HIV-1 protein coding sequences are present in relevant bacteria

Hector F. Pelaez-Prestel[*], Juan Mozas-Gutierrez[*], Esther M. Lafuente and Pedro A. Reche[†]

Laboratory of Immunomedicine & Immunoinformatics, Department of Immunology, Ophthalmology and ORL (IO2), School of Medicine, Complutense University of Madrid, Pza Ramon y Cajal, s/n, 28040, Madrid, Spain.

[*]Authors contributed equally

[†] Correspondence author: Dr. Pedro A Reche; EM: parecheg@med.ucm.es

ABSTRACT

Human Immunodeficiency Virus (HIV) is a retrovirus that attacks the human immune system and can cause acquired immunodeficiency syndrome (AIDS). Early diagnosis and treatment of HIV infected individuals are considered key to reduce HIV transmission and developing AIDS. Therefore, HIV diagnostics play an important role in the battle against AIDS. Most HIV tests are aimed to detect HIV antibodies and/or antigens, particularly p24, but there also tests that detect HIV genomic material. HIV tests are regarded as reliable and very specific. However, false-positives are known to occur, usually caused by infections with unrelated pathogens leading to cross-reactive antibodies. In this work, we found through TBLASTN searches that the genome of several bacterial species, most frequently *Klebsiella pneumoniae and Escherichia coli*, contain segments matching HIV-1 proteins, including p24 and other relevant proteins for HIV testing, with a high degree of similarity (up to 100 % sequence identity). The presence of HIV-1 in these bacteria of the human micriobiota does not appear to be an artifact since HIV-1 proteins were detected in different strains/isolates of the same species. The proteome of other common viruses, particularly Influenza A virus and Hepatitis B virus, were also detected in bacterial genomes, but to a much lesser degree. Overall, our findings support that some bacteria can acquire HIV-1 genetic material and could interfere with HIV-1 testing, causing false-positives.

Keywords: HIV, HIV diagnostics, Bacteria, *Klebsiella pneumoniae, Escherichia coli,* Cross-reactivity

34 **1. INTRODUCTION**

35

36 Human Immunodeficiency Virus (HIV) is an enveloped retrovirus that attacks $CD4^+$ T cells,
37 leading to their gradual depletion and the development of acquired immunodeficiency syndrome
38 (AIDS) (1,2). HIV transmission occurs mainly through sexual contact, often between men who
39 have sex with men but in Africa, where heterosexual sex transmission is dominant (3). The
40 virus can be categorized into 2 distinct subtypes: HIV type 1 (HIV-1) and HIV type 2 (HIV-2).
41 Both viruses are transmitted similarly, show the same tropism and cause AIDS (4). However,
42 infections by HIV-1 are much more common and progress quicker to AIDS than HIV-2
43 infections (4).

44

45 AIDS is the last stage of HIV infection, where the clinical symptoms of a weakened immune
46 system manifest (5,6). The first stage is the acute phase, which can pass unnoticed or manifest
47 with flu-like symptoms. During this phase, lasting 2-4 weeks after infection, the virus replicates
48 quickly, spreading through the body, and HIV genome and antigens, notably HIV-1 p24, can be
49 detected in blood. The second stage is the chronic or asymptomatic phase, in which the host
50 starts producing antibodies (Ab) and becomes HIV seropositive. In this stage, HIV antigens
51 drop and may be undetectable, yet viral replication and destruction of $CD4^+$ T cells continue and
52 patients eventually enter the AIDS stage unless they receive antiretroviral therapy (ART) (5,6).

53

54 There is evidence that starting ART early in the course of HIV infection has long-term health
55 benefits for patients and decreases the risk of transmission (7,8). Therefore, ART is currently
56 recommended for all HIV positive people, regardless of $CD4^+$ T cell count (9). Moreover, in
57 order to decrease the risk of infection, antiretroviral prophylaxis is also recommended for
58 perinatally-exposed newborns and people who has been exposed to contagious HIV-positive
59 material (9). Despite the noted benefits, ARTs do not eliminate the virus and have a
60 considerably degree of toxicity and long-term adverse effects (10). Thereby, patients in ART are
61 continuously monitored, and may require adjustments of the ART regime and additional

62 supportive treatments, all of which represent a considerable burden to both, patients and health
63 care systems (10).
64
65 HIV diagnostics play a relevant role in the battle against HIV/AIDS, facilitating early initiation
66 of HIV treatments. Moreover, they are also useful to stage and monitor patients, and for
67 outbreak surveillance. HIV tests have historically relied on commercial enzyme immunoassays
68 (EIAs) that detect antibodies to HIV antigens in serum (5,11). Current generation of commercial
69 HIV EIA tests (4-th and 5-th generation), detects Abs against HIV-1 and HIV-2 as well as HIV-
70 1 p24 antigen, which allows for earlier diagnosis of HIV-1 infection (5). A list of FDA approved
71 laboratory Antibody/Antigen (Ab/Ag) HIV tests are available at
72 https://www.cdc.gov/hiv/partners/testing/laboratorytests.html (accessed on October/10/2024).
73 In these tests, HIV-1 p24 is detected by using Abs generated against this protein or fragments
74 derived from it, while HIV antibodies are detected by using HIV antigens and/or antigen
75 fragment cocktails. The following antigens are generally considered for HIV-1 testing (coding
76 gene in italic): gp160 (envelope protein precursor, *env*), gp120 (extracellular region of envelope
77 protein), gp41 (transmembrane region of envelope protein), gp31 (immunodominant region on
78 integrase, *pol*), p51 (reverse transcriptase, *pol*), p24 (core protein, *gag*) and p17 (matrix protein,
79 *gag*). In the laboratory, detection of Ab is sufficient to confirm HIV infection but differentiation
80 IEA tests may be necessary to distinguish between HIV-1 and HIV-2 infection. If the Ag test for
81 HIV-1 p24 is positive but the Ab is negative, supplemental confirmatory IEA tests (e.g.
82 western-blotting), or RNA tests will be necessary to confirm or negate HIV-1 infection (12).
83 HIV RNA tests involve reverse transcription of HIV genetic material into DNA and are aimed
84 to detect conserved regions in *gag* or *pol* genes (5).
85
86 HIV tests are regarded as very reliable and specific. However, false-positives (FPs) can occur
87 due to technical reasons (e.g. misinterpretation and clerical errors) and more importantly under
88 certain biological or medical conditions (13-15). Relevant conditions that have been reported to
89 interfere with HIV IEA tests and produce FP results include: pregnancy (16), recent vaccination

