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 immune system, causing 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. HIV tests are regarded as reliable and very specific. However, false-positive results 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 proteins relevant for HIV testing, with a high degree of similarity (sequence identity > 95 %). The presence of HIV-1 in these bacteria of the human microbiota does not appear to be an artifact, since HIV-1 proteins were detected in different isolates of the same species. The proteome of other common viruses, particularly Influenza A virus and Hepatitis B virus, was also detected in bacterial genomes, but to a much lesser extent. Overall, our findings support that some bacteria can acquire HIV-1 genetic material and could interfere with HIV testing, causing false-positives.

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

31 **1. INTRODUCTION**

32

33 Human Immunodeficiency Virus (HIV) is an enveloped retrovirus that attacks $CD4^+$ T cells, leading
34 to their gradual depletion and the development of acquired immunodeficiency syndrome (AIDS)
35 (1,2). HIV transmission occurs mainly through sexual contact, often between men who have sex with
36 men but in Africa, where heterosexual sex transmission is dominant (3). The virus can be categorized
37 into 2 distinct subtypes: HIV type 1 (HIV-1) and HIV type 2 (HIV-2). Both viruses are transmitted
38 similarly, show the same tropism and cause AIDS (4). However, infections by HIV-1 are much more
39 common and progress quicker to AIDS than HIV-2 infections (4).

40

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

49

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

59

60 HIV diagnostics play a relevant role in the battle against HIV/AIDS, facilitating early initiation of
61 HIV treatments. Moreover, they are also useful to stage and monitor patients, and for outbreak
62 surveillance. HIV tests have historically relied on commercial enzyme immunoassays (EIAs) that
63 detect antibodies to HIV antigens in serum (5,11). Current generation of commercial HIV EIA tests
64 (4-th and 5-th generation), detects Abs against HIV-1 and HIV-2 as well as HIV-1 p24 antigen, which
65 allows for earlier diagnosis of HIV-1 infection (5). A list of FDA approved laboratory
66 Antibody/Antigen (Ab/Ag) HIV tests are available at
67 https://www.cdc.gov/hiv/partners/testing/laboratorytests.html (accessed on October/10/2024). In
68 these tests, HIV-1 p24 is detected by using Abs generated against this protein or fragments derived
69 from it, while HIV antibodies are detected by using HIV antigens and/or antigen fragment cocktails.
70 The following antigens are generally considered for HIV-1 testing (coding gene in italic): gp160
71 (envelope protein precursor, *env*), gp120 (extracellular region of envelope protein), gp41
72 (transmembrane region of envelope protein), p31 (integrase, *pol*), p51 (reverse transcriptase, *pol*), p24
73 (core protein, *gag*) and p17 (matrix protein, *gag*). In the laboratory, detection of Ab is sufficient to
74 confirm HIV infection but differentiation IEA tests may be necessary to distinguish between HIV-1
75 and HIV-2 infection. If the Ag test for HIV-1 p24 is positive but the Ab is negative, supplemental
76 confirmatory IEA tests (e.g. western-blotting), or RNA tests will be necessary to confirm or negate
77 HIV-1 infection (12). HIV RNA tests involve reverse transcription of HIV genetic material into DNA
78 and are aimed to detect conserved regions in *gag* or *pol* genes (5).

79

80 HIV tests are regarded as very reliable and specific. However, false-positives (FPs) can occur due to
81 technical reasons (e.g. misinterpretation and clerical errors) and more importantly under certain
82 biological or medical conditions (13-15). Relevant conditions that have been reported to interfere
83 with HIV IEA tests and produce FP results include: pregnancy (16), recent vaccination (e.g. flu, Tdap
84 or Hepatitis B vaccine) (17-19), cancer (19), and autoimmune diseases like lupus (20) or rheumatoid
85 arthritis (20). Likewise, various infectious diseases (e.g. schistosomiasis, malaria, COVID-19 and

