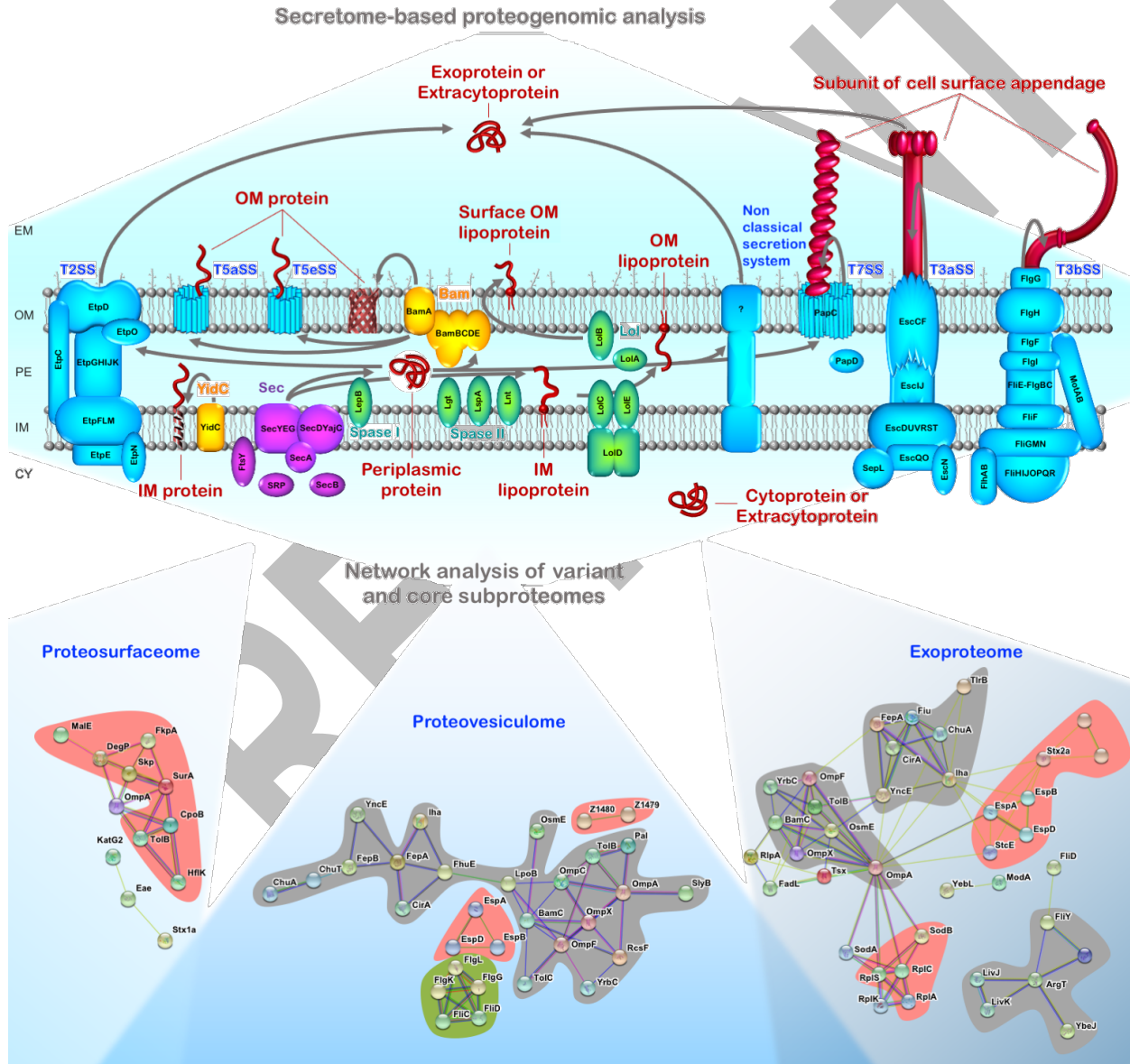


1 **The Secretome Landscape of *Escherichia coli* O157:H7: Deciphering the Cell-Surface, Outer**
 2 **Membrane Vesicle and Extracellular Subproteomes**

3 Ricardo MONTEIRO, Ingrid CHAFSEY, Valentin AGEORGES, Sabine LEROY, Christophe
 4 CHAMBON, Michel HÉBRAUD, Valérie LIVRELLI, Mariagrazia PIZZA, Alfredo PEZZICOLI,
 5 Mickaël DESVAUX*

6 **GRAPHICAL ABSTRACT**

Deciphering the extracytoplasmic subproteomes in *E. coli* O157:H7



7

8 **The Secretome Landscape of *Escherichia coli* O157:H7:**
9 **Deciphering the Cell-Surface, Outer Membrane Vesicle and**
10 **Extracellular Subproteomes**

11 Ricardo MONTEIRO ^{1,2}, Ingrid CHAFSEY ¹, Valentin AGEORGES ¹, Sabine LEROY ¹,
12 Christophe CHAMBON ³, Michel HÉBRAUD ^{1,3}, Valérie LIVRELLI ⁴, Mariagrazia PIZZA ²,
13 Alfredo PEZZICOLI ², Mickaël DESVAUX ^{1,*}

14 ¹ Université Clermont Auvergne, INRAE, MEDiS, F-63000 Clermont-Ferrand, France.

15 ² GSK, Siena, Italy.

16 ³ INRAE, Plate-Forme d'Exploration du Métabolisme, QuaPA, F-63122 Saint-Genès
17 Champanelle, France.

18 ⁴ Université Clermont-Auvergne, INSERM, INRAE, M2ISH, F-63000 Clermont-ferrand, France.

19 *Corresponding author: Dr Hab Mickaël DESVAUX-LENÔTRE, INRAE (Institut National de
20 Recherche pour l'Agriculture, l'Alimentation et l'Environnement) Auvergne-Rhône-Alpes,
21 Centre de Recherche de Clermont-Ferrand, UMR454 MEDiS (Microbiologie, Environnement
22 Digestif, Santé), Site de Theix, F-63122 Saint-Genès Champanelle, France. Tel.: +33 (0)4 73 62
23 47 23, Fax: +33 (0)4 73 62 45 81, E-mail: mickael.desvaux@inrae.fr.

24 **Abstract**

25 Among diarrheagenic *E. coli* (DEC), enterohaemorrhagic *E. coli* (EHEC) are the most virulent
26 anthropozoonotic agents. The ability of bacterial cells to functionally interact with their
27 surrounding essentially relies on the secretion of different protein effectors. To experimentally
28 determine the repertoire of extracytoproteins in *E. coli* O157:H7, a subproteomic analysis was
29 performed not only considering the extracellular milieu but the cell surface and outer membrane
30 vesicles. Following a secretome-based approach, the proteins trafficking from the interior to the
31 exterior of the cell were depicted considering cognate protein transport systems and subcellular
32 localisation. Label-free quantitative analysis of the proteosurfaceome, proteovesiculome and
33 exoproteome from *E. coli* O157:H7 grown in three different nutrient media revealed differential
34 protein expression profiles and allowed defining the core and variant subproteomes. Network
35 analysis further revealed the higher abundance of some protein clusters in chemically defined
36 medium over rich complex medium, especially related to some outer membrane proteins, ABC
37 transport and Type III secretion systems. This first comprehensive study of the EHEC secretome
38 unravels the profound influence of environmental conditions on the extracytoplasmic proteome,
39 provides new insight in the physiology of *E. coli* O157:H7 and identifies potentially important
40 molecular targets for the development of preventive strategies against EHEC/STEC.

