [14][15][16]. Metagenomic
78 sequencing in the past has been mainly oriented at profiling of bacterial genomes in the context of
79 microbiome comparisons in research settings, and most bioinformatic tools currently available have
80 been designed for that specific purpose [17][18]. Some of the previously bacterial oriented classifiers
81 are now being used for other domains, including viruses. However, viral metagenomics for pathogen
82 detection has specific challenges such as the low abundancy of viral sequences for some targets, and
83 incomplete or inaccurate reference sequences. The high diversity of viral sequences due to the high
84 mutation rate of RNA viruses further complicates accurate detection and identification [19]. While
85 an increasing number of benchmarking studies is being published on the use of metagenomic
86 classifiers for bacterial abundancy profiling, studies on the performance of classifiers for virus

87 pathogen detection are scarce. Publications on the performance of the computational analysis of
88 viral metagenomics are usually limited to *in silico* analysis of artificial sequence data [14][20][21] or
89 mock samples [22][23]. Though both sensitivity and specificity can be deduced when using
90 simulated datasets, they usually do not represent the complexity of data sets from clinical samples
91 which typically contain sequences from wet lab reagents that have been referred to as the ‘kitome’
92 [22][24][25]. These factors can affect the sensitivity and specificity of the overall procedure, and may
93 result in incorrect diagnoses. In contrast, performance studies that use real-world samples are
94 usually hindered by the huge number of negative metagenomic findings in the absence of gold
95 standard results for validation. Therefore, the performance parameters typically reported are recall
96 (sensitivity), precision (positive predictive value), and F1 (the harmonic mean of recall and
97 precision); while specificity is usually not assessed because negative findings by metagenomics are
98 poorly defined.

99 Here, we perform a comparison of five taxonomic classifiers: Centrifuge [26], Clark [18], Kaiju [27],
100 Kraken 2 [28] and Genome Detective [29]. The classifiers were tested using metagenomic shotgun
101 sequencing data obtained from a cohort of chronic obstructive pulmonary disease patients (COPD)
102 with a clinical exacerbation and therefore suspected of a respiratory infection. For these samples,
103 1,144 PCR test results were used as gold standard to infer both sensitivity and specificity of the
104 classifiers. For each classifier, we present appropriate benchmark scores for virus classification in the
105 diagnostic setting.

106

107 **Materials and Methods**

108 Clinical samples and PCR results

109 Clinical respiratory samples were used to obtain metagenomic data sets. In total 88 nasal washings
110 were taken from 63 patients with COPD suspected for respiratory infection as described previously
111 [30]. Each sample was tested using a respiratory PCR panel resulting in 1,144 real-time positive and
112 negative PCR results for 13 viral respiratory targets as described previously [30]. The respiratory
113 viruses addressed by this respiratory panel and cohort prevalence are shown in **Table 1**.

114

115 Ethical approval

116 Ethical approval for metagenomic sequencing of the cohorts was obtained from the medical ethics
117 review committee of the Leiden University Medical Center, the Netherlands, (CME number
118 B16.004).

119

120 Metagenomic next-generation sequencing (mNGS)

121 The metagenomic datasets used for comparison were generated as described before [30]. In short,
122 clinical samples were spiked with equine arteritis virus (EAV) and phocine herpesvirus 1 (PhHV-1), as
123 positive controls for RNA and DNA detection. Subsequently, extraction of nucleic acids was
124 performed using the Magnapure 96 DNA and Viral NA Small volume extraction kit on the MagnaPure
125 96 system (Roche, Basel, Switzerland). Library preparation was performed utilizing the NEBNext
126 Ultra II Directional RNA Library prep kit for Illumina (New England Biolabs, Ipswich, MA, USA) using
127 single, unique adaptors and a protocol optimized for processing RNA and DNA simultaneously in a
128 single tube [25]. Sequencing was performed on an Illumina NextSeq 500 sequencing system
129 (Illumina, San Diego, CA, USA) at GenomeScan BV (Leiden, Netherlands), obtaining approximately 10
130 million 150 bp paired-end reads per sample.

131

132 Pre-processing of data

133 Illumina 150 bp paired-end sequence reads were demultiplexed by standard Illumina software
134 followed by trimming, adapter clipping, and filtering of low-complexity reads using Trimmomatic [v.
135 0.36] [31]. Human reads were excluded after mapping them to the human genome GRCh38 [32]
136 using Bowtie2 [33].

137

138 Metagenomic classifiers

139 Bioinformatic metagenomics tools designed for taxonomic classification were selected for
140 benchmarking based on the following criteria: applicable for viral metagenomics for pathogen
141 detection; available either as download or webserver; and it is either widely used or showed
142 potential of being adopted for diagnostics in the future. Some tools considered were excluded due
143 to lack of support or details on how to use the tool, or non-functioning webserver. An overview of
144 characteristics of the selected classifiers can be found in **Table 2**.

145

146 Reference database

147 For comparison of classification performance, a single database was used as starting point for the
148 classifiers Centrifuge, Clark, Kaiju, and Kraken 2: viral genomes from NCBI/RefSeq [34] (downloaded
149 on 2020-12-27). Genome Detective was used as a service, it uses its own database generated on 3
150 March 2020 (version 1.130) by Genome Detective.

151

152 Metagenomic classifiers and characteristics

153 Centrifuge

154 Classification with Centrifuge (version 1.0.4) [26] is based on exact matches of at least 22 base pair
155 nucleotide sequences with the reference index, using k -mers of user-defined length. Centrifuge by
156 default allows five classification labels per sequence read. For a realistic comparison, in the current
157 study, this setting was adapted to maximum one label per sequence (the lowest common ancestor)

158 to mimic results of Kraken2 and other classifiers where only one label per sequence read is given.