86 Epstein-Barr virus mononucleosis) have been shown to produce FPs (21-24), likely due to cross-
87 reactivity.
88
89 In this work, we investigated potential cross-reactivity between bacteria and HIV-1, using TBLASTN
90 searches against bacterial genomes with the HIV-1 proteome. Interestingly, we detected the presence
91 of HIV-1 protein coding sequences in many bacterial species but more frequently in *Klebsiella*
92 *pneumoniae* and *Escherichia coli,* which are part of the human gut microbiota. Sequence identity
93 between HIV-1 and bacterial genomes was not only high (up to 100%) but reached the entire HIV-1
94 proteome, including p24 and other relevant proteins in HIV-1 testing. Other common viruses were
95 also detected in bacterial genomes, but to a much lesser degree. These results support that some
96 bacteria can acquire HIV-1 sequences, which could potentially induce immune cross-reactivity to
97 HIV-1, causing FP results in HIV tests. Ruling out this possibility will reduce the chance of putting
98 healthy individuals on ART and alleviate the economic burden of HIV treatments on health care
99 systems.
100

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

121 saved in Excel (Microsoft). The taxonomic TBLASTN information tabulated the bacterial species
122 with TBLASTN hits, indicating the number of hits and top bit-score of all hits *per* bacteria as well.
123
124 Similar TBLASTN searches were carried out to determine bit-scores of viral proteomes/proteins with
125 themselves but restricting the search in RefSeq Genome Database to Viruses (taxid:10239). As result,
126 for each protein a self bit-score resulting from a hit alignment with itself was obtained and recorded.
127 Bit-scores only depend on the hit sequence alignment, not the size of the database. Hence, the self bit-
128 score (sBS) of a viral sequence query corresponds to the maximum bit-score (mBS) that the query
129 sequence can reach.
130
131 **Relative bit-scores**
132
133 Relative bit-scores (RBSs) of TBLASTN hits between viral protein queries and subject bacterial
134 genomic sequences were obtained by dividing bit-scores (BS) with the corresponding mBS, as
135 indicated in Equation 1.
136

$$RBS = \frac{BS}{mBS} \times 100$$

137 Eq. 1
138
139 Where mBS is the maximum bit-score of the relevant viral protein sequence obtained from a self-hit
140 alignment, as defined previously. Since a given viral protein query sequence can produce numerous
141 TBLASTN hits to bacterial sequences from the same species, the top RBS for each viral protein per
142 bacterial species was computed after the best hit, that with the top bit-score.
143
144 **Viral proteome coverage**
145

146 The viral proteome coverage (VPC) was introduced as an estimation of the overall presence and
147 similarity of an entire viral proteome (*p*) in a given bacterial species (*b*). For any given viral proteome
148 used in TBLASTN searches, VPC (*p,b*) was obtained using Equation 2, which tallies the top RBS
149 (tRBS) from the best hit of each protein in the viral proteome (*p:1 .. N*) *per* bacterial species (b),
150 divided by the total number of proteins, N, in the viral proteome.
151

152 $$VPC_{(p,b)} = \frac{\sum_{p=N}^{p=1} tRBS_{(p,b)}}{N}$$ Eq.2

153

154 **Other procedures**

155

156 TBLASTN results were parsed using PERL scripts to compute RBSs of hit alignments and data was
157 processed in Excel (Microsoft, Corp.) Multiple sequence alignments between HIV-1 proteins and
158 bacterial TBLASTN hits were generated using T-COFFEE (25). All graphics were generated using
159 GraphPad Prism (GraphPad Software, Inc.).
160

161 **3. RESULTS**

162

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

164

165 We investigated bacteria as potential sources of immune cross-reactivity with HIV-1 by queering
166 bacterial genomes with HIV-1 protein sequences, using TBLASTN (details in Methods). The results
167 revealed that all HIV-1 proteins produced highly significant TBLASTN hits in the genome of various
168 bacterial species (TBLASTN output is provided in Supplementary File 1). The single most significant
169 hits with bacteria were produced by p24, Vpr, Integrase, Protease and Reverse Transcriptase (RT)
170 p51, which reached a relative bit-score (RBS) of > 97 % (Fig 1). The RBS of the hits between HIV-1
171 proteins and bacterial genomes were computed by dividing the bit-score of the hit alignments with the
172 maximum possible bit-score of the relevant HIV-1 proteins, obtained from a self-hit (details in
173 Methods). An RBS of 100% for a given HIV-1 protein indicates that the protein produces a hit
174 alignment over their entire length with 100% sequence identity. As shown in Figure 1A, all HIV-1
175 proteins produced a considerable number of hits with bacteria genomes with RBS ≥ 50 % and the
176 RBS of the best scoring hit was ≥ 58 % for all HIV-1 proteins. Interestingly, only a handful of
177 bacterial species, including *Escherichia coli* and *Klebsiella pneumoniae*, were the subject of the best
178 scoring hits with HIV-1 proteins (Fig.1). These best scoring hits between HIV-1 proteins and bacteria
179 had a sequence identity ≥ 63 % (value from Tat) and covered at least 65% of the length of query
180 (value from Vpu) (Table 2). Multiple sequence alignments of HIV-1 proteins with bacterial genomes
181 are provided in Supplementary File 2.