90 (e.g. flu, Tdap or Hepatitis B vaccine) (17-19), cancer (19), and autoimmune diseases like lupus
91 (20) or rheumatoid arthritis (20). Likewise, various infectious diseases (e.g. schistosomiasis,
92 malaria, COVID-19 and Epstein-Barr virus mononucleosis) have been shown to produce FPs
93 (21-24), likely due to cross-reactivity.
94
95 In this work, we investigated potential cross-reactivity between bacteria and HIV-1, using
96 TBLASTN searches against bacterial genomes with the HIV-1 proteome. Interestingly, we
97 detected the presence of HIV-1 protein coding sequences in many bacterial species but more
98 frequently in *Klebsiella pneumoniae* and *Escherichia coli,* which are part of the human
99 microbiota. Sequence identity between HIV-1 and bacterial genomes was not only high (up to
100 100%) but reached the entire HIV-1 proteome, including p24 and other relevant proteins in
101 HIV-1 testing. Other common viruses were also detected in bacterial genomes, but to a much
102 lesser degree. These results support that some bacteria can acquire HIV-1 sequences, which
103 could potentially induce immune cross-reactivity to HIV-1, causing FP results in HIV tests.
104 Ruling out this possibility will reduce the chance of putting healthy individuals on ART and
105 alleviate the economic burden of HIV treatments on health care systems.
106

2. METHODS

Genomes and proteins

The HIV-1 reference genome (HIV-1 group M isotype B, isolate HXB2) was obtained from NCBI (Accession: NC_001802), and amino acid coding sequences (CDS) and mature proteins were collected upon it in a FASTA file. The entire proteomes of nine common human viruses, consisting of the CDS in the relevant NCBI genomic sequences, were also collected in FASTA files (Table 1).

Table 1. Proteomes of common human viruses considered in this study

Accession	Virus	CDS	Amino Acids
NC_038311	Human rhinovirus A	1	2157
NC_009996	Human rhinovirus C	1	2144
NC_038235	Respiratory Syncitial virus A	11	4536
NC_001781	Respiratory Syncitial virus B	11	4540
NC_026422_29*	Influenza A virus	12	4784
NC_002204_11*	Influenza B virus	10	4609
NC_003977	Hepatitis B virus	8	2299
M62321	Hepatitis C virus	1	3011
NC_027779	Papiloma virus	6	2309

* Proteome is contained in various independent accessions varying the last two digits and the range is indicated. The corresponding author will provide the FASTA files upon written request.

Microbial genome BLAST searches

Sequence similarity searches of viral proteins in microbial genomes were carried online at NCBI BLAST site (https://blast.ncbi.nlm.nih.gov/Blast.cgi), selecting TBLASTN. The amino acid sequences of viral proteomes/proteins were entered as query and the RefSeq Genome Database was selected as the target database, restricting the search to Bacteria (taxid:2). TBLASTN searches were carried out with default settings but increasing the maximum target to

1000. The raw output from searches was downloaded as text and BLAST summary results with taxonomic information were saved in Excel (Microsoft). The taxonomic TBLASTN information tabulated the bacterial species with TBLASTN hits, indicating the number of hits and top bit-score of all hits *per* bacteria as well.

Similar TBLASTN searches were carried out to determine bit-scores of viral proteomes/proteins with themselves but restricting the search in RefSeq Genome Database to Viruses (taxid:10239). As result, for each protein a self bit-score resulting from a hit alignment with itself was obtained and recorded. Bit-scores only depend on the hit sequence alignment, not the size of the database. Hence, the self bit-score (sBS) of a viral sequence query corresponds to the maximum bit-score (mBS) that the query sequence can reach.

Relative bit-scores

Relative bit-scores (RBSs) of TBLASTN hits between viral protein queries and subject bacterial genomic sequences were obtained by dividing bit-scores (BS) with the corresponding mBS, as indicated in Equation 1.

$$RBS = \frac{BS}{mBS} \times 100 \qquad \text{Eq. 1}$$

Where mBS is the maximum bit-score of the relevant viral protein sequence obtained from a self-hit alignment, as defined previously. Since a given viral protein query sequence can produce numerous TBLASTN hits to bacterial sequences from the same species, the top RBS for each viral protein per bacterial species was computed after the best hit, that with the top bit-score.

Viral proteome coverage

154

155 The viral proteome coverage (VPC) was introduced as an estimation of the overall presence and

156 similarity of an entire viral proteome (*p*) in a given bacterial species (*b*). For any given viral

157 proteome used in TBLASTN searches, VPC (*p,b*) was obtained using Equation 2, which tallies

158 the top RBS (tRBS) from the best hit of each protein in the viral proteome (*p:1 .. N*) *per*

159 bacterial species (b), divided by the total number of proteins, N, in the viral proteome.

160

161 $$VPC_{(p,b)} = \frac{\sum_{p=N}^{p=1} tRBS_{(p,b)}}{N} \qquad \text{Eq.2}$$

162

163 **Other procedures**

164

165 TBLASTN results were parsed using PERL scripts to compute RBSs of hit alignments and data

166 was processed in Excel (Microsoft, Corp.). All graphics were generated using GraphPad Prism

167 (GraphPad Software, Inc.).

168

169 **3. RESULTS**

170

171 **Detection of HIV-1 proteins in bacterial genomes**

172

173 We investigated bacteria as potential sources of immune cross-reactivity with HIV-1 by
174 queering bacterial genomes with HIV-1 protein sequences, using TBLASTN (details in
175 Methods). The results revealed that all HIV-1 proteins produced highly significant TBLASTN
176 hits in the genome of various bacterial species (TBLASTN output description summary is
177 provided in Supplementary file 1). The single most significant hits with bacteria were produced
178 by p24, Vpr and Protease, which reached a relative bit-score (RBS) of 100% (Fig 1). The RBS
179 of a hit is computed by dividing the bit-score of the corresponding alignment between the viral
180 protein query and the subject genome by the maximum possible bit-score of the protein query,
181 obtained from a self-hit (details in Methods). Thus, an RBS of 100% for p24, Vpr and Protease
182 indicates that these proteins produce hit alignments spanning their entire length with 100%
183 sequence identity. As shown in Figure 1A, all HIV-1 proteins produced a considerable number
184 of hits with bacteria genomes with RBS ≥ 50 % and the RBS of the best hit was ≥ 60% for all
185 HIV-1 proteins. Interestingly, only a handful of bacterial species, including *Escherichia coli*
186 and *Klebsiella pneumoniae*, were the subject of the best hits with HIV-1 proteins (Fig.1). These
187 best hits between HIV-1 proteins and bacteria had a sequence identity ≥ 66.7 % and covered at
188 least 60% of the length of query (Vpu) (Table 1).