41 **Keywords:** Shiga toxin-producing *E. coli* (STEC); Subproteomic analysis; Secretome; Outer
42 membrane vesicle; Cell-surface proteins; Extracellular proteins

43 **Introduction**

44 *Escherichia coli* are an inhabitant of warm-blooded animals intestines and typically coexist with
45 the host as a commensal with mutual benefit for both organisms [1]. However, some highly
46 adapted *E. coli* strains have acquired specific virulence attributes, which allow them causing a
47 broad spectrum of diseases. Based on clinical syndromes, these pathogenic strains can be
48 discriminated into extraintestinal pathogenic *E. coli* (ExPEC) responsible for urinary tract
49 infections, sepsis and/or meningitis [2] and intestinal pathogenic *E. coli* (InPEC), especially the
50 diarrheagenic *E. coli* (DEC) subset. DEC represent a leading cause for traveller's diarrhoea but
51 also pediatric disease in developing countries, and an emerging source of diarrhoea in
52 industrialised countries [3]. Enterohaemorrhagic *E. coli* (EHEC), which belong to the group of
53 shigatoxin encoding *E. coli* (STEC), are the most severe etiologic agent for *E. coli* diarrhoea in
54 humans, with a wide spectrum of diseases ranging from aqueous or bloody diarrhoea to severe
55 cases of haemorrhagic colitis to haemolytic uremic syndrome (HUS), the leading cause of chronic
56 renal failure in children [4, 5]. Over the 400 distinct serotypes of STEC reported to date [6], only
57 a handful of serogroups are reported in human outbreaks [7], included EHEC O157:H7 considered
58 as one of the most severe [8].

59 While shigatoxin (Stx) is regarded as the main virulence factor for EHEC, it is neither
60 essential in DEC nor sufficient in EHEC. Actually, the ability of EHEC to colonise the intestinal
61 epithelial surface would also be a major factor in tuning the virulence level and thus the severity
62 of the infection, *e.g.* the attaching and effacing lesions (A/E lesions) [9, 10]. Most of the genes
63 necessary to form A/E lesions are encoded by a 35-kb chromosomal pathogenicity island (PAI)
64 known as the locus for enterocyte effacement (LEE), particularly genes encoding a Type III,
65 subtype a, secretion system (T3aSS) and the cognate effector proteins [11, 12]. Besides infection,

66 some virulence factors could play a role as niche factors in various environmental conditions
67 encounters by bacterial cells along the food chain [13, 14], from soil, plants, animal carriage, food
68 matrices to human gastrointestinal tract when considering EHEC [15, 16]. While the main secreted
69 virulence factors are now well characterised in EHEC O157:H7 [17-19], additional genes are most
70 probably induced in the course of an infection [20] but their expression at a protein level and in
71 other environmental conditions remain elusive. In lipopolysaccharidic (LPS) diderm bacteria (*i.e.*
72 archetypal Gram-negative bacteria) [21], nine protein secretion systems, from Type I to Type IX
73 (T1SS-T9SS) have been uncovered to date but, surprisingly enough, some proteins can be found
74 as localised extracellularly without any obvious cognate secretion pathways. Such transport *via*
75 unknown protein trafficking route is referred as non-classical secretion (NCS) [22, 23], where
76 different molecular mechanisms can be at play, *e.g.* controlled lysis, piggybacking or outer
77 membrane vesicles (OMVs) [24-26]. As secretion refers to the active transport from the interior to
78 the exterior of the cell compartment [21], secreted proteins can have four main possible fates in
79 LPS-diderm bacteria, *i.e.* to (i) remain associated with the bacterial outer membrane (OM), (ii) be
80 released into the extracellular milieu, (iii) be vectorised in OMVs, or (iv) be injected into a target
81 cell (either a eukaryotic or bacterial cell) [27, 28]. Environmental growth conditions are known to
82 influence the secretion of some classically and non-classically secreted proteins in *E. coli* O157:H7
83 [29], however, the occurrence at a global secretome scale remains to be explored. Considering the
84 poor correlation between mRNA and protein abundance, which is even more exacerbated when
85 dealing with secreted proteins [30-32], proteomics is a pre-request to functional analyses.

86 Growth conditions are known to deeply influence the physiology of *E. coli* O157:H7, for
87 instance in the ability to adhere to different extracellular matrix components [33], the adherence
88 to intestinal epithelial cells [34, 35] or the virulence with the LEE expression [36, 37]. However,

89 there is still a gap of knowledge in apprehending the whole proteins secreted by *E. coli* O157:H7.
90 For some times, the investigations of the extracellular proteome (exoproteome) were often
91 restricted to the use of chemically defined medium due to the limit imposed by two-dimensional
92 electrophoresis (2-DE) approaches, especially with the isoelectric focusing (IEF) in the first
93 dimension, and the access to the cell-surface proteome (proteosurfaceome) proven difficult
94 especially due to the hydrophobicity of outer membrane proteins (OMPs). Nowadays, the
95 development of shotgun off-gel proteomics has clearly exceeded in-gel 2-DE approaches to
96 circumvent these limitations, together with cell-surface biotin labelling to efficiently investigate
97 the proteosurfaceome [38, 39]. The present study aims at deciphering the secretome landscape of
98 *E. coli* O157:H7 by using a range of growth media most likely or known to induce changes in the
99 protein secretion profiles, namely (i) M9 as a minimal medium forcing the bacterial cell to activate
100 its full anabolic potential, (ii) DMEM (Dulbecco's modified Eagle's medium) as a chemically
101 defined medium known to induce the expression of several virulence factors, and (iii) BHI (brain
102 heart infusion) as a rich complex medium classically use for the growth of microbial pathogens.
103 To get a whole picture of proteins secreted by *E. coli* O157:H7, three subproteomes were
104 considered and characterised by relative quantification analysis, namely the proteosurfaceome
105 (proteome of the cell surface), exoproteome (proteome of the extracellular milieu) and
106 proteovesiculome (proteome of the membrane vesicles).

107 **Material and Methods**

108 **Bacterial growth conditions**

109 *Escherichia coli* O157:H7 strain EDL933 was used for this study [40]. Pre-cultures were grown
110 aerobically in M9, DMEM (Dulbecco's modified eagle medium, Gibco), or BHI (brain-heart
111 infusion, Becton-Dickinson) liquid media at 37°C under orbital shaking (150 rpm) in triplicates.
112 After 1:100 (v/v) dilution, bacterial cultures were grown in the same conditions until late
113 exponential phase and sampled at about 3, 5 and 8 h from BHI, DMEM and M9, respectively.
114 Samples corresponded to the same OD_{600 nm} (0.5) and bacterial cells were harvested by
115 centrifugation (4 000 g, 5 min, 4°C) and the culture supernatant recovered.

116 **Isolation of bacterial cell-surface, outer membrane vesicle, and extracellular** 117 **proteins**

118 Bacterial cell-surface proteins were isolated following biotinylation and protein affinity
119 purification as previously described in details [39] using sulfo-NHS-SS-biotin (sulfo-succinimidyl
120 biotin-amidoethyl dithio-propionate) for protein labelling and neutravidin for purification. Briefly,
121 bacterial cells were washed with 10 mM PBS pH 8.0 containing 1 mM phenylmethylsulfonyl
122 fluoride (PMSF) and centrifuged as above. Pelleted cells were weighed to ensure the same amount
123 of wet cells (100 mg) was labelled with sulfo-NHS-SS-biotin (1.5 mM) for 1 min at room
124 temperature. After quenching the excess of biotinylation reagent with glycine buffer (500 mM)
125 and centrifugation, cells were resuspended in Triton X-100 lysis buffer (1 %, 1 mM PMSF, 10 mM
126 PBS pH 8). Cells were disrupted using FastPrep-24 classic (MP Biomedicals) with two steps of
127 20 s at 6 m/s, cell debris discarded by centrifugation (20 000 g, 30 min, 4°C) and exactly the same
128 volume of each biotinylated protein samples (500 µl) was used for affinity purification using
129 neutravidin agarose resin (Pierce Thermo Scientific, 400 µl). After washing (10 mM PBS, pH 8,

130 1 % NP-40), labelled proteins were eluted (2 % SDS, 20 % glycerol, 5 % β -mercaptoethanol,
131 50 mM dithiothreitol, 62.5 mM Tris-HCl).