182

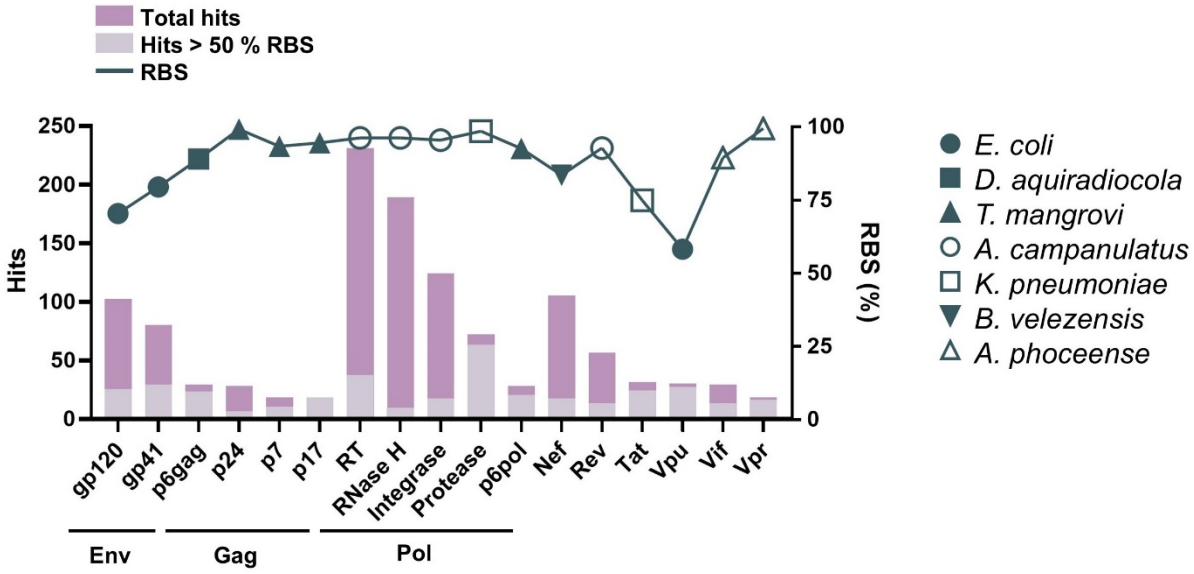

Figure 1. Presence of HIV-1 proteins in bacterial genomes. Chart showing for each HIV-1 protein (x-axis) the number of **TBLASTN** hits with bacterial genomes (left y-axis). Stacked columns in light grey show the number of hits with RBS ≥ 50%. The RBS (right y-axis) of the best hit for each HIV-1 protein 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 2. Statistics of best scoring TBLASTN hits between HIV-1 proteins and bacterial genomes