189

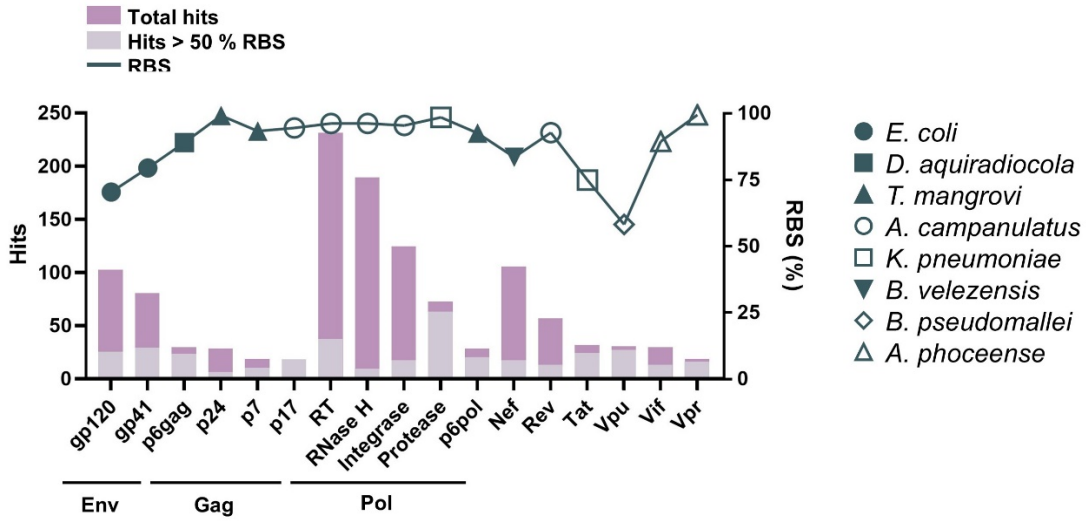

Figure 1. Presence of HIV-1 proteins in bacterial genomes. Chart showing for each HIV-1 protein (x-axis) the number of **TBLASTN** hits with bacterial genomes (left y-axis). Stacked columns in light grey show the number of hits with RBS ≥ 50%. The RBS (right y-axis) of the best hit for each HIV-1 protein is also plotted in the chart, using different symbols that represent the bacterial species, indicated on the right, subject of the hit. HIV-1 proteins that derive from Env, Gag and Pol are underlined.

Table 1. Statistics of best TBLASTN hits between HIV-proteins and bacterial genomes

Viral protein	Bacteria	E-value	Bit-Score	mBS[1]	RBS[2]	Q-coverage	Identity
gp120	*E. coli*	0	703	999	70.37%	99%	70.88%
gp41	*E. coli*	4×10−147	428	551	79.49%	92%	67.79%
p24	*T. mangrovi*	1×10−159	478	482	99.17%	100%	98.27%
p17	*A. campanulatus*	9×10−64	223	236	94.49%	100%	80.92%
p7	*T. mangrovi*	2×10−26	112	120	93.33%	100%	90.91%
p6gag	*D. aquiradiocola*	4×10−20	94.4	106	89.06%	100%	90.38 %
RT p51	*A. campanulatus*	0	1073	1096	97.9%	100%	94.29%
Protease	*K. pneumoniae*	3×10−56	199	202	98.51%	100%	97.98%
Integrase	*A. campanulatus*	0	591	602	98.17%	100%	97.22%
p15	*A. campanulatus*	5×10−67	231	240	96.25%	100%	95.83%
p6pol	*T. mangrovi*	9×10−20	92.8	99.8	92.98%	97%	95.74%
Nef	*B. velezensis*	9×10−108	325	389	83.54%	100%	73.79%
Rev	*A. campanulatus*	2×10−54	180	194	92.78%	100%	78.45%
Tat	*K. pneumoniae*	5×10−23	97.4	130	74.92%	83%	63.89%
Vif	*A. phoceense*	7×10−112	333	372	89.52%	98%	83.68%
Vpr	*A. phoceense*	2×10−58	193	193	100%	100%	96.88%
Vpu	*B. pseudomallei*	1×10−9	63.5	109	58.25%	65%	66.67%

[1] mBS is the maximum possible bit-score (BS) obtained from a self-hit alignment (details in Methods). [2] RBS, relative bit-score obtained dividing the BS by the corresponding mBS.

The genome database used in TBLASTN searches included not one but many genomes from the same bacterial species, usually different strains/isolates. Thereby, it is worth noting that HIV-1 proteins produced very similar hits in different representative genomes of a particular bacterial species (see Supplementary file 1). The number of hits per bacterial species with RBS ≥ 50% corresponding to HIV-1 proteins relevant for HIV-1 testing is shown in Figure 2 along with the RBS of the best hit. The selected proteins include gp120 and gp41, p24 and p17, p31 and p51.

201 The first two are part of the HIV-1 envelope glycoprotein (gp160 or Env) and derive by
202 cleavage, while p24 and p17 (matrix) derive from Gag protein, and p31 (integrase) and p51
203 (reverse transcriptase, RT) from Pol. The bacterial species whose genomes resulted in
204 significant hits varied with the distinct HIV-1 proteins (Fig.2). However, *E. coli* and *K.*
205 *pneumoniae* were the subject of the most significant hits with the majority of HIV-1 proteins.
206 Only p17 (matrix) did not yield any significant hit with these two bacterial species.