132 Recovery of OMV and extracellular proteins were adapted from previously described protocols
133 [41]. Briefly, the same volume of culture supernatant corresponding to the same amount of
134 bacterial cells was first filtered through a 0.45 μ m membrane under vacuum (to eliminate residual
135 cells), and ultrafiltrated on a polyethersulfone (PES) membrane at a 500 kDa cut-off (to
136 concentrate the supernatant); the ultrafiltrate was kept and corresponded to the extracellular
137 fraction, whereas the retentate was further processed to get the OMV fraction. In the ultrafiltrate,
138 0.2 mM PMSF was added to inhibit protease activity as well as sodium deoxycholate
139 (0.2 mg.mL⁻¹) to support protein precipitation with trichloroacetic acid (10 % w/v). After
140 centrifugation (20 000 g, 30 min, 4°C), the precipitate was washed with ice-cold acetone and all
141 samples were solubilised in the same volume of PBS. The retentate was further ultracentrifuged
142 (100 000 g, 3 h, 4°C) and all samples were resuspended in the same volume of PBS. Protein
143 concentration was determined by the method of Bradford using bovine serum albumin as the
144 standard [34].

145 **Protein preparation for LC-MS/MS analysis**

146 To avoid bias in the comparison of protein abundance by relative quantification analyses, the
147 bacterial cell surface, extracellular milieu and OMV fractions, respectively, were adjusted
148 respective to the same amount of biomass in the different growth conditions as indicated above.

149 To eliminate interfering molecules (such as salts) and concentrate proteins in a single band of
150 polyacrylamide gel, each sample was submitted to a short SDS-PAGE (sodium dodecyl sulfate–
151 polyacrylamide gel electrophoresis) migration (T=12.5 %, C= 3.3%) in a BioRad Mini Protean II
152 unit. After the addition of the same volume of Laemmli buffer (2 % SDS, 25 % glycerol, 5 % β -

153 mercaptoethanol, 0.005 % bromphenol blue and 62.5 mM Tris-HCl, final) and heating at 95°C for
154 5 min, 15 µl of protein extracts from each subcellular fraction were migrated and just allowed to
155 concentrate at the junction between stacking and resolving gel after a few minutes migration
156 (25 mA/gel at 100 V). Gel was then stained with Coomassie Brilliant Blue G250 and the single
157 bands in each lane were excised. The proteins in the gel were reduced in 100 mM ammonium
158 bicarbonate with 45 mM dithiothreitol for 45 min at 50°C, and alkylated with 100 mM
159 iodoacetamide for 20 min at room temperature in the dark. The gel was washed once in 25 mM
160 ammonium bicarbonate in 5 % acetonitrile for 30 min and twice in 25 mM ammonium bicarbonate
161 in 50 % acetonitrile for 30 min each. The gel was dehydrated with 100 % acetonitrile, then the
162 samples were hydrolysed with 400 ng of trypsin (Promega) for 5 h and the peptides extracted with
163 100 % acetonitrile. After air drying, the hydrolysed samples were resuspended in the same volume
164 of a 0.1 % trifluoroacetic acid aqueous solution before injection (3 µl) and nanoHPLC separation
165 (Ultimate 3000, Dionex) with a C18 column (LC Acclaim PepMap, ThermoScientific, 75 µm,
166 15 cm) with acetonitrile solution gradient (99.9 % acetonitrile, 0.1 % water) from 10 to 40 %
167 during 50 min (as validated for samples with at least less than 500 different proteins). The peptides
168 separated on the nanoHPLC column were nanoelectrosprayed *via* a nanoESI source in the mass
169 spectrometer (LTQ Velos IonTrapOrbitrap, ThermoScientific) used in data dependant mode with
170 full MS in the range 400-1 400 m/z with a resolution set up at 30 000 in the orbitrap. For MS/MS,
171 top ten mode was used (full scan MS followed by ten MS/MS analysis on major peptides in the
172 full scan) with 2 repeat counts, dynamic exclusion set up at 80 s, and CID in ion trap set up at 37 %
173 in collision energy. Each condition was performed in triplicates (biological replicates) with two
174 runs per sample (technical repeats).

175 **Data processing and bioinformatic analysis**

176 The proteins identification was performed from *E. coli* O157:H7 strain EDL933 database
177 (KEGG Genome T00044, 5449 protein entries) using Mascot search engine. Peptides were
178 validated for a Mascot score permitting to obtain a false discovery rate (FDR) at below 1 % and
179 proteins were identified when at least two peptides matched significantly (score higher than 20) in
180 the database with search parameters set to 10 ppm for peptide and 0.5 Da for fragment mass
181 tolerance. Comparison of protein abundance was performed following the standard workflow
182 provided by Progenesis QI for proteomics applying the label-free relative quantification method
183 using Hi-N mode (Nonlinear Dynamics, Waters); basically, the MS1 survey scans are used for ion
184 abundance quantification where peptide ions falling within the robust estimation limits are used to
185 calculate the normalisation factor and selected as normalisation reference to compare peptidic
186 signals between runs. For each subproteome, differential protein quantification was considered
187 significant for fold abundance greater than 1.5 and *p*-value lower than 0.05. FDR (false discovery
188 rate) approach was applied to overcome the issue of multiple hypothesis testing and *q*-values
189 (adjusted *p*-values) were further calculated.

190 Secretomic analysis was performed for LPS-derm bacteria [42] using different software
191 tools to predict protein transport pathways, subcellular localisations and protein categories. *In*
192 *silico* predictions of protein subcellular localisation (SCL) were accessed using PSORTb v.3.0.2
193 [43] combined with SignalP v5.0 to identify export signal peptides [44] and SecretomeP v2.0 for
194 prediction of non-classically secreted proteins [23] as well as KEGG (Kyoto encyclopedia of genes
195 and genomes) [45] and literature survey for protein secretion pathways [14, 46]. Identified proteins
196 predicted as located at the cytoplasm were filtered for moonlight activity using MoonProt v2.0
197 [47]. Combining these different predictions with literature survey and Uniprot information [48],

198 protein transport route was predicted for each protein, including export system across the
199 cytoplasmic membrane, secretion system across the outer membrane, translocational and/or post-
200 translocational maturation systems (Figure 1). The information of gene ontology (GO) for
201 molecular function and biological process was retrieved from UniprotKB and QuickGo [49, 50]
202 and enrichment of overrepresented GO terms was performed using Panther up to the ancestor level
203 2 [51]. Network analysis to identify protein-protein links was performed using STRING v11.0 [52,
204 53].

PRE-PRINT

205 **Results**

206 **Mapping the protein secretion routes following secretomic analysis of the** 207 **proteosurfaceome, proteovesiculome and exoproteome of *E. coli* O157:H7**

208 To experimentally determine the repertoire of proteins with an extracytoplasmic localisation and
209 transported from the interior to the exterior of the cell in *E. coli* O157:H7, a subproteomic analysis
210 was performed not only considering the extracellular milieu (exoproteome) but also the proteins
211 present at the cell surface (proteosurfaceome) and in outer membrane vesicles (proteovesiculome).
212 Besides, three growth media most likely to induce differential protein expression profiles in *E. coli*
213 O157:H7 were used, namely (i) a bacteria-culture minimal medium (M9), (ii) a bacteria-culture
214 rich complex medium (BHI) and (iii) a tissue-culture chemically defined medium (DMEM). In
215 order to map the transport routes of the total 154 distinct proteins experimentally identified in these
216 conditions, a secretomic analysis was performed (Figure 1 and supplementary material Table S1).

217 Scanning the amino acid sequence of these extracytoproteins, 69 proteins were predicted
218 to possess a N-terminal signal peptide (SP) and thus exported across the inner membrane (IM) by
219 the Sec system (Figures 1, 2 and supplementary material Table S1). Among them, 14 harboured a
220 SP of type II (SPII) and were therefore predicted to be supported by the lipoprotein maturation
221 pathway, that is processing by the lipoprotein diacylglyceryl transferase (Lgt), cleaved by a signal
222 peptidase of type II (SPase II) and then the lipoprotein N-acyl transferase (Lnt) [54]; half of them
223 were predicted to be anchored to the IM and thus to be inner membrane lipoproteins (IMLs),
224 whereas the other half were predicted to be outer membrane lipoproteins (OMLs) translocated to
225 the OM across the periplasm by the Lol (lipoprotein outer membrane localisation) pathway [55]
226 and two of them would be surface exposed lipoproteins (SOMLs) following translocation through
227 Bam (β -barrel assembly machinery) [56, 57]. Among the 55 proteins possessing a SPI and thus

228 cleaved by a SPase I, (i) 33 were predicted as simply exported into the periplasm, *i.e.* some
229 periplasmic proteins (PPs), (ii) one as anchored to the IM by an α -helical transmembrane domain
230 through the insertase YidC [58], *i.e.* as an IMP (inner membrane protein), and (ii) 16 as anchored
231 to the OM by β -barrel upon translocation through the insertase Bam [59], *i.e.* some OMPs (outer
232 membrane proteins). The remaining proteins with a SPI were predicted as secreted across the OM
233 through the Type II secretion system (T2SS), T5aSS, T7SS or by NCS, and were either some
234 exoproteins (soluble extracellular proteins) or subunit of cell-surface appendage (SCSA). Of note,
235 no protein exported across the IM by the alternative Tat (twin-arginine translocation) system and
236 thus exhibiting the cognate SP could be here identified.