Viral protein	Bacteria	E-value	Bit-Score	mBS[1]	RBS[2]	Q-coverage	Identity
gp120	*Escherichia coli*	0	703	999	70.37%	99%	70.88%
gp41	*Escherichia coli*	4×10e−147	438	551	79.49%	92%	67.79%
p6gag	*Deinococcus aquiradiocola*	4×10e−20	94.4	106	89.06%	100%	90.38 %
p24	*Thalassovita mangrovi*	1×10e−159	478	482	99.17%	100%	98.27%
p7	*Thalassovita mangrovi*	2×10e−26	112	120	93.33%	100%	90.91%
p17 (Matrix)	*Thalassovita mangrovi*	3×10e−64	223	236	94.49%	100%	80.92%
RT (p51)	*Actinoplanes campanulatus*	0	1073	1096	97.90%	100%	94.29%
RNAse H (p15)	*Actinoplanes campanulatus*	5×10e−67	231	240	96.25%	100%	95.83%
Integrase (p31)	*Actinoplanes campanulatus*	0	591	602	98.17%	100%	97.22%
Protease	*Klebsiella pneumoniae*	2×10e−59	199	202	98.51%	100%	97.98%
p6pol	*Thalassovita mangrovi*	9×10e−20	92.8	99.8	92.98%	97%	95.74%
Nef	*Bacillus velezensis*	9×10e−108	325	389	83.54%	100%	73.79%
Rev	*Actinoplanes campanulatus*	2×10e−54	180	194	92.78%	100%	78.45%
Tat	*Klebsiella pneumoniae*	5×10e−23	97.4	130	74.92%	83%	63.89%
Vpu	*Escherichia coli*	7×10e−9	63.5	109	58.25%	65%	66.67%
Vif	*Aeromicrobium phoceense*	7×10e−112	333	372	89.24%	98%	83.68%
Vpr	*Aeromicrobium phoceense*	2×10e−58	193	196	98.46%	100%	96.88%

[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. Selected best TBLASTN hits were those with the largest BS.

The genome database used in TBLASTN searches included not one but many genomes from the same bacterial species, usually different strains/isolates. Therefore, 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. The first two are part of the HIV-1 envelope glycoprotein (gp160 or Env) and derive by cleavage, while p24 and p17 (matrix) derive from Gag protein, and p31 (integrase) and p51 (reverse transcriptase, RT) from Pol. The bacterial species whose genomes resulted in significant hits varied with the distinct HIV-1 proteins (Fig.2). However, *E. coli* and *K. pneumoniae* were the subject of the most significant hits with the majority of HIV-1 proteins. Only p17 (matrix) did not yield any significant hit with these two bacterial species.

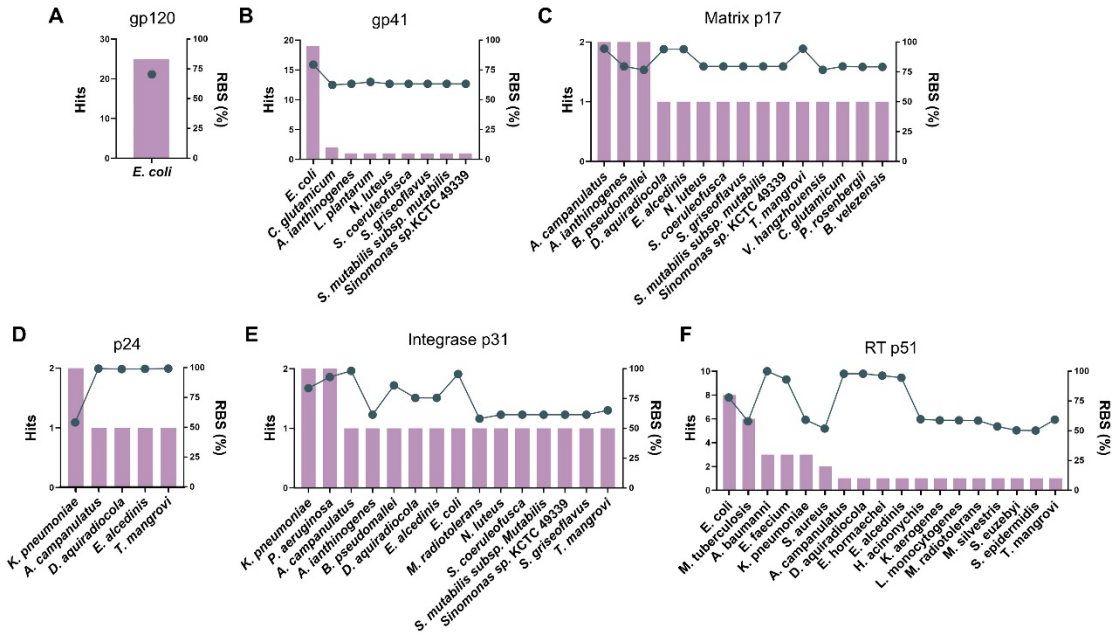

Figure 2. HIV-1 protein hits per bacterial species. The graphs show the number of hits on bacterial genomes (RBS ≥ 50%) of gp120 (**A**), gp41 (**B**), p17 (**C**), p24 (**D**), p51 (**E**), and p31 (**F**), stratified by bacterial species. The graphs also depict the RBS (line and dots) of the best scoring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