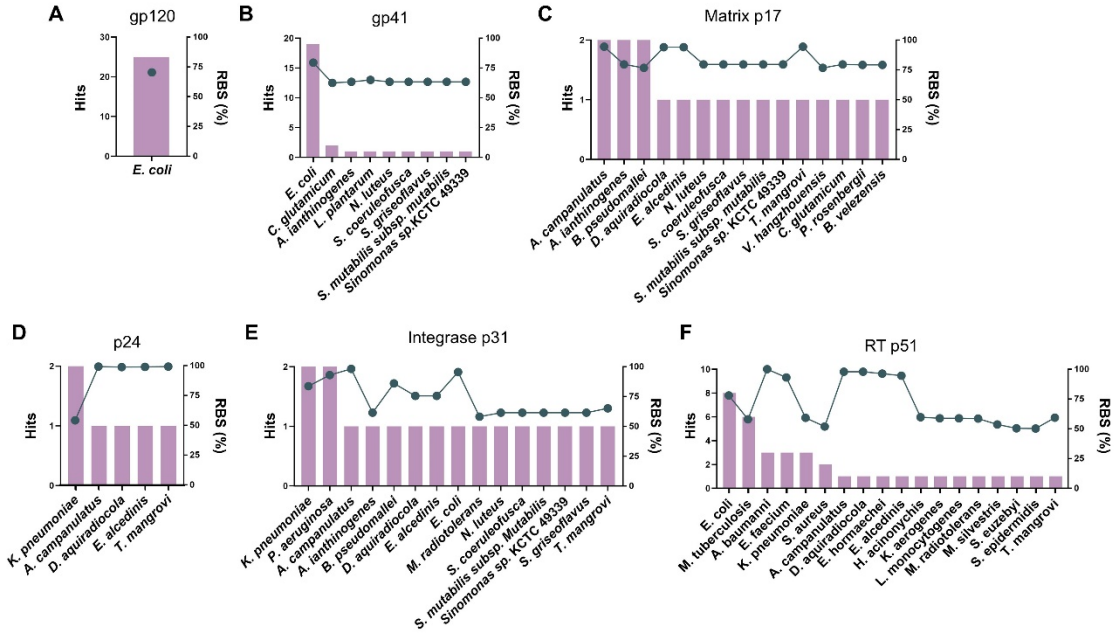

Figure 2. HIV-1 protein hits per bacterial species. The graphs show the number of hits on bacterial genomes (RBS ≥ 50%) of gp120 (**A**), gp41 (**B**), p17 (**C**), p24 (**D**), p51 (**E**), and p31 (**F**), stratified by bacterial species. The graphs also depict the RBS (line and dots) of the best hit per bacterial species.

207

208 **HIV-1 proteome coverage in bacterial genomes**

209 Since many HIV-1 proteins produced significant hits with the same bacterial species, we
210 introduced the viral proteome coverage (VPC) to estimate how much of the HIV-1 proteome,
211 not just a single protein, is present in the genome of a specific bacterial species. The VPC is
212 computed by adding up the RBS value from the best hit of each viral protein to the genome of a
213 specific bacterial species and dividing by the total number of viral proteins used in TBLASTN
214 searches (detailed in Methods). A VPC of 100% will indicate that all proteins included in the
215 viral proteome produce 100% identity hit alignments spanning their entire length. As shown in
216 Figure 3, HIV-1 reached the largest VPC in *K. pneumoniae* (67.74%), followed by *E. coli*

217 (63.72%), *Actinoplanes campanulatus* (60.12 %), *Enterococcus alcedinis* (59.91%) and
218 *Deinococcus aquiradiocola* (59.87%) However, while *A. campanulatus*, *E. alcedinis* and *D.*
219 *aquiradiocola* were the subject of only a few, yet highly significant hits*, K. pneumoniae and E.*
220 *coli* were the subject of the majority of significant hits, indicating that HIV-1 proteome appears
221 repeatedly in these two species and this presence is unlikely the result of contamination or
222 experimental artifacts. The data used to compute VPC values is provided in Supplementary
223 Table S1.

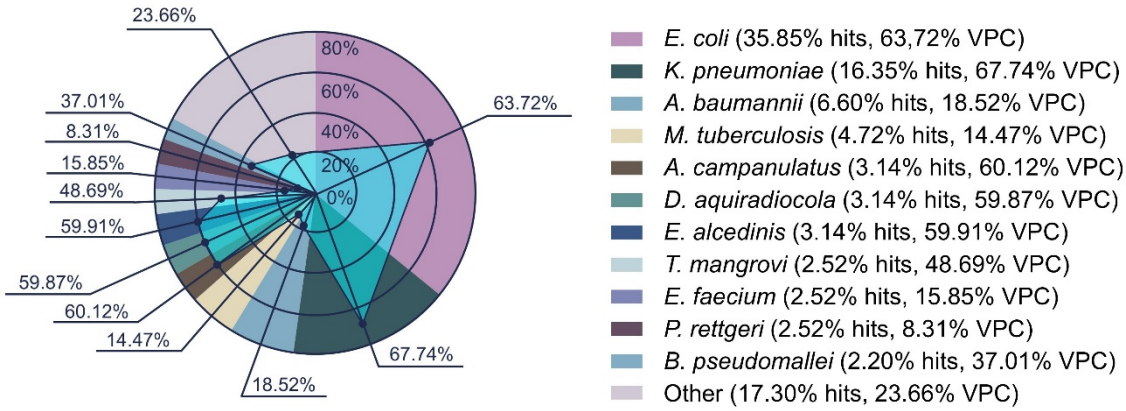

Figure 3. HIV-1 proteome in bacterial species. Parts-to-whole chart show the percentage of significant TBLASTN hits (RBS ≥ 50%) to each bacterial species considering all HIV-1 proteins. Total number of hits with ≥ 50% was 318. The spider charts represent the VPC per bacterial species computed after the best hits. The VPC shown in *Other* corresponds to the VPC determined on *Pseudomonas aeruginosa*, that reaches the largest VPC of all bacterial species included under *Other*.

224

225 **Proteome coverage of common viruses in bacterial genomes**

226

227 Given the extensive presence of HIV-1 in bacterial genomes, we examined whether the same
228 was true for other viruses. To that end, we carried out TBLASTN searches against bacterial
229 genomes with the proteome of common viruses and analyzed hits computing their RBS. The
230 VPC of the viruses per bacterial species was also computed after the RBS of the best hits. The
231 virus proteomes included Human Rhinovirus A (HRVA) and C (HRVC), Respiratory Syncitial
232 virus A (RSVA) and B (RSVB), Influenza A virus (IAV), Influenza B virus (IBV), Hepatitis B

233 virus (HBV), Hepatitis C virus (HCV) and Human Papiloma virus (HPV). The results revealed
234 that, all these viruses, with the exception of IAV, produced much fewer significant hits (RBS ≥
235 50%) in bacterial genomes than HIV-1 (Fig.4A). IAV produced a considerable number of
236 highly significant hits and reached a VPC of ≥ 75% in *Listeria monocytogenes* (Fig.4A). HBV
237 produced few but yet highly significant hits and actually reached a VPC of 100% in
238 *Streptomyces sp. 2BB-J2* (Fig.4). Supplementary Table S2 and S3 collect the statistics of the
239 best TBLASTN hits between IAV and HBV proteins with bacterial genomes, respectively.
240 Interestingly, we also noted that IAV produced a significant number of hits with *E. coli* reaching
241 a VPC ~60% (Fig.3B). Likewise, HBV reached a VPC of ~60% in *K. pneumoniae* (Fig.4C). In
242 supplementary Table S4 and S5, we provide the VPC data of IAV with *E. coli* and of HBV with
243 *K. pneumoniae*, respectively.