159 Preceding classification, Centrifuge builds small reference indexes based on adapted versions of the

160 Burrows-Wheeler transform (BWT) [35] and the Ferragina-Manzini (FM) index [36] resulting in a

161 compressed index of only unique genomic sequences.

162

163 Clark

164 Clark (version 1.2.6.1) [18] is a taxonomic classifier based on reduced *k*-mers using nucleotide-level

165 classification. It uses a compressed index database containing unique target specific *k*-spectrum of

166 target sequences. For the current comparison the default execution mode was used.

167

168 Kaiju

169 Kaiju (version 1.7.3)[27] is a taxonomic classifier that assigns sequence reads using amino acid-level

170 classification. Sequence reads are translated into six possible open reading frames and split into

171 fragments according to the detected stop codons. Classification with Kaiju can be performed using

172 two settings, both based on an adjusted backward alignment search algorithm of BWT [35]. For the

173 current comparison study, the greedy mode was used providing high sensitivity because it allows up

174 to five mismatches to further increase the highest scoring matches. In this mode Kaiju assesses six

175 possible ORF's using the amino acid scores of Blosum62 [37] to obtain the highest scoring match.

176

177 Kraken 2

178 Kraken 2 (version 2.0.8-beta) [28] is a classifier designed to improve the large memory requirements

179 of the former version of Kraken [17], resulting in a reduction of in general 85% of the size of the

180 index database. Kraken 2 uses a probabilistic, compact hash table to map minimizers to the lowest

181 common ancestors (LCA), and stores only minimizers from the reference sequence library in its index

182 reference [28].

183

184 Genome Detective

185 Genome Detective [29] is a commercially-available bioinformatic pipeline that includes the entire
186 workflow from automated quality control, *de novo* assembly of reads and classification of viruses.
187 After adapter trimming and filtering low-quality reads using Trimmomatic [31], viral reads are
188 selected based on Diamond [38] protein alignment using as reference protein sequences from
189 Swissprot Uniref 90 [39]. Viral reads are sorted in buckets, after which all sequences in one bucket
190 are *de novo* assembled into contigs using SPAdes [40] or metaSPAdes [41]. Subsequently, contigs are
191 processed by BLASTx and BLASTn [42] against a databases containing NCBI Refseq [34] sequences
192 and some additional virus sequences. Potential hits represented by the contigs are assigned to
193 individual species using the Advanced Genome Aligner [43], and coverage the viral genomes is
194 calculated.

195

196 Performance, statistical analysis and ROC

197 Sensitivity and specificity were calculated for the classifiers based on application of PCRs (designed
198 for detection of 13 targets) for 88 samples with 24 PCR positive and 1,120 PCR negative results.
199 Receiver Operating Characteristic (ROC) curves were generated for results of classification at
200 species, genus, and family levels, by varying the number of sequence read counts used as cut-off for
201 defining a positive result (resolution: 1,000 steps from 1 read to the maximum number of sequence
202 reads for each PCR target per sample). Area under the curve (AUC), the ROC distance to the closest
203 error- free point (0,1, informedness) curve, positive and negative predictive values were calculated.
204 Furthermore, correlation (R^2) of sequence read counts with PCR cycle threshold (Ct) value were
205 analysed.

206

207 **Results**

208

209 **Performance: sensitivity, specificity, and ROC**

210 The performance of the selected taxonomic classifiers Centrifuge, Clark, Kaiju, Kraken 2, and
211 Genome Detective for metagenomic virus pathogen detection was assessed using datasets from 88
212 respiratory samples with 24 positive and 1,120 negative PCR results available as gold standard. To
213 exclude variability based on different default databases provided with the classifiers, a single
214 database of reference genome sequences was used in combination with a standardized dataset for
215 all classifiers. Raw NGS reads were filtered and classified, both prior and after the exclusion of
216 human sequence reads, and after exclusion of human reads combined with normalization of reads
217 based on the target viral genome length. ROC curves are shown for all classifiers, for assignments at
218 species, genus and family level for the NGS data in **Figure 1**, and **Supplementary Table 1**. Detection
219 parameters (ROC distance to the upper left corner of the graph, sensitivity and selectivity, and AUC)
220 at three taxonomic levels calculated for the NGS data, before and after exclusion of human reads,
221 with or without normalization of assigned reads by corresponding genome sequence lengths are
222 additionally shown in **Figure 2**. Overall, sensitivity, specificity, and AUC ranged 83-100%, 90-99%, 91-
223 98%, respectively, and varied per level of taxonomic classification, per classifier, and with the
224 exclusion of human reads prior to classification. Classification at species and genus levels tended to
225 result in lower sensitivity and higher ROC distances, but higher selectivity when compared with
226 family level classification, for most of the classifiers evaluated. Extraction of human sequence reads
227 prior to classification resulted in comparable sensitivity at all levels of assignment for all classifiers
228 except CLARK for which sensitivity plummeted at species and genus levels. Selectivity was mainly
229 increased after extraction of human reads, for classification at all levels, except for Kaiju and
230 Kraken2, for which decreased selectivity values at family level were observed. Extraction of human
231 reads reduced the differences in selectivity between the classifiers that were observed at genus and
232 family level prior to extraction. The ROC distances were overall smallest, and the AUC highest when

233 using amino-acid based classifier Kaiju, the latter at species and family levels and was comparable
234 with Kraken2 at genus level. Normalization of assigned read counts by corresponding genome length
235 resulted in minor changes in performance when considering 1 read as threshold for defining positive
236 results. Sensitivity was dramatically reduced to 13-33% at species level after read normalization
237 when a threshold of 10 reads was applied, while without read normalization in combination with a
238 threshold of 10 reads, sensitivity was 75-88% (**Supplementary Table 1**). This indicates that
239 normalization of reads can negatively affect the detection of targets with read counts around
240 detection level.