HIV-1 proteome coverage in bacterial genomes

Since many HIV-1 proteins produced significant hits with the same bacterial species, we introduced the viral proteome coverage (VPC) to estimate how much of the HIV-1 proteome, not just a single protein, is present in the genome of a specific bacterial species. The VPC was computed by adding up the RBS value from the best scoring hit of each viral protein to the genome of a specific bacterial species and dividing by the total number of viral proteins, 17 for HIV-1, used in TBLASTN searches (details in Methods). A VPC of 100% will indicate that all proteins included in the viral proteome

209 produce 100% identity hit alignments spanning their entire length. As shown in Figure 3, HIV-1
210 reached the largest VPC in *K. pneumoniae* (67.74%), followed by *E. coli* (63.72%), *Actinoplanes*
211 *campanulatus* (60.12 %), *Enterococcus alcedinis* (59.91%) and *Deinococcus aquiradiocola* (59.87%)
212 However, while *A. campanulatus*, *E. alcedinis* and *D. aquiradiocola* were the subject of only a few,
213 yet highly significant hits, *K. pneumoniae and E. coli* were the subject of the majority of significant
214 hits, indicating that HIV-1 proteome appears repeatedly in these two species and this presence is
215 unlikely the result of contamination or experimental artifacts. The data used to compute VPC values is
216 provided in Supplementary Table S1.

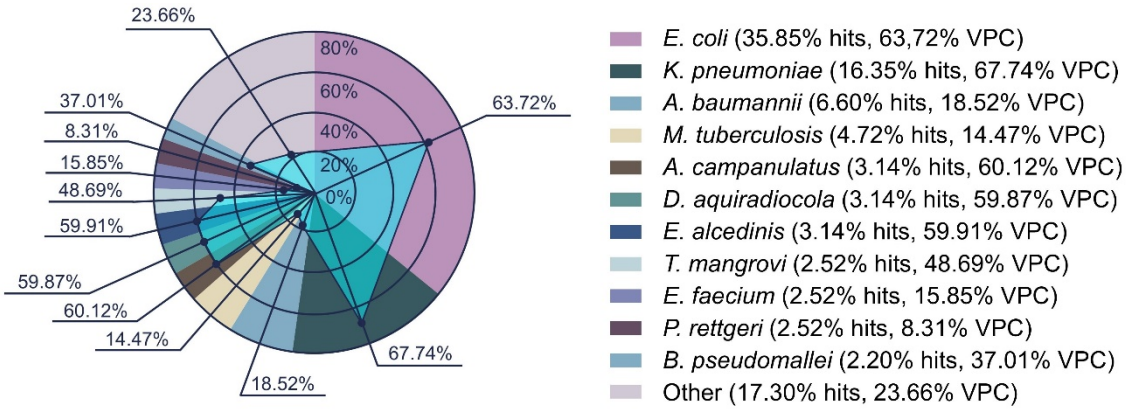

Figure 3. HIV-1 proteome in bacterial species. Parts-to-whole chart show the percentage of significant TBLASTN hits (RBS ≥ 50%) to each bacterial species considering all HIV-1 proteins. Total number of hits with ≥ 50% was 318. The spider charts represent the VPC per bacterial species computed after the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

217

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

219

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

226 virus (HCV) and Human Papiloma virus (HPV). The results revealed that, all these viruses, with the
227 exception of IAV, produced much fewer significant hits (RBS ≥ 50%) in bacterial genomes than HIV-
228 1 (Fig.4A). IAV produced a considerable number of highly significant hits and reached a VPC of ≥
229 75% in *Listeria monocytogenes* (Fig.4A). HBV produced few but yet highly significant hits and
230 actually reached a VPC of 100% in *Streptomyces sp. 2BB-J2* (Fig.4). Supplementary Table S2 and S3
231 collect the statistics of the best TBLASTN hits between IAV and HBV proteins with bacterial
232 genomes, respectively. Interestingly, we also noted that IAV produced a significant number of hits
233 with *E. coli* reaching a VPC ~60% (Fig.3B). Likewise, HBV reached a VPC of ~60% in *K.*
234 *pneumoniae* (Fig.4C). In supplementary Table S4 and S5, we provide the VPC data of IAV with *E.*
235 *coli* and of HBV with *K. pneumoniae*, respectively.