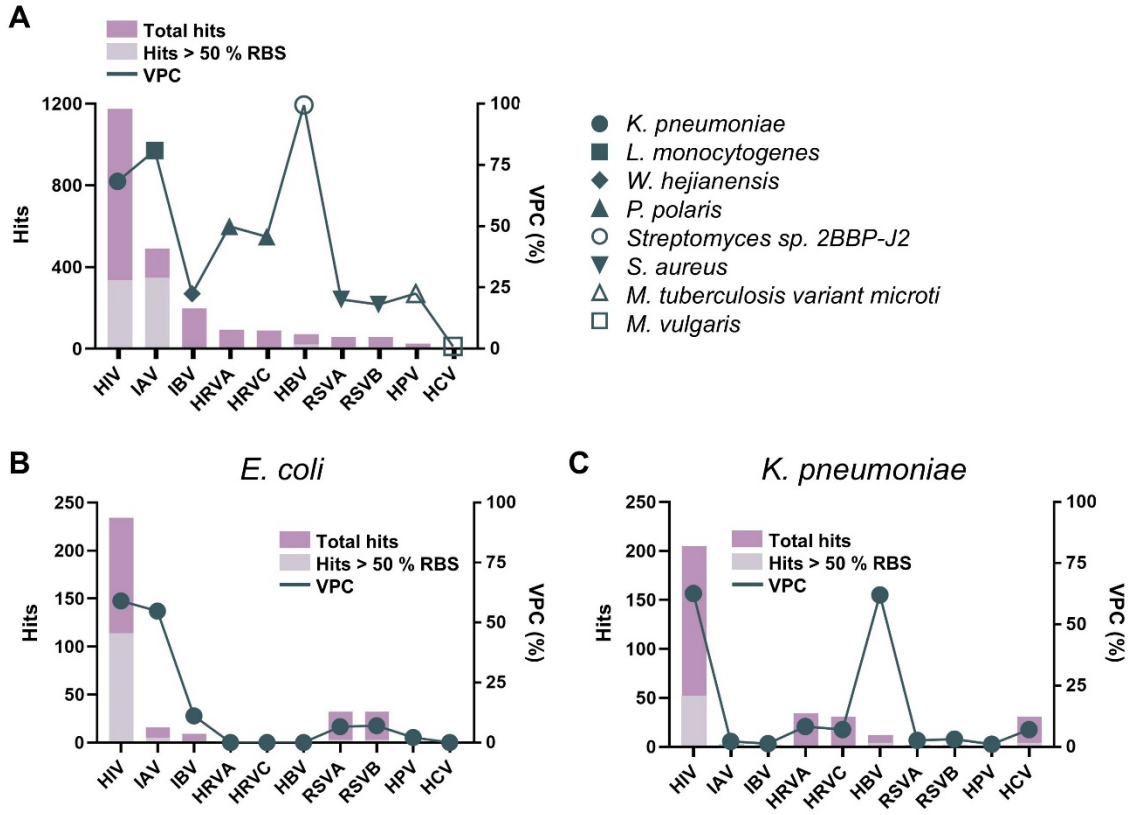

Fig.4. Presence of common virus proteomes in bacterial species. A) Chart showing for each virus (HIV-1, IAV IBV, HRVA, HRVC, RSVA, RSVB, HPV and HCV) (x-axis) the total number of TBLASTN hits with bacterial genomes (left y-axis). Stacked columns in light grey show the number of hits with RBS ≥ 50%. The largest VPC value (right y-axis) reached by the viruses is also plotted in the chart (right y-axis), using different symbols that represent the bacterial species (shown on the right) where that particular value was attained. **B**) and **C**) show the same data as in panel A but limited to TBLASTN hits with *E. coli* and *K. pneumoniae*,

respectively.

244

244 **4. DISCUSSION**

245

246 HIV testing is an important part in the fight against AIDS since early diagnosis and treatment
247 maximize the benefits of ART therapy (7,8). Thereby, major attention is paid to minimize false-
248 negatives in HIV tests results that could delay ART therapy. However, false-positive (FP)
249 results in HIV testing are not without consequences, including those derived from exposing
250 healthy individuals to the adverse effects of lifelong treatments, not to mention the significant
251 social and emotional impact on a person. Fortunately, FPs are considered rare, yet they are
252 known to occur. Therefore, it is important to investigate and know the circumstances under
253 which they may arise.

254

255 FP HIV results are more commonly reported in relation to infections that induce antibodies
256 cross-reacting with HIV antigens (15). These cross-reactive antibodies likely result from
257 antigens showing molecular mimicry with HIV-1 proteins, which can be readily detected by
258 sequence similarity (25-27). However, the infectious agents that can potentially induce cross-
259 reactive antibodies with HIV remain to be investigated and are unknown. In this work, we
260 investigated bacteria as potential sources of immune cross-reactivity to HIV-1 due to molecular
261 mimicry, using TBLASTN searches with HIV-1 proteins against bacterial genomes. The results
262 were surprising and unexpected.