241 Overall, Kaiju outperformed all classifiers when ROC distance, AUC, and sensitivity were considered,
242 but had consistent lower values of selectivity than Centrifuge and Genome Detective.

243 In this patient cohort, with an incidence of 21% (24/88 samples) of respiratory viruses, the positive
244 and negative predictive values at species levels were 42-67% and 99-100% respectively (see
245 **Supplementary Table 1**).

246

247 **Correlation read counts and Ct-values**

248 The correlation between sequence read counts at and Ct-value for corresponding PCR target viruses
249 for all classifiers is shown in **Figure 3 and supplementary Table 2**. Correlation (R^2 , %), linear
250 regression slope and interception varied per virus species, per taxonomic classifier, and was
251 dependent on the extraction of human reads. Correlation ranged from 15.1% for CLARK (no
252 exclusion of human reads, species level) to 62.7% for Kaiju-based classification at species level (after
253 exclusion of human reads with normalization of assigned reads by corresponding genome sequence
254 lengths). The most consistent results (when comparing R^2 prior and after human reads exclusion,
255 and after normalization) were demonstrated by Kaiju and Genome Detective with overall
256 outperformance of Kaiju classifier at all classification levels (61.8-62.7% versus 42.3-43.9% for
257 Centrifuge). Reads assigned to Rhinoviruses were most common outliers in relation to Ct-value and
258 varied up to 3 \log_{10} reads difference from the predicted read count (LR), possibly resulting from their

259 high within species divergence. This was is in contrast to read counts of other viruses (for example
260 Influenza virus), which were closer to the predicted correlation line. Extraction of human sequence
261 reads resulted in an increase in R^2 for CLARK classifier at species and family level, a decrease for
262 Centrifuge and Kraken at all levels, and resulted in minor changes for amino acid-based classifiers
263 Genome Detective and Kaiju at all levels. Decrease in absolute or relative number of total reads after
264 pre-processing (extraction of human reads in combination with normalization of assigned reads by
265 corresponding genome lengths) led to decrease in intercept values for all classifiers.
266 These data support that a more accurate taxonomic classification assists semi-quantitative
267 performance of metagenomic classification tools.
268
269

270 Discussion

271 In this study, we compared the performance of five taxonomic classification tools for virus pathogen
272 detection, using datasets from well-characterized clinical samples. In contrast to previously reported
273 comparisons with datasets from real samples, both sensitivity and specificity could be assessed using
274 a unique set of 1,144 PCR results as gold standard. A uniform database was created to exclude
275 variability based on differences in availability of genomes in databases provided with the classifiers.
276 In general, sensitivity and specificity were within ranges applicable to diagnostic practice. Exclusion
277 of human reads generally resulted in increased specificity. Normalization of read counts for genome
278 length negatively affected the detection of targets with read counts around detection level. The
279 correlation of sequence read counts with PCR Ct-values was highest for viruses with relatively lower
280 sequence diversity.

281 Previous studies have benchmarked metagenomic profilers, mainly for use of bacterial profiling and
282 DNA-to-DNA and DNA-to-protein classification methods were among the best-scoring methods in
283 comparison with DNA-to-marker (16S) methods [22][27][44][45][46][47][48]. In a study with
284 simulated bacterial datasets comparing the performance of CLARK, Kraken and Kaiju, sensitivity and
285 precision were 75% and 95% and decreased when a lower number of reference genomes was
286 available for the specific target [27]. Since in our study the same reference database was used by all
287 classifiers, the only determining factors would be the index database built from the reference
288 database and the classification algorithm. DNA-to-DNA methods have been applied in hundreds of
289 published microbiome studies (e.g., Kraken: 1,438 citations; Kraken 2: 204 citations, by March 2021,
290 according to their official websites [48]). Centrifuge was designed as a follow-up of Kraken with
291 enhanced features, though misclassifications have also been reported in a comparison with
292 simulated datasets [22]. DNA-to-protein methods are generally more sensitive to novel and highly
293 variable sequences due to lower mutation rates of amino acid compared to nucleotide sequences
294 [22], [27] as was seen in our study when classifying rhinoviruses by Kaiju. The difference was
295 especially visible in genera with limited availability of genomes in reference databases [27].