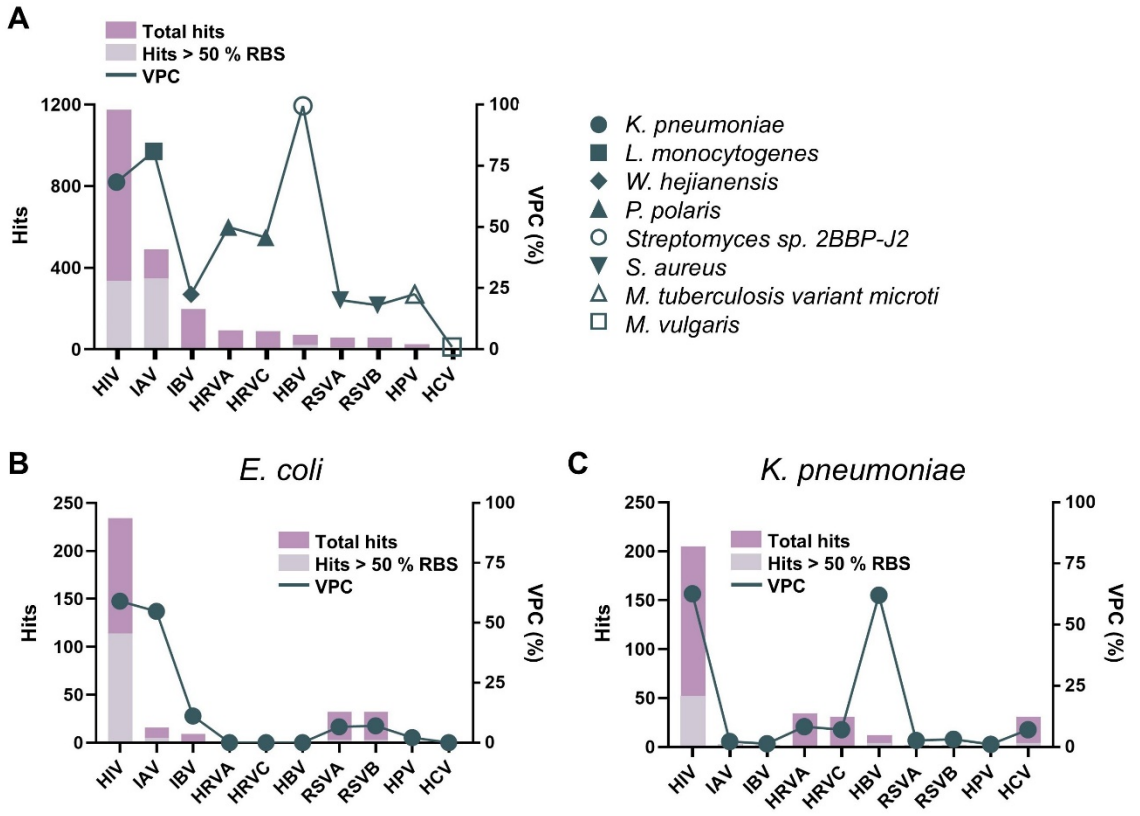

Fig.4. Presence of common virus proteomes in bacterial species. A) Chart showing for each virus (HIV-1, IAV IBV, HRVA, HRVC, RSVA, RSVB, HPV and HCV) (x-axis) the total number of TBLASTN hits with bacterial genomes (left y-axis). Stacked columns in light grey show the number of hits with RBS ≥ 50%. The 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.