263

264 We found that HIV-1 proteins appear to be present in quite a few bacterial species (Fig.1, Fig.2
265 and Fig.3). Proteins like p24, Vpr and Integrase were found to produce 100% sequence identity
266 hit alignments over their entire length of the sequences (relative bit-score, RBS, of 100 %), and
267 hit alignments produced by other HIV-1 proteins were also highly significant (RBS ≥ 60 %)
268 (Fig. 1). *K. pneumoniae* and *E. coli* were the subject of the best hits with HIV-1 proteins that
269 are relevant for HIV testing like gp120, gp41, p24, p31 and p51 (Fig. 2). *K. pneumoniae* and *E.*
270 *coli* also produced highly significant hits with the remaining HIV-1 proteins, reaching a VPC of
271 67.74% and 63.72%, respectively. Only three other bacterial species, *A. campanulatus, E.*

272 *alcedinis and D. aquiradiocola*, reached a comparable VPC of 60.12 %, 59.91%and 59.87%,
273 respectively.
274
275 Arguably, the detection of bacterial genomic sequences that are highly similar (if not identical)
276 to HIV-1 proteins or derived fragments could be attributed to contamination or some other
277 experimental artifact. However, the fact that the same/similar highly significant HIV-1 hits
278 were detected in different strains/isolates of the same bacterial species, such as in the case of *K.*
279 *pneumoniae* and *E. coli,* makes this possibility unlikely. Furthermore, the presence of HIV-1
280 related sequences in bacterial genomes might be greater if sequences that are not aligned with
281 the relevant taxa are excluded during annotation and curation processes. In comparison with
282 HIV-1, the proteome of other common viruses was seldom detected in bacterial genomes with
283 the exception of Influenza virus A (IAV) and Hepatitis B virus (HBV) (Fig. 4). Interestingly, *K.*
284 *pneumoniae* and *E. coli* also capitalized a relevant number of highly significant hits with IAV
285 and HBV but with differences. Thus, IAV reached a VPC of ~ 60% in *E. coli* and of 20% in *K.*
286 *pneumoniae.* In contrast, HPV reached a VPC of ~60% in *K. pneumoniae*, while in *E. coli* was ≤
287 20%. Overall, these data support that bacteria including *K. pneumoniae and E. coli* can acquire
288 viral genomic sequences with perhaps certain preferences. This possibility has actually been
289 verified experimentally for HIV-1. Thus, Zajac *et al.* (28) reported that DNA isolated from
290 intestinal bacteria of HIV/AIDS patients bears sequences that are more than 90% identical with
291 those of HIV-1. Of note, *E. coli* and *K. pneumoniae* are both part of the human gut microbiota,
292 where they are usually harmless (29,30). However, there are strains of *E. coli* that are
293 pathogenic (31) and *K. pneumoniae* can colonize other tissues and produce an ample range of
294 diseases (30). In general, bacteria in the gut microbiota can cause harm if they enter the
295 organism and are confronted by the immune system. Indeed, antibodies against the gut
296 microbiota are very common in human sera (32), which surely result after occasional breaches
297 in the gut mucosa leading to the entrance of bacteria and subsequent immune recognition.
298

299 Viruses, particularly retroviruses, are well known vectors of horizontal gene transfer, posing an
300 important evolutionary force in most life forms, including bacteria (33,34). However, how
301 bacteria can integrate sequences from viruses that infect humans is unclear –more so in the case
302 of RNA viruses– and deserves to be investigated. Nonetheless, human commensal bacteria,
303 including *E. coli,* are known to interact with enteric viruses (35) and likely with other viruses.
304 Hence, we can speculate that such close interactions at the gut, or sites of infection, may also
305 facilitate gene transfer from viruses to bacteria. Alternatively, viral gene transfer may occur
306 between bacteria and virus-infected cells (33). Integration of genomic material from RNA
307 viruses into bacterial genome will require prior reverse transcription to DNA. Incidentally, *E.*
308 *coli* has been found to express a functional reverse transcriptase (36) and likewise *K.*
309 *pneumoniae* (37), which is part of a viral defense mechanism (38).
310
311 Integration of HIV-1 sequences in *K. pneumoniae* and *E. coli* can have wide implications, since
312 those bacteria are part the gut microbiota and can be transmitted through person-to-person
313 contact or contact with contaminated sources. Bacteria carrying HIV-1 sequences are very
314 unlikely to shed infectious viral particles and contribute to HIV-1 transmission. However, *K.*
315 *pneumoniae* and *E. coli* specimens bearing HIV-1 sequences could serve as transmission vectors
316 of HIV positivity in HIV tests, provided that the bacteria can express HIV-1 proteins.
317 Interestingly, this appears to be the case. By using HIV-1 antibodies against various Gag
318 derived proteins, Hainova *et al.* (39) detected a protein in bacteria isolated from HIV-1 patients
319 with a size comparable to the Gag precursor. It could be argued that bacteria expressing HIV-1
320 proteins will unlikely secrete them, which may be anticipated as necessary to induce HIV-1
321 antibodies and transmit HIV positivity. However, once bacteria enter the body, there are
322 processes that destroy them and result in the release of hidden antigens and epitopes, which can
323 then induce specific antibodies. Indeed, a recent work has revealed that the majority of
324 antibodies elicited during infection are directed against hidden antigens and epitopes (40).
325 Relevant immune processes that would lead to particulate/degrade bacteria and antigens include
326 complement activation (41) and phagocytosis (42).

327 **5. CONCLUSION**

328

329 Our work supports that highly common and clinically relevant bacteria such as *E. coli* and *K.*
330 *pneumoniae* can acquire HIV-1 sequences and may transmit HIV positivity to people. Since
331 these bacteria are present in the gut microbiota, transmission of HIV positivity without actual
332 HIV infection may be common: it could occur through person-to-person contacts, from
333 contaminated sources and procedures such as microbiota transplantation, which demands testing
334 the presence of HIV sequences in microbiota samples. The possibility that bacteria such as *E.*
335 *coli* and *K. pneumoniae* could transmit HIV positivity deserves serious consideration to prevent
336 treating healthy individuals with life-long medications that carry significant toxicity, and take a
337 heavy toll on health care systems. Therefore, it is crucial to count with accurate HIV diagnostics
338 guaranteeing that life-saving ART is only received by HIV infected patients.