296 Misclassification of human genomic sequence reads has been reported for most DNA classifiers [22].
297 Protein-based classifiers had higher misclassification ranges of human genome sequences (up to
298 15%), partially due to the larger number of target sequences in their default databases [22].
299 Inclusion of the human genome in the reference database, which is by default the case for
300 Centrifuge and KrakenUniq [49] reduced the rate of misclassification to negligible [22]. In line, in our
301 study, exclusion of human sequence reads prior to classification reduced misclassifications for all
302 classifiers. In general, reduction of false-positive hits can be achieved by assembly of sequences (for
303 example, by Genome Detective) and thus reducing the number of hits based on short nucleotide
304 sequences used by *k*-mer based methods. Inclusion of genome coverage of mapped reads, as
305 adopted by Genome Detective and KrakenUniq [49], also can reduce false-positive hits.
306 To our knowledge, a limited number of studies have published on the benchmarking of tools for viral
307 metagenomics for pathogen detection. In a Swiss-wide ring trial based on spiked plasma samples,
308 median F1 scores ranged from 70-100% for the different pipelines, though since the entire workflow
309 was analysed, no conclusions on specific classifiers could be drawn [15]. A series of tools and
310 programs were analysed in a COMPARE virus proficiency test using a single *in silico* dataset [14]. For
311 Kraken discrepant classification results were observed, likely due to differences in the databases
312 used by the participants. A recent European benchmark of thirteen bioinformatic pipelines currently
313 in use for metagenomic virus diagnostics used datasets from clinical samples [16] Analyses using
314 Centrifuge, and Genome Detective software resulted in sensitivities of 93% and 87% respectively.
315 In conclusion, sensitivity and specificity of the classifiers evaluated in this study was within the
316 ranges that may be applied in clinical diagnostic settings. Performance testing for viral
317 metagenomics for pathogen detection is intrinsically different from benchmarking of bacterial
318 profiling, and should incorporate parameters that are inherent to clinical diagnostic use such as
319 specificity calculations, sensitivity for divergent viruses and variants, and importantly, a determined
320 cut-off for defining a positive result for each workflow. Taking these factors into account during

321 validation and implementation of viral metagenomics for pathogen detection contributes to optimal
322 performance and applicability in clinical diagnostic settings.

323

324 **Data access**

325 NGS data used in this study have been submitted (after removal of human reads) to the NCBI's

326 Sequence Read Archive (<http://www.ncbi.nlm.nih.gov>; accession number SRX6713943-SRX6714030).

327

328 **Disclosures**

329 None

330

331 References

- 332 [1] M. R. Wilson *et al.*, 'Clinical Metagenomic Sequencing for Diagnosis of Meningitis and
333 Encephalitis', *N. Engl. J. Med.*, vol. 380, no. 24, pp. 2327–2340, Jun. 2019, doi:
334 10.1056/NEJMoa1803396.
- 335 [2] F. X. López-Labrador *et al.*, 'Recommendations for the introduction of metagenomic high-
336 throughput sequencing in clinical virology, part I: Wet lab procedure', *J. Clin. Virol.*, vol. 134, p.
337 104691, Jan. 2021, doi: 10.1016/j.jcv.2020.104691.
- 338 [3] J. J. C. de Vries *et al.*, 'Recommendations for the introduction of metagenomic next-generation
339 sequencing in clinical virology, part II: bioinformatic analysis and reporting', *J. Clin. Virol.*, vol.
340 138, p. 104812, May 2021, doi: 10.1016/j.jcv.2021.104812.
- 341 [4] E. C. Carbo, I. Blankenspoor, J. J. Goeman, A. C. M. Kroes, E. C. J. Claas, and J. J. C. De Vries,
342 'Viral metagenomic sequencing in the diagnosis of meningoencephalitis: a review of technical
343 advances and diagnostic yield', *Expert Rev. Mol. Diagn.*, vol. 21, no. 11, pp. 1139–1146, Nov.
344 2021, doi: 10.1080/14737159.2021.1985467.
- 345 [5] C. Y. Chiu and S. A. Miller, 'Clinical metagenomics', *Nat. Rev. Genet.*, vol. 20, no. 6, pp. 341–
346 355, Jun. 2019, doi: 10.1038/s41576-019-0113-7.
- 347 [6] W. Gu, S. Miller, and C. Y. Chiu, 'Clinical Metagenomic Next-Generation Sequencing for
348 Pathogen Detection', *Annu. Rev. Pathol. Mech. Dis.*, vol. 14, no. 1, pp. 319–338, Jan. 2019, doi:
349 10.1146/annurev-pathmechdis-012418-012751.
- 350 [7] A. Reyes *et al.*, 'Viral metagenomic sequencing in a cohort of international travellers returning
351 with febrile illness', *J. Clin. Virol.*, vol. 143, p. 104940, Oct. 2021, doi:
352 10.1016/j.jcv.2021.104940.
- 353 [8] J. R. Brown, T. Bharucha, and J. Breuer, 'Encephalitis diagnosis using metagenomics:
354 application of next generation sequencing for undiagnosed cases', *J. Infect.*, vol. 76, no. 3, pp.
355 225–240, Mar. 2018, doi: 10.1016/j.jinf.2017.12.014.
- 356 [9] E. C. Carbo *et al.*, 'Improved diagnosis of viral encephalitis in adult and pediatric hematological
357 patients using viral metagenomics', *J. Clin. Virol.*, p. 104566, Jul. 2020, doi:
358 10.1016/j.jcv.2020.104566.
- 359 [10] C. Y. Chiu *et al.*, 'Diagnosis of Fatal Human Case of St. Louis Encephalitis Virus Infection by
360 Metagenomic Sequencing, California, 2016', *Emerg. Infect. Dis.*, vol. 23, no. 10, pp. 1964–1968,
361 Oct. 2017, doi: 10.3201/eid2310.161986.
- 362 [11] M. Christopheit *et al.*, 'Suspected encephalitis with *Candida tropicalis* and *Fusarium* detected by
363 unbiased RNA sequencing', *Ann. Hematol.*, vol. 95, no. 11, pp. 1919–1921, Nov. 2016, doi:
364 10.1007/s00277-016-2770-3.
- 365 [12] A. W. D. Edridge *et al.*, 'Novel Orthobunyavirus Identified in the Cerebrospinal Fluid of a
366 Ugandan Child With Severe Encephalopathy', *Clin. Infect. Dis.*, vol. 68, no. 1, pp. 139–142, Jan.
367 2019, doi: 10.1093/cid/ciy486.
- 368 [13] E. C. Carbo *et al.*, 'Coronavirus discovery by metagenomic sequencing: a tool for pandemic
369 preparedness', *J. Clin. Virol.*, vol. 131, p. 104594, Oct. 2020, doi: 10.1016/j.jcv.2020.104594.
- 370 [14] A. Brinkmann *et al.*, 'Proficiency Testing of Virus Diagnostics Based on Bioinformatics Analysis
371 of Simulated *In Silico* High-Throughput Sequencing Data Sets', *J. Clin. Microbiol.*, vol. 57, no. 8,
372 pp. e00466-19, /jcm/57/8/JCM.00466-19.atom, Jun. 2019, doi: 10.1128/JCM.00466-19.
- 373 [15] Junier *et al.*, 'Viral Metagenomics in the Clinical Realm: Lessons Learned from a Swiss-Wide
374 Ring Trial', *Genes*, vol. 10, no. 9, p. 655, Aug. 2019, doi: 10.3390/genes10090655.
- 375 [16] J. J. C. de Vries *et al.*, 'Benchmark of thirteen bioinformatic pipelines for metagenomic virus
376 diagnostics using datasets from clinical samples', *J. Clin. Virol.*, Aug. 2021 doi:
377 10.1016/j.jcv.2021.104908.
- 378 [17] D. E. Wood and S. L. Salzberg, 'Kraken: ultrafast metagenomic sequence classification using
379 exact alignments', *Genome Biol.*, vol. 15, no. 3, p. R46, 2014, doi: 10.1186/gb-2014-15-3-r46.