237 Regarding the proteins lacking a SP, 8 were predicted as SCSA and thus cell-surface
238 exposed following secretion and assembly by the T3aSS to form the injectisome needle or by the
239 T3bSS to form flagella (Figure 1 and supplementary material Table S1); two were IMPs inserted
240 *via* YidC to the IM. Following the secretome concept, the presence of the cognate transport systems
241 and post-translocation maturation pathways was checked out (Figure 2); for each of the above
242 proteins actively translocated, exported and/or secreted, the proteins associated to the different
243 transport routes were indeed encoded in the genome of *E. coli* O157:H7 EDL933. The remaining
244 74 proteins lacking a SP were primarily predicted as cytoplasmic but 65 % of them (48 out of 74)
245 could also have an extracytoplasmic localisation upon NCS. Among these extracytoplasmic
246 cytoproteins (ECPs), glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is known to be found
247 in the extracellular milieu and at the cell surface of EHEC where it can moonlight by binding to
248 plasminogen and fibrinogen [60]; in this species, GAPDH was further demonstrated to be
249 piggybacked through the T3aSS [61]. Additional glycolytic enzymes known to be cell-surface
250 exposed and to moonlight as adhesins in different bacterial species were here identified, including

251 enolase, pyruvate kinase or triosephosphate isomerase [62]. Some chaperones and elongation
252 factors are also known to be cell-surface exposed, such as the elongation factor TU then
253 participating to bacterial aggregation in *E. coli* [63].

254 From the 154 distinct proteins identified, 42, 58 and 127 were found as part of the
255 proteosurfaceome, proteovesiculome and exoproteome, respectively (supplementary material
256 Table S1). While 46 proteins were found in common between the proteovesiculome and
257 exoproteome, essentially some PPs and OMPs (37 and 24 %, respectively), only 14 proteins from
258 the proteovesiculome (essentially some PPs, 79 %) and 25 proteins from the exoproteome
259 (essentially some ECPs and PPs, 44 and 40 %, respectively) were common to the
260 proteosurfaceome. With regards to the secretome analysis (Figure 2), the large majority of these
261 proteins experimentally identified were predicted as extracytoplasmic, namely 91 % (38 out of
262 42), 98 % (57 out of 58) and 80 % (102 out of 127) for the proteosurfaceome, proteovesiculome
263 and exoproteome, respectively (supplementary material Table S1). In the proteosurfaceome and
264 proteovesiculome, the majority of the proteins (60 and 90 %, respectively) were predicted as
265 associated with the cell envelope, namely in the IM, periplasm, OM, or cell surface (Figure 2 and
266 supplementary material Table S1); 60 and 79 % of these proteins were exported *via* Sec for the
267 proteosurfaceome and proteovesiculome, respectively, and were mainly predicted as PPs or
268 associated to the OM *via* Bam as OMPs or OMLs (Figure 2). While 39 % of the proteins were
269 predicted as exported *via* Sec in the exoproteome, only a few exoproteins were identified as
270 secreted by T2SS or T3SS but rather through NCS (Figure 2).

271 **Growth media induce shifting in extracytoprotein abundance and allow**
272 **identifying core and variant proteosurfaceomes, proteovesiculomes and**
273 **exoproteomes of *E. coli* O157:H7**

274 To evaluate the influence of the different culture media on proteins transported from the interior
275 to the exterior of *E. coli* O157:H7 cells, a label-free quantitative proteomic analysis was applied
276 to determine the relative protein abundance in these three subproteomes, namely the
277 proteosurfaceome, proteovesiculome and exoproteome. Shifting in protein abundance was
278 analysed individually for each subproteome comparing all three culture media simultaneously
279 (Figure 3 and Tables 1, 2 and 3). Proteins demonstrating significant differences in their abundance
280 were further referred as differential proteins and constituted the variant subproteome; at the
281 opposite, proteins found in the three growth media with no significant differences in their
282 abundance were referred as common proteins and constituted the core subproteome.

283 In the proteosurfaceome, 55 % (23 out of 42) of total identified proteins in this subproteome
284 showed a change in their relative abundance (Figure 3 and Table 1); this variant proteosurfaceome
285 was essentially constituted of PPs exported by Sec system. The majority of the differential proteins
286 (78%, 18 out of 23) presented the highest abundance when bacteria were grown in DMEM; in
287 contrast, the lowest abundance of differential proteins was observed in BHI (Table 1). Regarding
288 the core proteosurfaceome, the identified proteins were essentially ECPs transported by NCS
289 (supplementary material Table S1). Concerning the proteovesiculome, 84 % (49 out of 58) of the
290 identified proteins constituted the variant proteome as they showed a differential abundance in a
291 media-dependent manner (Figure 3 and Table 2); the differential proteins were essentially
292 exported *via* Sec to the periplasm and corresponded to PPs or further translocated to the OM as
293 OMPs, OMLs, SOMLs or SCSAs. M9 appears as the culture condition where differential proteins

294 were higher abundant, *i.e.* 78 % (38 out of 49) comparing to the other two growth media. On the
295 contrary, only around 34 % (43 out of 127) of the identified proteins constituted the variant
296 exoproteome (Figure 3 and Table 3); these differential proteins were essentially PPs exported *via*
297 Sec, OMLs and OMPs translocated to the OM, or SCSAs and exoproteins secreted across the OM.
298 The core exoproteome corresponded in the vast majority (75 %, 62 out of 83) to cytoproteins and
299 also ECPs transported by NCS (supplementary material Table S1). As for the proteovesiculome
300 (but in contrast to the proteosurfaceome), M9 condition was again the culture condition for the
301 exoproteome in which most differential proteins were identified (59 %, 26 out of 44), whereas the
302 condition showing the lowest differential abundance was the BHI (Figure 3 and Table 3).

303 To gain further insight in these subproteomes, differential and common proteins were
304 categorised according to the prediction of gene ontology (GO) classes respective to their
305 subcellular localisation (SCL), biological processes and /or molecular function (Figure 4).
306 Concerning the SCL prediction, the variant proteosurfaceome and proteovesiculome presented
307 similar profiles but differing from the exoproteome (Figure 4A). Changes in proteins abundance
308 were more perceptible on proteins predicted as located in the periplasm (GO: 0042597) and at the
309 OM (GO: 0036406) with both SCL representing more than 50 % of all differential proteins.
310 Proteins predicted as extracellular (GO: 0005576) represented more than 10 % of all differential
311 proteins in both the proteosurfaceome and proteovesiculome (Figure 4A). For the exoproteome,
312 proteins predicted as located in the periplasm represented the majority of the differential proteins
313 (74 %), whereas both exoproteins and OMPs represented less than 20 %. Proteins predicted as
314 cytoplasmic (GO: 0005737) were not identified in the variant subproteomes (Figure 4A). In
315 contrast, some cytoplasmic proteins (cytoproteins) were identified in all the core subproteomes

316 (Figure 4A); in the exoproteome, 63 % of the common proteins were ECPs trafficking through
317 NCS, some of them predicted with moonlighting activities (supplementary material Table S1).