236

236 **4. DISCUSSION**

237

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

245

246 FP HIV results are more commonly reported in relation to infections that induce antibodies cross-
247 reacting with HIV antigens (15). These cross-reactive antibodies likely result from antigens showing
248 molecular mimicry with HIV-1 proteins, which can be readily detected by sequence similarity (26-
249 28). However, the infectious agents that can potentially induce cross-reactive antibodies with HIV
250 remain to be investigated and are unknown. In this work, we investigated bacteria as potential sources
251 of immune cross-reactivity to HIV-1 due to molecular mimicry, using TBLASTN searches with HIV-
252 1 proteins against bacterial genomes. The results were surprising and unexpected.

253

254 We found that HIV-1 proteins appear to be present in quite a few bacterial species (Fig. 1, Fig. 2 and
255 Fig. 3). Proteins like p24, Integrase, Protease, R and Vpr and were found to produce hit alignments
256 with ≥ 96% sequence identity spanning their entire length (relative bit-score, RBS, > 97%). Moreover,
257 hit alignments produced by other HIV-1 proteins were also highly significant (RBS ≥ 58 %) (Fig. 1
258 and Table 1). *K. pneumoniae* and *E. coli* were the subject of the best hits with HIV-1 proteins that are
259 relevant for HIV testing like gp120, gp41, p24, p31 and p51 (Fig. 2). *K. pneumoniae* and *E. coli* also
260 produced highly significant hits with the remaining HIV-1 proteins, reaching a VPC of 67.74% and
261 63.72%, respectively. Only three other bacterial species, *A. campanulatus, E. alcedinis and D.*
262 *aquiradiocola*, reached a comparable VPC of 60.12 %, 59.91% and 59.87%, respectively.

263

264 Arguably, the detection of bacterial genomic sequences that are highly similar (if not identical) to
265 HIV-1 proteins or derived fragments could be attributed to contamination or some other experimental
266 artifact. However, the fact that the same/similar highly significant HIV-1 hits were detected in
267 different strains/isolates of the same bacterial species, such as in the case of *K. pneumoniae* and *E.*
268 *coli,* makes this possibility unlikely. Furthermore, the presence of HIV-1 related sequences in
269 bacterial genomes might be greater if sequences that are not aligned with the relevant taxa are
270 excluded during annotation and curation processes. In comparison with HIV-1, the proteome of other
271 common viruses was seldom detected in bacterial genomes with the exception of Influenza virus A
272 (IAV) and Hepatitis B virus (HBV) (Fig. 4). Interestingly, *K. pneumoniae* and *E. coli* also capitalized
273 a relevant number of highly significant hits with IAV and HBV but with differences. Thus, IAV
274 reached a VPC of ~ 60% in *E. coli* and of 20% in *K. pneumoniae.* In contrast, HPV reached a VPC of
275 ~60% in *K. pneumoniae*, while in *E. coli* was ≤ 20%. Overall, these data support that some bacteria
276 (*e.g. K. pneumoniae and E. coli*) can acquire viral genomic sequences, with perhaps certain
277 preferences. The presence of HIV-1 like sequences have been indeed confirmed in bacteria isolated
278 from the gut and respiratory tract of HIV/AIDS patients (29,30). Of note, *E. coli* and *K. pneumoniae*
279 are both part of the human gut microbiota, where they are usually harmless (31,32). However, there
280 are strains of *E. coli* that are pathogenic (33) and *K. pneumoniae* can colonize other tissues and
281 produce an ample range of diseases. Moreover, bacteria in the gut microbiota can cause harm if they
282 enter the organism, in which case the immune system responds and eliminates them. Indeed,
283 antibodies against the gut microbiota are very common in human sera (34), which surely result after
284 occasional breaches in the gut mucosa leading to the entrance of bacteria and subsequent immune
285 recognition.
286
287 Viruses, particularly retroviruses, are well known vectors of horizontal gene transfer, posing an
288 important evolutionary force in most life forms, including bacteria (35,36). However, how bacteria
289 can acquire sequences from viruses that infect humans is unclear –more so in the case of RNA
290 viruses– and deserves to be investigated. Nonetheless, human commensal bacteria, including *E. coli,*
291 are known to interact with enteric viruses (37) and likely with other viruses. Hence, we can speculate