- 380 [18] R. Ounit, S. Wanamaker, T. J. Close, and S. Lonardi, 'CLARK: fast and accurate classification of
381 metagenomic and genomic sequences using discriminative k-mers', *BMC Genomics*, vol. 16, no.
382 1, p. 236, Dec. 2015, doi: 10.1186/s12864-015-1419-2.
- 383 [19] P. Simmonds *et al.*, 'Virus taxonomy in the age of metagenomics', *Nat. Rev. Microbiol.*, vol. 15,
384 no. 3, pp. 161–168, Mar. 2017, doi: 10.1038/nrmicro.2016.177.
- 385 [20] S. Nooij, D. Schmitz, H. Vennema, A. Kroneman, and M. P. G. Koopmans, 'Overview of Virus
386 Metagenomic Classification Methods and Their Biological Applications', *Front. Microbiol.*, vol.
387 9, p. 749, Apr. 2018, doi: 10.3389/fmicb.2018.00749.
- 388 [21] A. Escobar-Zepeda *et al.*, 'Analysis of sequencing strategies and tools for taxonomic
389 annotation: Defining standards for progressive metagenomics', *Sci. Rep.*, vol. 8, no. 1, p. 12034,
390 Dec. 2018, doi: 10.1038/s41598-018-30515-5.
- 391 [22] S. H. Ye, K. J. Siddle, D. J. Park, and P. C. Sabeti, 'Benchmarking Metagenomics Tools for
392 Taxonomic Classification', *Cell*, vol. 178, no. 4, pp. 779–794, Aug. 2019, doi:
393 10.1016/j.cell.2019.07.010.
- 394 [23] N. Couto *et al.*, 'Critical steps in clinical shotgun metagenomics for the concomitant detection
395 and typing of microbial pathogens', *Sci. Rep.*, vol. 8, no. 1, p. 13767, Dec. 2018, doi:
396 10.1038/s41598-018-31873-w.
- 397 [24] M. Asplund *et al.*, 'Contaminating viral sequences in high-throughput sequencing viromics: a
398 linkage study of 700 sequencing libraries', *Clin. Microbiol. Infect.*, vol. 25, no. 10, pp. 1277–
399 1285, Oct. 2019, doi: 10.1016/j.cmi.2019.04.028.
- 400 [25] S. van Boheemen *et al.*, 'Retrospective Validation of a Metagenomic Sequencing Protocol for
401 Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric
402 Patients', *J. Mol. Diagn.*, vol. 22, no. 2, pp. 196–207, Feb. 2020, doi:
403 10.1016/j.jmoldx.2019.10.007.
- 404 [26] D. Kim, L. Song, F. P. Breitwieser, and S. L. Salzberg, 'Centrifuge: rapid and sensitive
405 classification of metagenomic sequences', *Genome Res.*, vol. 26, no. 12, pp. 1721–1729, Dec.
406 2016, doi: 10.1101/gr.210641.116.
- 407 [27] P. Menzel, K. L. Ng, and A. Krogh, 'Fast and sensitive taxonomic classification for metagenomics
408 with Kaiju', *Nat. Commun.*, vol. 7, no. 1, p. 11257, Sep. 2016, doi: 10.1038/ncomms11257.
- 409 [28] D. E. Wood, J. Lu, and B. Langmead, 'Improved metagenomic analysis with Kraken 2', *Genome
410 Biol.*, vol. 20, no. 1, p. 257, Dec. 2019, doi: 10.1186/s13059-019-1891-0.
- 411 [29] M. Vilsker *et al.*, 'Genome Detective: an automated system for virus identification from high-
412 throughput sequencing data', *Bioinformatics*, vol. 35, no. 5, pp. 871–873, Mar. 2019, doi:
413 10.1093/bioinformatics/bty695.
- 414 [30] A. L. van Rijn *et al.*, 'The respiratory virome and exacerbations in patients with chronic
415 obstructive pulmonary disease', *PLOS ONE*, vol. 14, no. 10, p. e0223952, Oct. 2019, doi:
416 10.1371/journal.pone.0223952.
- 417 [31] A. M. Bolger, M. Lohse, and B. Usadel, 'Trimmomatic: a flexible trimmer for Illumina sequence
418 data', *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014, doi:
419 10.1093/bioinformatics/btu170.
- 420 [32] 'GRCh38'. [Online]. Available: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/
- 421 [33] B. Langmead and S. L. Salzberg, 'Fast gapped-read alignment with Bowtie 2', *Nat. Methods*, vol.
422 9, no. 4, pp. 357–359, Apr. 2012, doi: 10.1038/nmeth.1923.
- 423 [34] N. A. O'Leary *et al.*, 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic
424 expansion, and functional annotation', *Nucleic Acids Res.*, vol. 44, no. D1, pp. D733–D745, Jan.
425 2016, doi: 10.1093/nar/gkv1189.
- 426 [35] Burrows, M.; Wheeler, David J., 'A block-sorting lossless data compression algorithm.', vol.
427 Digital Equipment Corporation, no. Technical Report 124, 1994.
- 428 [36] P. Ferragina and G. Manzini, 'Opportunistic data structures with applications', in *Proceedings
429 41st Annual Symposium on Foundations of Computer Science*, Redondo Beach, CA, USA, 2000,
430 pp. 390–398. doi: 10.1109/SFCS.2000.892127.