318 To categorise which molecular functions and biological processes were affected in the three
319 studied subproteomes, a enrichment of GO terms was performed. To facilitate the data presentation
320 in Figure 4, GO terms were elevated to the ancestor level 2 (supplementary material Table S1).
321 While only a few proteins had no attributed molecular function (supplementary material Table S1),
322 proteins associated to binding (GO: 0005488) represented the majority of the identified proteins
323 in variant and core subproteomes, *i.e.* around 50 % (Figure 4B). Catalytic activity (GO: 0003824)
324 and transporter activity (GO: 0005212) represented similar percentage of identified proteins in the
325 different subproteomes, whereas structural molecule activity (GO: 0005198) represented around
326 10 % in the variant proteovesiculome and exoproteome but was completely absent in
327 proteosurfaceome. The majority of the common proteins in all core subproteomes were associated
328 with binding and catalytic activity (GO: 0005488 and 0003824). A significant number of proteins
329 related with transport activities (GO: 0005212) were identified in the variant subproteomes but not
330 in the core subproteomes.

331 The enrichment for GO terms related to biological process presented a completely different
332 trend between subproteomes (Figure 4C). Proteins identified in proteosurfaceome were essentially
333 associated with cellular adaptation to the environment. While differential proteins were associated
334 with localisation (GO: 0051179), cellular process (GO: 0009987), over 60 % of common proteins
335 were essentially related to metabolic pathways (GO: 0008152) in the proteosurfaceome and
336 exoproteome or localisation in the proteovesiculome. Remarkably, only proteins identified in the
337 variant subproteomes were associated with processes involved in interactions with other organisms
338 (GO: 0051704), cellular organisation (GO: 0071840) and biological adhesion (GO: 0022610). In

339 addition, most differential proteins from the variant exoproteome were associated to response to
340 stimulus (GO: 0050896). In the core exoproteome, 65 % of the common proteins were associated
341 with metabolic process (GO: 0008152), whereas more than the half of the core proteovesiculome
342 was represented by common proteins related to localisation process (GO: 0051179).

343 **Network analysis of the variant and core proteosurfaceomes,**
344 **proteovesiculomes and exoproteomes of *E. coli* O157:H7 highlighted functional**
345 **protein clusters related to some outer membrane proteins, ABC transport and**
346 **Type III secretion systems**

347 To investigate the links between the proteins identified in each of the variant and core
348 subproteomes of *E. coli* O157:H7 EDL933, a network analysis was performed to profile potential
349 associations and several functional clusters could be highlighted (Figure 5).

350 In the variant proteosurfaceome, several differential proteins were positively influenced in
351 their abundance in DMEM over the two other growth media M9 and BHI, especially several
352 molecular periplasmic chaperones, namely DegP, Skp, SurA, HflK and FkpA (Figure 5A,
353 Table 1). Some proteins associated to cell division as part of the Tol-Pal system were also
354 highlighted in DMEM, namely TolB and CpoB (Figure 5A and Table 1). In the variant
355 proteovesiculome, some flagellar proteins secreted and assembled through the T3bSS were clearly
356 pinpointed with a higher abundance (Figure 5A and Table 2); as such, FlgG, FlgL, FliD, FliC and
357 FlgK were up to 40 fold more abundant in BHI compared to DMEM or M9. Interestingly, in both
358 the variant proteovesiculome and exoproteome, it appeared that some proteins secreted through
359 the T3aSS showed highest abundances when bacteria were grown in DMEM, especially EspA,
360 EspB and EspD showing up to 50-fold higher in abundance in DMEM compared to BHI or M9
361 (Figure 5A). Similarly, phage major capsid proteins Z1480 and Z1479 were highlighted in this

362 growth condition. In both the proteovesiculome and exoproteome, several porins (namely OmpA,
363 OmpF and OmpX) were noticeably highlighted as up to 10-fold more abundant in M9 compared
364 to DMEM or BHI (Figure 5A). For the exoproteome, another clearly defined functional cluster
365 with higher protein abundance in M9 was related to some ABC transporters, namely with
366 periplasmic binding proteins LivJ, LivK, FliY, GlnH and YbeJ (Figure 5A). Regarding the variant
367 exoproteome and besides the virulence factor StcE secreted by a T2SS, a group of ribosomal
368 subunits transported through NCS and potentially moonlighting, *i.e.* RplC, RplS and RplA, was
369 highlighted to cluster with high abundance in DMEM (Figure 5A).

370 By definition, proteins of the core subproteomes were not influenced in their abundance by
371 the growth media, nonetheless, connectivity network analysis revealed several functional clusters
372 (Figure 5B). In the core proteosurfaceome, a clear cluster of proteins involved in protein
373 biosynthesis and folding was highlighted, namely with the elongation factor TufA and FusA, the
374 ribosomal proteins RlpA, RlpE and RpsC, molecular chaperone DnaK and GroEL, as well as FtsH
375 involved in the quality control of membrane proteins. In addition, an association network of several
376 proteins primarily related to the central catabolism was disclosed, especially with several
377 glycolytic enzymes, namely the glyceraldehyde-3-phosphate dehydrogenase GapA, the fructose-
378 bisphosphate aldolase FbaA, the pyruvate dehydrogenase components AceE and AceF, as well as
379 the pyruvate formate lyase PflB (Figure 5B). While it was scarce for the core proteovesiculome, a
380 dense protein-protein association network was observed for the core exoproteome. In the latter,
381 two main functional clusters regrouping primarily cytoproteins could be underscored, one
382 involving numerous ribosomal proteins (namely RplE, RplD, RpoA, RpsC, RpsM, RplX, RpsR,
383 RplQ and RplJ) and one involving several enzymes from central metabolic pathways, especially
384 glycolytic enzymes (*e.g.* Eno, GpmA, Pgi, LpdA, GapA, FbaA or AceE) (Figure 4B).

385 **Discussion**

386 While numerous proteomic analyses have been dedicated *E. coli* strain K12, whose secretome and
387 extracytoproteome is relatively well characterised [64-69], much fewer investigations have been
388 dedicated to the secretome of EHEC strains [70]. Bearing in mind that domesticated laboratory
389 strains of *E. coli* should not be considered as archetypical models [71], proteomic analyses of the
390 secretome appears much more relevant in commensal or pathogenic strains of *E. coli* to apprehend
391 the ecophysiology of this bacterial species. Respective to EHEC O157:H7, whole-cell or
392 cytoplasmic proteome has been studied on numerous occasions, however, only a handful of reports
393 have dedicated to the exoproteome [72-74] and, surprisingly enough, none to the proteosurfaceome
394 or proteovesiculome. From the first investigation following a 2DE approach, 29 exoproteins were
395 identified in *E. coli* O157:H7 strain EDL933 grown in LB [72]; using a LC/LC-MS/MS approach,
396 the latest study allowed to reproducibly identify 65 exoproteins from bacteria grown in M9 [74].
397 Compared to the total of 127 exoproteins presently identified from this same *E. coli* O157:H7
398 strain using three very different growth media, our work outperformed previous investigations but
399 more importantly our subcellular fraction procedure allowed to discriminate free soluble
400 extracellular proteins in the culture supernatant (true exoproteome) from proteins vectorised by
401 OMVs (proteovesiculome), which are generally mixed up in other studies of the so-called
402 exoproteome, also often misleadingly called secretome [21].

403 From the premise that secreted proteins reflect the lifestyle of bacterial cells in given
404 environmental conditions, the secretome of *E. coli* O157:H7 EDL933 was investigated using
405 distinct growth culture media most likely to induce differential protein expression profiles.
406 Following the secretome concept aiming at apprehending comprehensively the protein trafficking
407 from the interior to the exterior of the a cell considering both the secreted proteins and their cognate

408 secretion systems, the extracytoplasmic proteome of *E. coli* O157:H7 EDL933 was for the first time
409 experimentally and comprehensively deciphered considering the proteosurfaceome,
410 proteovesiculome and exoproteome. From (i) the export across or insertion in the IM, (ii) the post-
411 translocational maturations in the periplasm, and (iii) the secretion across, anchoring to or insertion
412 in the OM, the different systems involved in protein translocation and transport were identified by
413 proteogenomic analyses. This analytical approach further allowed to attribute each of the
414 extracytoplasmic proteins experimentally identified with the respective transport pathways and
415 thus provided a map of the protein trafficking routes. Besides, this secretome-based analysis
416 permitted to ascribe each of the experimentally identified proteins to a category and different SCL.