339

REFERENCES

1. Gallo, RC and Montagnier, L. The discovery of HIV as the cause of AIDS. *N Engl J Med. (2003)* **349**: 2283-2285. doi: 2210.1056/NEJMp038194.
2. Schnittman, SM and Fauci, AS. Human immunodeficiency virus and acquired immunodeficiency syndrome: an update. *Adv Intern Med. (1994)* **39**: 305-355.
3. Mody, A, Sohn, AH, Iwuji, C, Tan, RKJ, Venter, F and Geng, EH. HIV epidemiology, prevention, treatment, and implementation strategies for public health. *Lancet. (2024)* **403**: 471-492. doi: 410.1016/S0140-6736(1023)01381-01388. Epub 02023 Nov 01330.
4. Nyamweya, S, Hegedus, A, Jaye, A, Rowland-Jones, S, Flanagan, KL and Macallan, DC. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Rev Med Virol. (2013)* **23**: 221-240. doi: 210.1002/rmv.1739. Epub 2013 Feb 1026.
5. Parekh, BS, Ou, CY, Fonjungo, PN, Kalou, MB, Rottinghaus, E, Puren, A, Alexander, H, Hurlston Cox, M and Nkengasong, JN. Diagnosis of Human Immunodeficiency Virus Infection. *Clin Microbiol Rev. (2018)* **32**: e00064-00018. doi: 00010.01128/CMR.00064-00018. Print 02019 Jan.
6. Swinkels, HM, Justiz Vaillant, AA, Nguyen, AD and Gulick, PG. (2024). StatPearls Publishing, Treasure Island (FL).
7. Lundgren, JD, Babiker, AG, Sharma, S, Grund, B, Phillips, AN, Matthews, G, Kan, VL, Aagaard, B, Abo, I, Alston, B *et al.* Long-Term Benefits from Early Antiretroviral Therapy Initiation in HIV Infection. *NEJM Evid. (2023)* **2**: 10.1056/evidoa2200302. doi: 2200310.2201056/evidoa2200302. Epub 2202023 Feb 2200327.
8. Babiker, AG, Emery, S, F√§tkenheuer, G, Gordin, FM, Grund, B, Lundgren, JD, Neaton, JD, Pett, SL, Phillips, A, Touloumi, G *et al.* Considerations in the rationale, design and methods of the Strategic Timing of AntiRetroviral Treatment (START) study. *Clin Trials. (2013)* **10**: S5-S36. doi: 10.1177/1740774512440342. Epub 1740774512442012 Apr 1740774512440330.
9. Kemnic, TR and Gulick, PG. (2022) HIV Antiretroviral Therapy. https://www.ncbi.nlm.nih.gov/books/NBK513308/.
10. Chawla, A, Wang, C, Patton, C, Murray, M, Punekar, Y, de Ruiter, A and Steinhart, C. A Review of Long-Term Toxicity of Antiretroviral Treatment Regimens and Implications for an Aging Population. *Infect Dis Ther. (2018)* **7**: 183-195. doi: 110.1007/s40121-40018-40201-40126. Epub 42018 May 40114.
11. Bean, P. The history of HIV-1 testing. *Am Clin Lab. (2000)* **19**: 4.
12. Huynh, K and Kahwaji, CI. (2023) HIV Testing. *StatPearls*.
13. Reid, J, Van Zyl, G, Linstr√∂m, M, Korsman, S, Marais, G and Preiser, W. High positive HIV serology results can still be false positive. *IDCases. (2020)* **21:e00849.**: 10.1016/j.idcr.2020.e00849. eCollection 02020.
14. Colbrunn, DK, Jacks, C, Curry, SR, Gebregziabher, M and Meissner, EG. Outcomes of discordant HIV screening test results at a southern academic medical center. *Aids. (2024)* **38**: 1181-1185. doi: 1110.1097/QAD.0000000000003884. Epub 0000000000002024 Mar 0000000000003813.
15. Klarkowski, D, O'Brien, DP, Shanks, L and Singh, KP. Causes of false-positive HIV rapid diagnostic test results. *Expert Rev Anti Infect Ther. (2014)* **12**: 49-62. doi: 10.1586/14787210.14782014.14866516.
16. G√ºler, E, Ar*f*±kan, A, Abobakr, M, Sayan, M, S√ºer, K and ≈ûanl*f*±da*f*ü, T. Positive Anti-HIV ELISA Results in Pregnancy: Is It Reliable? *Infect Dis Obstet Gynecol. (2022)* **2022:1157793.**: 10.1155/2022/1157793. eCollection 1152022.
17. Simonsen, L, Buffington, J, Shapiro, CN, Holman, RC, Strine, TW, Grossman, BJ, Williams, AE and Schonberger, LB. Multiple false reactions in viral antibody screening assays after influenza vaccination. *Am J Epidemiol. (1995)* **141**: 1089-1096. doi: 1010.1093/oxfordjournals.aje.a117374.

18. Yu, KOA, Glassman, SR and Link, HM. False-Positive Human Immunodeficiency Virus-1 Test Results With Rapid Seroreversion After Third-Trimester Tdap Booster Vaccination. *Obstet Gynecol (2024)* **13**: 0000000000005726.
19. Quigley, J, Hussain, T and Arthur, C. False-positive HIV screening test in a healthcare student. *Occup Med (Lond). (2024)* **74**: 251-253. doi: 210.1093/occmed/kqae1015.
20. Banday, AZ, Nataraj, L, Jindal, AK, Kaur, H, Gummadi, A, Sharma, M, Pandiarajan, V and Rawat, A. False-positive HIV serology, Candida lusitaniae pneumonia, and a novel mutation in the CYBB gene. *Immunobiology. (2021)* **226**: 152110. doi: 152110.151016/j.imbio.152021.152110. Epub 152021 Jul 152112.
21. Everett, DB, Baisely, KJ, McNerney, R, Hambleton, I, Chirwa, T, Ross, DA, Changalucha, J, Watson-Jones, D, Helmby, H, Dunne, DW *et al.* Association of schistosomiasis with false-positive HIV test results in an African adolescent population. *J Clin Microbiol. (2010)* **48**: 1570-1577. doi: 1510.1128/JCM.02264-02209. Epub 02010 Feb 02224.
22. Ainoda, Y, Komaki-Yasuda, K, Kano, S and Hirai, Y. False-positive fourth-generation HIV test result in a woman with Plasmodium malariae malaria. *Trans R Soc Trop Med Hyg. (2023)* **117**: 147-148. doi: 110.1093/trstmh/trac1098.
23. Gudipati, S, Shallal, A, Peterson, E, Cook, B and Markowitz, N. Increase in False-Positive Fourth-Generation Human Immunodeficiency Virus Tests in Patients With Coronavirus Disease 2019. *Clin Infect Dis. (2023)* **77**: 615-619. doi: 610.1093/cid/ciad1264.
24. Ladizinski, B and Sankey, C. Acute Epstein-Barr virus infection and human immunodeficiency virus antibody cross-reactivity. *Am J Med. (2014)* **127**: e9-10. doi: 10.1016/j.amjmed.2014.1002.1039. Epub 2014 Mar 1016.
25. Bodas-Pinedo, A, Lafuente, EM, Pelaez-Prestel, HF, Ras-Carmona, A, Subiza, JL and Reche, PA. Combining different bacteria in vaccine formulations enhances the chance for antiviral cross-reactive immunity: a detailed in silico analysis for influenza A virus. *Front Immunol. (2023)* **14:1235053.**: 10.3389/fimmu.2023.1235053. eCollection 1232023.
26. Fernandez, SA, Pelaez-Prestel, HF, Fiyouzi, T, Gomez-Perosanz, M, Rein√©, J and Reche, PA. Tetanus-diphtheria vaccine can prime SARS-CoV-2 cross-reactive T cells. *Front Immunol. (2024)* **15:1425374.**: 10.3389/fimmu.2024.1425374. eCollection 1422024.
27. Reche, PA. Potential Cross-Reactive Immunity to SARS-CoV-2 From Common Human Pathogens and Vaccines. *Front Immunol. (2020)* **11:586984.**: 10.3389/fimmu.2020.586984. eCollection 582020.
28. Zajac, V, Stevurkova, V, Matelova, L and Ujhazy, E. Detection of HIV-1 sequences in intestinal bacteria of HIV/AIDS patients. *Neuro Endocrinol Lett. (2007)* **28**: 591-595.
29. Martinson, JNV and Walk, ST. Escherichia coli Residency in the Gut of Healthy Human Adults. *EcoSal Plus. (2020)* **9**: 10.1128/ecosalplus.ESP-0003-2020. doi: 1110.1128/ecosalplus.ESP-0003-2020.
30. Yang, J, Li, Y, Tang, N, Li, J, Zhou, J, Lu, S, Zhang, G, Song, Y, Wang, C, Zhong, J *et al.* The human gut serves as a reservoir of hypervirulent Klebsiella pneumoniae. *Gut Microbes. (2022)* **14**: 2114739. doi: 2114710.2111080/19490976.19492022.12114739.
31. Alhadlaq, MA, Aljurayyad, OI, Almansour, A, Al-Akeel, SI, Alzahrani, KO, Alsalman, SA, Yahya, R, Al-Hindi, RR, Hakami, MA, Alshahrani, SD *et al.* Overview of pathogenic Escherichia coli, with a focus on Shiga toxin-producing serotypes, global outbreaks (1982-2024) and food safety criteria. *Gut Pathog. (2024)* **16**: 57. doi: 10.1186/s13099-13024-00641-13099.
32. Vogl, T, Klompus, S, Leviatan, S, Kalka, IN, Weinberger, A, Wijmenga, C, Fu, J, Zhernakova, A, Weersma, RK and Segal, E. Population-wide diversity and stability of serum antibody epitope repertoires against human microbiota. *Nat Med. (2021)* **27**: 1442-1450. doi: 1410.1038/s41591-41021-01409-41593. Epub 42021 Jul 41519.