- 431 [37] S. Henikoff and J. G. Henikoff, 'Amino acid substitution matrices from protein blocks.', *Proc.*
432 *Natl. Acad. Sci.*, vol. 89, no. 22, pp. 10915–10919, Nov. 1992, doi: 10.1073/pnas.89.22.10915.
- 433 [38] B. Buchfink, C. Xie, and D. H. Huson, 'Fast and sensitive protein alignment using DIAMOND',
434 *Nat. Methods*, vol. 12, no. 1, pp. 59–60, Jan. 2015, doi: 10.1038/nmeth.3176.
- 435 [39] B. E. Suzek, H. Huang, P. McGarvey, R. Mazumder, and C. H. Wu, 'UniRef: comprehensive and
436 non-redundant UniProt reference clusters', *Bioinformatics*, vol. 23, no. 10, pp. 1282–1288, May
437 2007, doi: 10.1093/bioinformatics/btm098.
- 438 [40] A. Bankevich *et al.*, 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-
439 Cell Sequencing', *J. Comput. Biol.*, vol. 19, no. 5, pp. 455–477, May 2012, doi:
440 10.1089/cmb.2012.0021.
- 441 [41] S. Nurk, D. Meleshko, A. Korobeynikov, and P. A. Pevzner, 'metaSPAdes: a new versatile
442 metagenomic assembler', *Genome Res.*, vol. 27, no. 5, pp. 824–834, May 2017, doi:
443 10.1101/gr.213959.116.
- 444 [42] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 'Basic local alignment search
445 tool', *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990, doi: 10.1016/S0022-2836(05)80360-
446 2.
- 447 [43] K. Deforche, 'An alignment method for nucleic acid sequences against annotated genomes',
448 *Bioinformatics*, preprint, Oct. 2017. doi: 10.1101/200394.
- 449 [44] K. Mavromatis *et al.*, 'Use of simulated data sets to evaluate the fidelity of metagenomic
450 processing methods', *Nat. Methods*, vol. 4, no. 6, pp. 495–500, Jun. 2007, doi:
451 10.1038/nmeth1043.
- 452 [45] F. Meyer, A. Bremges, P. Belmann, S. Janssen, A. C. McHardy, and D. Koslicki, 'Assessing
453 taxonomic metagenome profilers with OPAL', *Genome Biol.*, vol. 20, no. 1, p. 51, Dec. 2019,
454 doi: 10.1186/s13059-019-1646-y.
- 455 [46] A. Sczyrba *et al.*, 'Critical Assessment of Metagenome Interpretation—a benchmark of
456 metagenomics software', *Nat. Methods*, vol. 14, no. 11, pp. 1063–1071, Nov. 2017, doi:
457 10.1038/nmeth.4458.
- 458 [47] A. B. R. McIntyre *et al.*, 'Comprehensive benchmarking and ensemble approaches for
459 metagenomic classifiers', *Genome Biol.*, vol. 18, no. 1, p. 182, Dec. 2017, doi: 10.1186/s13059-
460 017-1299-7.
- 461 [48] Z. Sun *et al.*, 'Challenges in benchmarking metagenomic profilers', *Nat. Methods*, vol. 18, no. 6,
462 pp. 618–626, Jun. 2021, doi: 10.1038/s41592-021-01141-3.
- 463 [49] F. P. Breitwieser, D. N. Baker, and S. L. Salzberg, 'KrakenUniq: confident and fast metagenomics
464 classification using unique k-mer counts', *Genome Biol.*, vol. 19, no. 1, p. 198, Dec. 2018, doi:
465 10.1186/s13059-018-1568-0.
- 466

467 **Table 1.** Overview of respiratory PCR panel targets and their test results.