417 While secreted proteins generally require a N-terminal SP, *e.g.* proteins first exported by the
418 Sec system and then further secreted by the T2SS, T5SS or T7SS (Figure 1), some other proteins
419 lack such a SP, *e.g.* proteins secreted *via* a T3SS. From sequence similarity, such secreted proteins
420 could be deduced based on homology and a secretion pathway could then be attributed. Besides
421 these known secretion systems, however, no secretion system could be attributed to some
422 identified exoproteins. Such proteins could be transported by NCS, that is by unknown secretion
423 systems or mechanisms [22, 23]. It can be wise to further discriminate NCS from unconventional
424 protein trafficking (UCT), in the sense that NCS refers to an active protein transport mechanism
425 through a conducting channel, whereas UCT refers to mechanisms unrelated to secretion *per se*
426 but participating to protein trafficking from inside to outside the cell compartment. Different NCS
427 mechanisms can be at play, such as piggybacking as demonstrated in *E. coli* O157:H7 for the
428 GAPDH, which associates to T3aSS effectors and is thus non-specifically secreted *via* this system
429 [61]. Among ECPs, it is interesting to note that a number is well-known to associate to other
430 proteins, especially the molecular chaperone in the course of protein folding and quality check, or

431 ribosomal proteins in the course of protein synthesis, especially upon co-translational
432 translocational. Including glycolytic enzyme known to moonlight at the bacterial cell surface in
433 numerous bacterial species [62], the hypothesis that such ECPs could be secreted by piggybacking
434 would clearly necessitate further in-depth investigations. One of the most well-known UCT
435 mechanisms involve OMVs. Upon blebbing of the OM, OMVs carrying periplasmic and OM-
436 associated proteins are released in the culture supernatant; it cannot be excluded that some of them
437 burst out liberating their luminal content (with periplasmic proteins) as well as OM-associated
438 proteins in the extracellular milieu. Contrary to most exoproteomic analyses that just recover the
439 culture supernatant, the present investigation took a great care to differentiate the exoproteome
440 from the proteovesiculome by separating the culture supernatant into two fractions. As suggested
441 by the identification of phage capsid proteins, some ECPs could be released upon activation of
442 lysogenic bacteriophage entering into a lytic cycle. As an UCT mechanism, bacterial cell lysis
443 could also be induced in the course of control cell lysis by siblings, *i.e.* autolysis [75] or following
444 apoptosis upon holin activation [76].

445 Investigating the proteosurfaceome, exoproteome and proteovesiculome, a comprehensive
446 picture of the experimental extracytoplasmic proteome landscape of *E. coli* O157:H7 EDL933 was
447 obtained for the first time. Proteomic quantitative analysis revealed significant changes in all
448 subproteomes. Remarkably, 76 % of the total proteins identified in the extracellular milieu
449 changed their abundance when comparing the three culture media investigated. Actually, proteins
450 found in the exoproteome appeared with the highest abundance in minimal chemically defined
451 medium M9. While M9 has sometimes been considered as inhibiting protein secretion [77], it must
452 be stressed that minimal media are well-known in bacterial physiology to trigger the full anabolic
453 potential of a cell. Indeed, in M9 neither macromolecules (*e.g.* lipids, polysaccharides, proteins,

454 DNA) nor elementary bricks (*e.g.* fatty acids, glucides, amino acids, nucleotides) are present, thus
455 the cell needs to synthesise them all *de novo*. While BHI is clearly a rich medium providing a
456 whole set of biochemical macromolecules that needs first to be degraded (depolymerised) before
457 being use by the cell, DMEM can be viewed as an in-between since macromolecules are absent but
458 some elementary bricks (namely amino acids or some vitamins) are readily available for the cell;
459 in the latter and contrary to growth in M9, the cell does not necessarily trigger the basic
460 biosynthetic pathways to synthesise them but at least some polymerisation and assembly pathways.
461 In agreement with the fact that protein secretion is an essential anabolic cellular function, changes
462 in the extracytoproteome (including the proteosurfaceome, proteovesiculome and exoproteome)
463 were clearly demonstrated in M9 compared to the rich media. In agreement with active proteins
464 trafficking in the cell, the biological processes related to transport, tethering or maintenance of
465 molecules in a specific location (GO: 0051179) were the most affected by culture conditions in
466 these subproteomes. Considering the biochemical nature of molecules to be transported differs
467 between rich complex and chemically defined media (*i.e.* use of partially depolymerised
468 compounds versus biosynthesis of elementary bricks, respectively) to support bacterial
469 metabolism and growth, the differential bacterial physiology in these conditions could explain that
470 a significant number of proteins related to transport activities (GO: 0005212) were identified in
471 the variant subproteomes rather than in the core subproteomes. This investigation allowed
472 providing the first dynamic picture of extracytoplasmic proteins differentially or similarly
473 expressed in response to different culture conditions respective to their predicted SCL and
474 biological processes. All-in-all and by modulating the levels of protein expression in the
475 proteosurfaceome, proteovesiculome and exoproteome, *E. coli* O157 appeared to adapt its secretion
476 profile depending on the environmental conditions.

477 In allowing the transport of protein effectors, secretion systems are closely related to the
478 main virulence features in pathogenic bacteria. The present work revealed that DMEM induced
479 major changes in the proteosurfaceome comparing to the other two growth media. As expected
480 from the up-regulation of the LEE in DMEM [77, 78], the abundance of proteins related with the
481 injectisome (T3aSS) was higher in this medium. Interestingly, in this condition the abundance of
482 proteins related to flagellum (T3bSS) was lower, which could suggest some kind of counter-
483 regulation of these two types of T3SS. While such regulation trend has never been reported for
484 *E. coli* O157:H7 EDL933 and would require further in-depth investigations, in *Yersinia* and
485 *Shigella*, flagellar genes are induced at temperatures below 30°C but injectisome genes are
486 transcribed at 37°C [79-82]. In *E. coli* O157:H7, the induction of the LEE in DMEM would result
487 from the presence of sodium bicarbonate, which enhances the transcription of *ler* [36]. The effect
488 of NaHCO₃ on the expression of other genes and secretion of proteins remain unknown but would
489 undoubtedly require further investigations. Actually, differential expression of other virulence
490 factors in *E. coli* O157:H7 has been reported in the literature but generally relies on transcriptional
491 analysis [36, 37, 77, 78, 83-88] or on adhesion phenotype to various cell models [34, 83, 89, 90].
492 While transcriptomics is useful for initial information about the genetic expression in some given
493 conditions, the correlation between mRNA level and protein abundance is known to be poor and
494 at most indicates that a protein is likely expressed when mRNA is present [30]. As such,
495 proteomics remains the ultimate method of choice to ascertain that proteins are effectively
496 expressed, especially for secreted proteins since the regulation level is not only transcriptional but
497 can occur at pre-transcriptional, post-transcriptional, post-translational and/or post-translocational
498 levels. Besides the influence of the nutrient composition, the bacterial growth rate *per se* could
499 influence protein expression and secretion, such as the flagellar biosynthesis [57]; a rigorous

500 investigation of the influence of the bacterial growth rate in a given medium, thought, cannot be
501 performed in batch culture but in chemostat cultures [91, 92]. Additional environmental
502 parameters, especially temperature and pH, could also be tested to assay their effects on the
503 expression of the secreted proteins and thus broaden our knowledge of the secretome landscape.
504 In fact, the experimental identification of extracytoplasmic proteins is necessary to better
505 understand the physiology and ecophysiology of *E. coli* O157:H7 in various environments.
506 Extracytoproteins have also practical importance for the development of preventive strategies
507 against EHEC and/or DEC contamination along the food chain and ultimately human infection.
508 This can include the development of antagonists targeting important virulence or colonisation
509 factors but also the identification of potential antigens for vaccine development [93, 94].

510 **Author's contributions**

511 RM and MD conceptualise the overarching aims of the research study. RM, IC, SL, CC and MD
512 conceived and designed the experiments. RM, IC, VA and CC performed the experiments and data
513 acquisition. RM, IC, VA, SL, CC, MH, VL, MP, AP and MD analysed and interpreted the data.
514 MD had management as well as coordination responsibility for the execution of the research work.
515 MP and MD contributed to the acquisition of the financial supports and resources leading to this
516 publication. RM, IC, VA, SL, CC, MH, VL, MP, AP and MD wrote the article, including drafting
517 and revising critically the manuscript for important intellectual content.