292 that such close interactions at the gut, or sites of infection, may also facilitate gene transfer from
293 viruses to bacteria. Alternatively, viral gene transfer may occur between bacteria and virus-infected
294 cells (35). Integration of genomic material from RNA viruses into bacterial genome will require prior
295 reverse transcription to DNA. Incidentally, *E. coli* has been found to express a functional reverse
296 transcriptase (38) and likewise *K. pneumoniae* (39), which is part of a viral defense mechanism (40).
297
298 The presence of HIV-1 sequences in *K. pneumoniae* and *E. coli* can have wide implications, since
299 those bacteria are part the gut microbiota and can be transmitted through person-to-person contact or
300 contact with contaminated sources. Bacteria carrying HIV-1 sequences are very unlikely to shed
301 infectious viral particles and contribute to HIV-1 transmission. However, *K. pneumoniae* and *E. coli*
302 specimens bearing HIV-1 sequences could serve as transmission vectors of HIV seropositivity,
303 provided that the bacteria can express HIV-1 proteins. Interestingly, this appears to be the case. By
304 using HIV-1 antibodies against various Gag derived proteins, Hainova *et al.* (41) detected a protein in
305 bacteria isolated from HIV-1 patients with a size comparable to the Gag precursor. It could be argued
306 that bacteria expressing HIV-1 proteins will unlikely secrete them, which may be anticipated as
307 necessary to induce HIV-1 antibodies and transmit HIV seropositivity. However, once bacteria enter
308 the body, there are processes that destroy them and result in the release of hidden antigens and
309 epitopes, which can then induce specific antibodies. Indeed, a recent work has revealed that the
310 majority of antibodies elicited during infection are directed against hidden antigens and epitopes (42).
311 Relevant immune processes that would lead to particulate/degrade bacteria and antigens include
312 complement activation (43) and phagocytosis (44).
313
314

314 **5. CONCLUSION**
315

316 Our work supports that highly common and clinically relevant bacteria such as *E. coli* and *K.*
317 *pneumoniae* can acquire HIV-1 sequences and could transmit HIV seropositivity to people. Since
318 these bacteria are present in the gut microbiota, transmission of HIV positivity without actual HIV
319 infection might not be uncommon: it could occur through person-to-person contacts, from
320 contaminated sources and procedures such as microbiota transplantation, which demands testing the
321 presence of HIV sequences in microbiota samples. The possibility that microbiota bacteria could
322 transmit HIV seropositivity deserves serious consideration in order to prevent treating healthy
323 individuals with life-long medications that carry significant toxicity and take a heavy toll on health
324 care systems. Therefore, it is crucial to count with accurate HIV diagnostics guaranteeing that life-
325 saving ART is only received by HIV infected patients.
326

ACKNOWLEDGMENTS

We wish to thank to PAR lab members for comments and critical reading

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.

STATEMENTS

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.

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. Notredame, C, Higgins, DG and Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol (2000)* **302**: 205-217.
26. 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.
27. 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.
28. 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.
29. Zajac, V, Matelova, L, Liskova, A, Mego, M, Holec, V, Adamcikova, Z, Stevurkova, V, Shahum, A and Krcmery, V. Confirmation of HIV-like sequences in respiratory tract bacteria of Cambodian and Kenyan HIV-positive pediatric patients. *Med Sci Monit. (2011)* **17**: CR154-158. doi: 110.12659/msm.881449.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.

35. 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.
36. 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.
37. 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.
38. Lim, D and Maas, WK. Reverse transcriptase in bacteria. *Mol Microbiol. (1989)* **3**: 1141-1144. doi: 1110.1111/j.1365-2958.1989.tb00264.x.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. Braun, V and Niedergang, F. Linking exocytosis and endocytosis during phagocytosis. *Biol Cell. (2006)* **98**: 195-201. doi: 110.1042/BC20050021.

484 **SUPPLEMENTARY MATERIALS**

485

486 **Supplementary File 1:** Output of TBLASTN searches with HIV-1 proteins against bacterial genomes

487 **Supplementary File 2**: Supplementary File 2: Multiple Sequence Alignments of HIV-1 proteins and

488 top TBLASTN hits to bacterial genomes

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

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

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

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

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

494

495