PCR target viruses	Family	Genus	Species	Alternative naming	# PCR positive samples	# PCR negative samples	PCR Ct-values (range)
HRV	Picorna- viridae	Enterovirus	Rhinovirus A, B, C, Enterovirus D		14	74	19-38
PIV1, PIV3	Paramyxo- viridae	Respiro- virus	Human respirovirus 1	Human parainfluenza virus 1	-	88	-
			Human respirovirus 3	Human parainfluenza virus 3	2	86	26-36
PIV2, PIV4	Paramyxo- viridae	Ortho- rubulavirus	Human orthorubulavirus 2	Human parainfluenza virus 2	-	88	-
			Human orthorubulavirus 4	Human parainfluenza virus 4	1	87	24
INF	Orthomyxo- viridae	Alphainflue- nzavirus	Influenza A virus Influenza B virus		3	85	29-36
					-	88	-
ACoV	Corona- viridae	Alpha- coronavirus	Human coronavirus NL63		2	86	32
			Human coronavirus 229E		-	88	-
BCoV	Corona- viridae	Betacorona- virus	Human coronavirus HKU1, Betacoronavirus 1; Human coronavirus OC43		2	86	27
HMPV	Pneumo- viridae	Metapneu- movirus	Human metapneumo- virus		-	88	-
RSV	Pneumo- viridae	Orthopneu- movirus	Human orthopneumo- virus		-	88	-
Total			Total PCR results: 1,144 (13 targets tested in 88 samples)		24	1,120	19-38

468

469 **Table 2.** Overview of characteristics of the classifiers evaluated.

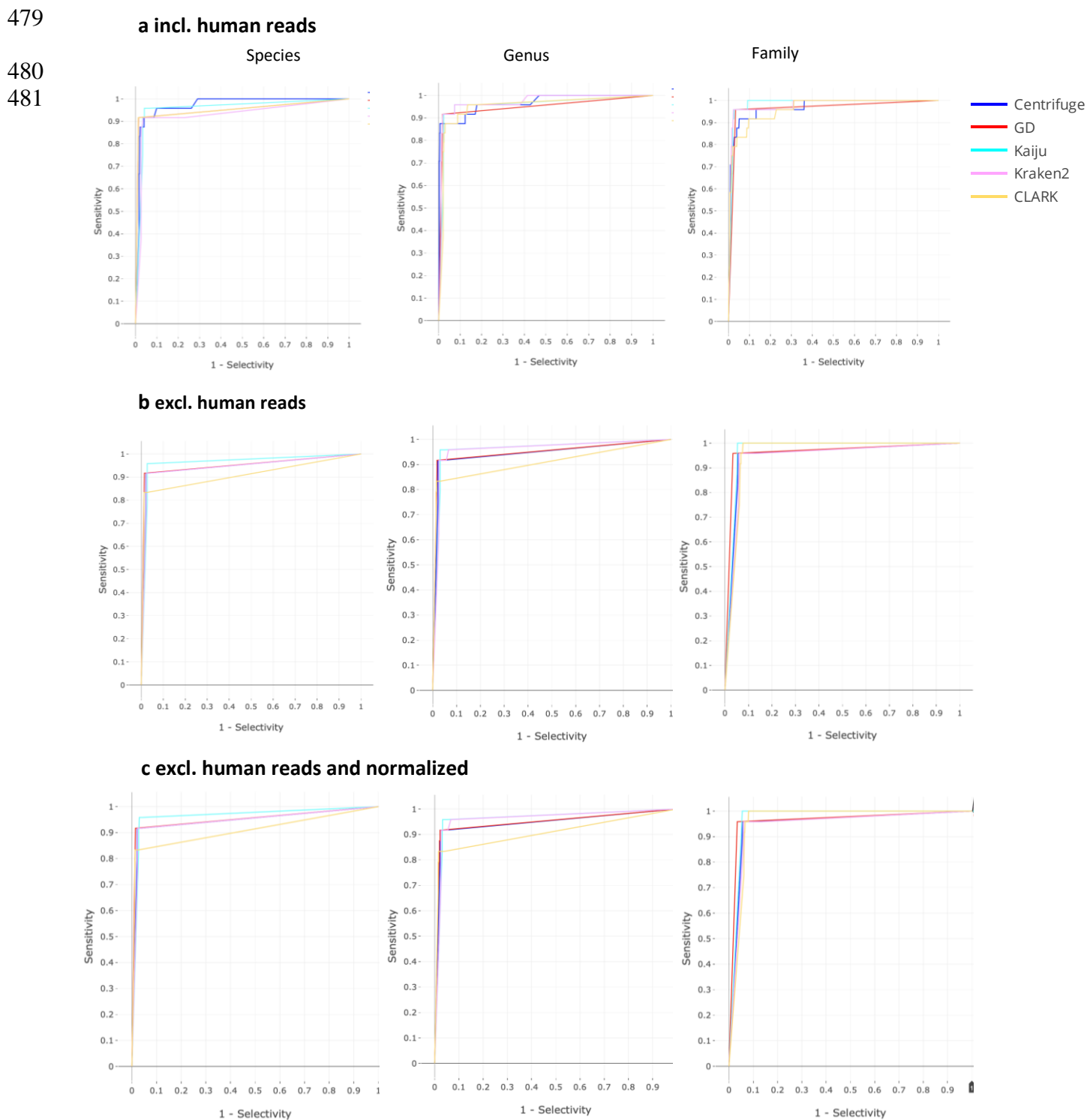
	Centrifuge [26]	Clark [18]	Kaiju [27]	Kraken 2 [28]	Genome Detective [29]
License	Open source	Open source	Open source	Open source	Commercial/free to use web application
Version	1.0.4	1.2.6.1	1.7.3	2.0.8-beta	1.126
Sequencing technology compatibility	Short/long reads	Short/long reads	Short/long reads	Short/long reads	Short reads (long reads experimentally)
Pre-processing	No	No	No	No	Yes
Type of alignment	NT	NT	AA	NT	NT/AA (DIAMOND [38]) including <i>de novo</i> assembly
Algorithm characteristics	Exact matches of 22 bp with target with default 5 labels per sequence, LCA optional	Exact matches of 31 bp with target with highest number of hits	Maximum exact matches (MEM) of AA, up to 5 mismatched optional*. LCA in case of multiple hits	Exact matches of 35 bp. LCA in case of multiple hits	Combined results of NT and AA hits based on scoring. LCA in case of multiple hits
Database (compression)	Compressed index NT database of only unique sequences	Compressed index NT database of only unique sequences	No compression, AA database	Compressed index NT database	No compression, viral subset of Swiss-Prot UniRef90 protein database