518

519 RM: Conceptualisation, Methodology, Formal analysis, Investigation, Writing - original draft,
520 Writing - review & editing, Visualisation. IC: Conceptualisation, Methodology, Formal analysis,
521 Investigation, Writing - original draft, Writing - review & editing, Visualisation, Supervision. VA:
522 Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualisation.
523 SL: Conceptualisation, Formal analysis, Writing - original draft, Writing - review & editing,
524 Visualisation. CC: Conceptualisation, Validation, Formal analysis, Investigation, Resources,
525 Writing - original draft, Writing - review & editing. MH: Formal analysis, Resources, Writing -
526 review & editing, Supervision. VL: Formal analysis, Writing - review & editing. MP: Formal
527 analysis, Resources, Writing - review & editing, Supervision, Funding acquisition. AP: Formal
528 analysis, Writing - review & editing, Supervision. MD Conceptualisation, Resources, Formal
529 analysis, Investigation, Writing - original draft, Writing - review & editing, Visualisation,
530 Supervision, Project administration, Funding acquisition.

531

532 **Competing interests**

533 MP and AP are permanent employees of the GSK group of companies. All other authors have
534 declared no competing interests.

535 **Acknowledgements**

536 This work was supported in part by (i) INRAE ("Institut National de Recherche pour l'Agriculture,
537 l'Alimentation et l'Environnement", previously called INRA, "Institut National de la Recherche
538 Agronomique"), (ii) EU FP7 Marie Curie Actions ITN (Initial Training Network) EID (European
539 Industrial Doctorates) DISCo (A multidisciplinary Doctoral Industrial School on novel preventive
540 strategies against *E. Coli* infections; <http://www.discoproject.eu/>; n°FP7-PEOPLE-607611)
541 project, (iii) CoMBa ("Colonisation de la Matrice extracellulaire Bacterienne") project funded by
542 the "Région Auvergne" FRI ("Fond Régional Innovation") IRP ("Institut de Recherche
543 Pharmabiotique") cluster, and (iv) ANR (Agence National de la Recherche Agronomique)
544 PathoFood (Deciphering the cell heterogeneity and the spatial distribution of foodborne bacterial
545 pathogens in food matrices in interaction with the microbial communities) project (n°ANR-17-
546 CE21-0002). The authors thank Nelly CACCIA for excellent technical support and Philippe RUIZ
547 for help and advices in statistical analysis. Dr Ricardo MONTEIRO was a Marie Curie PhD
548 Research Fellow granted by ITN EID DISCo. Dr Valentin AGEORGES was a PhD Research
549 Fellow granted by the "Conseil Régional d'Auvergne - FEDER (Fonds Européen de
550 Développement Régional)".

551 **Supplementary material**

552 **Table S1 Total proteins identified in the proteosurfaceome, proteovesiculome and**
553 **exoproteome of *E. coli* O157:H7 EDL933**

554 **References**

- 555 [1] L.V. Hooper, J.I. Gordon, Commensal host-bacterial relationships in the gut, *Science*
556 292(5519) (2001) 1115-8.
- 557 [2] J.P. Nataro, J.B. Kaper, Diarrheagenic *Escherichia coli*, *Clin Microbiol Rev* 11(1) (1998) 142-
558 201.
- 559 [3] M.B. Cohen, J.P. Nataro, D.I. Bernstein, J. Hawkins, N. Roberts, M.A. Staat, Prevalence of
560 diarrheagenic *Escherichia coli* in acute childhood enteritis: a prospective controlled study, *J*
561 *Pediatr* 146(1) (2005) 54-61.
- 562 [4] H. Karch, P.I. Tarr, M. Bielaszewska, Enterohaemorrhagic *Escherichia coli* in human
563 medicine, *Int J Med Microbiol* 295(6-7) (2005) 405-18.
- 564 [5] M. Ahmed, J. Van Velkinburgh, Enterohemorrhagic *Escherichia coli* O157 in North Africa
565 region: a threat require advanced investigation, *Pan Afr Med J* 19 (2014) 26.
- 566 [6] *Bergey's Manual of Systematic Bacteriology* 2nd ed., Springer US2005.
- 567 [7] M.A. Karmali, M. Mascarenhas, S. Shen, K. Ziebell, S. Johnson, R. Reid-Smith, J. Isaac-
568 Renton, C. Clark, K. Rahn, J.B. Kaper, Association of genomic O island 122 of *Escherichia coli*
569 EDL933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic
570 and/or serious disease, *J Clin Microbiol* 41(11) (2003) 4930-40.
- 571 [8] M.M. Kendall, Interkingdom chemical signaling in enterohemorrhagic *Escherichia coli*
572 O157:H7, *Adv Exp Med Biol* 874 (2016) 201-13.
- 573 [9] G. Frankel, A.D. Phillips, I. Rosenshine, G. Dougan, J.B. Kaper, S. Knutton, Enteropathogenic
574 and enterohaemorrhagic *Escherichia coli*: more subversive elements, *Mol Microbiol* 30(5) (1998)
575 911-21.

576 [10] M.J. Farfan, A.G. Torres, Molecular mechanisms that mediate colonization of shigatoxin-
577 producing *Escherichia coli* strains, *Infect Immun* 80(3) (2012) 903-13.

578 [11] S.J. Elliott, L.A. Wainwright, T.K. McDaniel, K.G. Jarvis, Y.K. Deng, L.C. Lai, B.P.
579 McNamara, M.S. Sonnenberg, J.B. Kaper, The complete sequence of the locus of enterocyte
580 effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69, *Mol Microbiol* 28(1) (1998)
581 1-4.

582 [12] T.K. McDaniel, K.G. Jarvis, M.S. Sonnenberg, J.B. Kaper, A genetic locus of enterocyte
583 effacement conserved among diverse enterobacterial pathogens, *Proc Natl Acad Sci U S A* 92(5)
584 (1995) 1664-8.

585 [13] C. Hill, Virulence or niche factors: what's in a name?, *J Bacteriol* 194(21) (2012) 5725-7.

586 [14] R. Monteiro, V. Ageorges, M. Rojas-Lopez, H. Schmidt, A. Weiss, Y. Bertin, E. Forano, G.
587 Jubelin, I.R. Henderson, V. Livrelli, A.P. Gobert, R. Rosini, M. Soriani, M. Desvaux, A secretome
588 view of colonisation factors in shigatoxin-encoding *Escherichia coli* (STEC): from
589 enterohaemorrhagic *E. coli* (EHEC) to related enteropathotypes, *FEMS Microbiol Lett* 363(16)
590 (2016).

591 [15] P. Vogeleer, Y.D. Tremblay, A.A. Mafu, M. Jacques, J. Harel, Life on the outside: role of
592 biofilms in environmental persistence of shigatoxin producing *Escherichia coli*, *Front Microbiol*
593 5 (2014) 317.

594 [16] S.M. Chekabab, J. Paquin-Veillette, C.M. Dozois, J. Harel, The ecological habitat and
595 transmission of *Escherichia coli* O157:H7, *FEMS Microbiol Lett* 341(1) (2013) 1-12.

596 [17] J. Sanchez-Villamil, F. Navarro-Garcia, Role of virulence factors on host inflammatory
597 response induced by diarrheagenic *Escherichia coli* pathotypes, *Future Microbiol* 10(6) (2015)
598 1009-33.