33. Moura de Sousa, J, Louren√ßo, M and Gordo, I. Horizontal gene transfer among host-associated microbes. *Cell Host Microbe. (2023)* **31**: 513-527. doi: 510.1016/j.chom.2023.1003.1017.
34. Arnold, BJ, Huang, IT and Hanage, WP. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol. (2022)* **20**: 206-218. doi: 210.1038/s41579-41021-00650-41574. Epub 42021 Nov 41512.
35. Karst, SM. The influence of commensal bacteria on infection with enteric viruses. *Nat Rev Microbiol. (2016)* **14**: 197-204. doi: 110.1038/nrmicro.2015.1025. Epub 2016 Feb 1038.
36. Lim, D and Maas, WK. Reverse transcriptase in bacteria. *Mol Microbiol. (1989)* **3**: 1141-1144. doi: 1110.1111/j.1365-2958.1989.tb00264.x.
37. Wilkinson, ME, Li, D, Gao, A, Macrae, RK and Zhang, F. Phage-triggered reverse transcription assembles a toxic repetitive gene from a noncoding RNA. *Science. (2024)* **386**: eadq3977. doi: 3910.1126/science.adq3977. Epub 2024 Oct 3974.
38. Tang, S, Conte, V, Zhang, DJ, ≈Ωedaveinyt*f*ó, R, Lampe, GD, Wiegand, T, Tang, LC, Wang, M, Walker, MWG, George, JT *et al.* De novo gene synthesis by an antiviral reverse transcriptase. *Science. (2024)* **386**: eadq0876. doi: 0810.1126/science.adq0876. Epub 2024 Oct 0874.
39. Hainova, K, Adamcikova, Z, Ciernikova, S, Stevurkova, V, Kr*f*çm√©ry, V and Zajac, V. Detection of protein homolougs with HIV-1 antigens in bacteria of positive patients - phase II. *Neuro Endocrinol Lett. (2014)* **35**: 110-115.
40. Ras-Carmona, A and Reche, PA. Analysis of Virus-Specific B Cell Epitopes Reveals Extensive Antigen Degradation Prior to Recognition. *Cells. (2024)* **13**: 1076. doi: 1010.3390/cells13131076.
41. Berends, ET, Kuipers, A, Ravesloot, MM, Urbanus, RT and Rooijakkers, SH. Bacteria under stress by complement and coagulation. *FEMS Microbiol Rev. (2014)* **38**: 1146-1171. doi: 1110.1111/1574-6976.12080. Epub 12014 Sep 12083.
42. Braun, V and Niedergang, F. Linking exocytosis and endocytosis during phagocytosis. *Biol Cell. (2006)* **98**: 195-201. doi: 110.1042/BC20050021.

SUPPLEMENTARY MATERIALS

Supplementary File 1: TBLASTN description summary of HIV-1 proteins against bacterial genomes.

Supplementary File 2: Supplementary Tables S1, S2, S3, S4 and S5.

Supplementary Table S1: HIV-1 proteome coverage in *E. coli* and *K. pneumoniae*.

Supplementary Table S2. IAV proteome coverage in *L. monocytogenes*.

Supplementary Table S3. HBV proteome coverage in *Streptomyces sp. 2BBP J2*.

Supplementary Table S4. IAV proteome coverage in *E. coli*.

Supplementary Table S5. HBV proteome coverage in *K. pneumoniae*

Author Contributions: Conceptualization: P.A.R; Methodology: H.F.P.-P. & J.M.-G.; Investigation: P.A.R & E.M.L; Data Analysis: H.F.P.-P. & J.M.-G; Writing-Original Draft: P.A.R; Final Writing & Editing: H.F.P.-P., J.M.-G, E.M.L & P.A.R. All authors have read and approved the final manuscript.

Funding: This research project was performed with no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Competing interest: The authors declare no competing interest

Author Approvals: All the individuals named as authors of this manuscript have approved its submission to bioRxiv