470 NT; nucleotide, AA; amino acid; LCA, lowest common ancestor

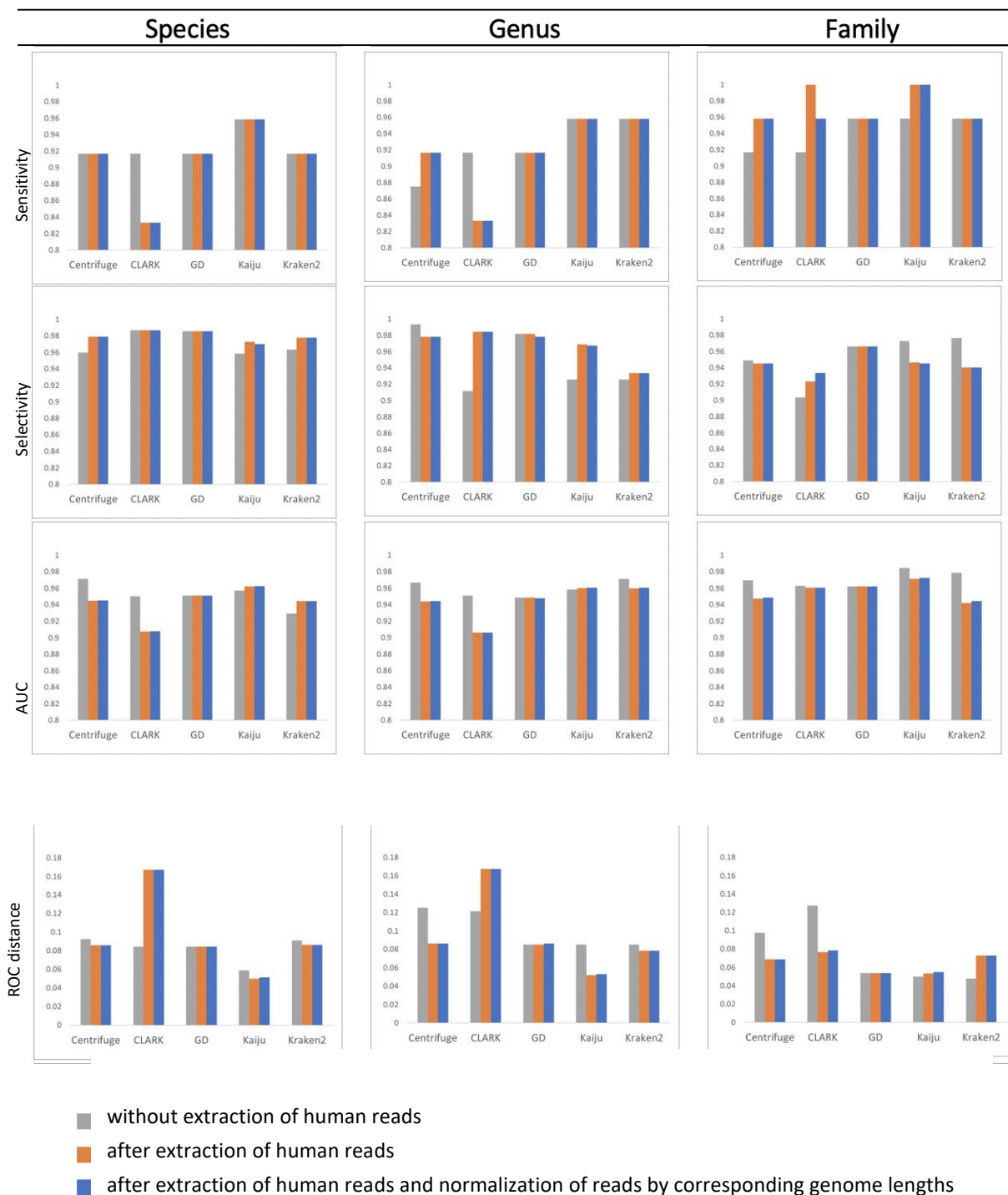
471 *Greedy-5 mode was used in the current study

472

473 **Figure 1.** ROC curves calculated based on reads of taxonomic assignment at three
474 taxonomic levels (species, genus, and family) by the five classifiers, based on PCR-targets,
475 **a.** without extraction of human reads and **b.** after extraction of human reads, **c.** after extraction of
476 human reads and normalization of reads by corresponding genome lengths (resolution of 1,000
477 steps from 1 read to the maximum number of sequence reads for each PCR target per sample).
478



482 **Figure 2.** Sensitivity, selectivity, AUC, and ROC distance calculated based on assignment at three
 483 taxonomic levels (species, genus, and family) by the five classifiers for three type of pre-processing
 484 of the NGS datasets, **a.** without extraction of human reads and **b.** after extraction of human reads, **c.**
 485 after extraction of human reads and normalization of reads by corresponding genome lengths.
 486
 487



488 **Figure 3.** Correlation between the number of sequence reads assigned (species level) and Ct-values
 489 of virus-specific PCRs, for the five taxonomic classifiers evaluated, **a.** without extraction of human
 490 reads and **b.** after extraction of human reads, **c.** after normalization of reads by corresponding
 491 genome lengths.
 492