- 599 [18] M.P. Stevens, G.M. Frankel, The Locus of Enterocyte Effacement and Associated Virulence
600 Factors of Enterohemorrhagic *Escherichia coli*, *Microbiol Spectr* 2(4) (2014) EHEC-0007-2013.
- 601 [19] J.J. LeBlanc, Implication of virulence factors in *Escherichia coli* O157:H7 pathogenesis, *Crit*
602 *Rev Microbiol* 29(4) (2003) 277-96.
- 603 [20] M. Gardette, S. Le Hello, P. Mariani-Kurkdjian, L. Fabre, F. Gravey, A. Garrivier, E.
604 Loukiadis, G. Jubelin, Identification and prevalence of *in vivo*-induced genes in
605 enterohaemorrhagic *Escherichia coli*, *Virulence* 10(1) (2019) 180-193.
- 606 [21] M. Desvaux, M. Hebraud, R. Talon, I.R. Henderson, Secretion and subcellular localizations
607 of bacterial proteins: a semantic awareness issue, *Trends Microbiol* 17(4) (2009) 139-45.
- 608 [22] J.D. Bendtsen, Wooldridge, K. G., Bacterial secreted proteins: Secretory mechanisms and role
609 in pathogenesis - Chapter 10: Non-classical secretion, Caister Academic Press 2009.
- 610 [23] J.D. Bendtsen, L. Kiemer, A. Fausboll, S. Brunak, Non-classical protein secretion in bacteria,
611 *BMC Microbiol* 5 (2005) 58.
- 612 [24] R. Lloubes, A. Bernadac, L. Houot, S. Pommier, Non classical secretion systems, *Res*
613 *Microbiol* 164(6) (2013) 655-63.
- 614 [25] G. Wang, H. Chen, Y. Xia, J. Cui, Z. Gu, Y. Song, Y.Q. Chen, H. Zhang, W. Chen, How are
615 the non-classically secreted bacterial proteins released into the extracellular milieu?, *Curr*
616 *Microbiol* 67(6) (2013) 688-95.
- 617 [26] Q. Kang, D. Zhang, Principle and potential applications of the non-classical protein secretory
618 pathway in bacteria, *Appl Microbiol Biotechnol* 104(3) (2020) 953-965.
- 619 [27] R.G. Gerlach, M. Hensel, Protein secretion systems and adhesins: the molecular armory of
620 Gram-negative pathogens, *Int J Med Microbiol* 297(6) (2007) 401-15.

621 [28] C. Schwechheimer, M.J. Kuehn, Outer-membrane vesicles from Gram-negative bacteria:
622 biogenesis and functions, *Nat Rev Microbiol* 13(10) (2015) 605-19.

623 [29] G. Wang, Y. Xia, X. Song, L. Ai, Common non-classically secreted bacterial proteins with
624 experimental evidence, *Curr Microbiol* 72(1) (2016) 102-11.

625 [30] C. Vogel, E.M. Marcotte, Insights into the regulation of protein abundance from proteomic
626 and transcriptomic analyses, *Nat Rev Genet* 13(4) (2012) 227-32.

627 [31] T. Maier, M. Guell, L. Serrano, Correlation of mRNA and protein in complex biological
628 samples, *FEBS Lett* 583(24) (2009) 3966-73.

629 [32] Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA
630 abundance, *Cell* 165(3) (2016) 535-50.

631 [33] C. Chagnot, A. Agus, S. Renier, F. Peyrin, R. Talon, T. Astruc, M. Desvaux, In vitro
632 colonization of the muscle extracellular matrix components by *Escherichia coli* O157:H7: the
633 influence of growth medium, temperature and pH on initial adhesion and induction of biofilm
634 formation by collagens I and III, *PLoS One* 8(3) (2013) e59386.

635 [34] X. Yin, Y. Feng, R. Wheatcroft, J. Chambers, J. Gong, C.L. Gyles, Adherence of *Escherichia*
636 *coli* O157:H7 to epithelial cells *in vitro* and in pig gut loops is affected by bacterial culture
637 conditions, *Can J Vet Res* 75(2) (2011) 81-8.

638 [35] T.E. Grys, M.B. Siegel, W.W. Lathem, R.A. Welch, The StcE protease contributes to intimate
639 adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells, *Infect Immun* 73(3)
640 (2005) 1295-303.

641 [36] H. Abe, I. Tatsuno, T. Tobe, A. Okutani, C. Sasakawa, Bicarbonate ion stimulates the
642 expression of locus of enterocyte effacement-encoded genes in enterohemorrhagic *Escherichia*
643 *coli* O157:H7, *Infect Immun* 70(7) (2002) 3500-9.

644 [37] F. Beltrametti, A.U. Kresse, C.A. Guzman, Transcriptional regulation of the *esp* genes of
645 enterohemorrhagic *Escherichia coli*, *J Bacteriol* 181(11) (1999) 3409-18.

646 [38] N. Solis, S.J. Cordwell, Current methodologies for proteomics of bacterial surface-exposed
647 and cell envelope proteins, *Proteomics* 11(15) (2011) 3169-89.

648 [39] R. Monteiro, I. Chafsey, S. Leroy, C. Chambon, M. Hebraud, V. Livrelli, M. Pizza, A.
649 Pezzicoli, M. Desvaux, Differential biotin labelling of the cell envelope proteins in
650 lipopolysaccharidic diderm bacteria: Exploring the proteosurfaceome of *Escherichia coli* using
651 sulfo-NHS-SS-biotin and sulfo-NHS-PEG4-bismannose-SS-biotin, *J Proteomics* 181 (2018) 16-
652 23.

653 [40] L.W. Riley, R.S. Remis, S.D. Helgerson, H.B. McGee, J.G. Wells, B.R. Davis, R.J. Hebert,
654 E.S. Olcott, L.M. Johnson, N.T. Hargrett, P.A. Blake, M.L. Cohen, Hemorrhagic colitis associated
655 with a rare *Escherichia coli* serotype, *N Engl J Med* 308(12) (1983) 681-5.

656 [41] S. Renier, I. Chafsey, C. Chambon, N. Caccia, A. Charbit, M. Hebraud, M. Desvaux,
657 Contribution of the multiple Type I signal peptidases to the secretome of *Listeria monocytogenes*:
658 deciphering their specificity for secreted exoproteins by exoproteomic analysis, *J Proteomics* 117
659 (2015) 95-105.

660 [42] M.F. Romine, Genome-wide protein localization prediction strategies for Gram negative
661 bacteria, *BMC Genomics* 12 Suppl 1 (2011) S1.

662 [43] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester,
663 L.J. Foster, F.S. Brinkman, PSORTb 3.0: improved protein subcellular localization prediction with
664 refined localization subcategories and predictive capabilities for all prokaryotes, *Bioinformatics*
665 26(13) (2010) 1608-15.

666 [44] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal
667 peptides from transmembrane regions, *Nat Methods* 8(10) (2011) 785-6.

668 [45] J. Du, Z. Yuan, Z. Ma, J. Song, X. Xie, Y. Chen, KEGG-PATH: Kyoto encyclopedia of genes
669 and genomes-based pathway analysis using a path analysis model, *Mol Biosyst* 10(9) (2014) 2441-
670 7.

671 [46] C. Chagnot, M.A. Zorgani, T. Astruc, M. Desvaux, Proteinaceous determinants of surface
672 colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion
673 perspective, *Front Microbiol* 4 (2013) 303.

674 [47] C. Chen, S. Zabad, H. Liu, W. Wang, C. Jeffery, MoonProt 2.0: an expansion and update of
675 the moonlighting proteins database, *Nucleic Acids Res* 46(D1) (2018) D640-D644.

676 [48] The UniProt Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Res*
677 45(D1) (2017) D158-D169.

678 [49] E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, A. Bairoch, UniProtKB/Swiss-Prot,
679 *Methods Mol Biol* 406 (2007) 89-112.

680 [50] D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan, R. Apweiler, QuickGO: a web-
681 based tool for Gene Ontology searching, *Bioinformatics* 25(22) (2009) 3045-6.

682 [51] P.D. Thomas, M.J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A.
683 Muruganujan, A. Narechania, PANTHER: a library of protein families and subfamilies indexed
684 by function, *Genome Res* 13(9) (2003) 2129-41.

685 [52] C. von Mering, L.J. Jensen, B. Snel, S.D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M.A.
686 Huynen, P. Bork, STRING: known and predicted protein-protein associations, integrated and
687 transferred across organisms, *Nucleic Acids Res* 33(Database issue) (2005) D433-7.

688 [53] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M.
689 Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering,
690 STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids*
691 *Res* 43(Database issue) (2015) D447-52.

692 [54] A. Kovacs-Simon, R.W. Titball, S.L. Michell, Lipoproteins of bacterial pathogens, *Infect*
693 *Immun* 79(2) (2011) 548-61.

694 [55] S.I. Narita, H. Tokuda, Bacterial lipoproteins; biogenesis, sorting and quality control, *Biochim*
695 *Biophys Acta Mol Cell Biol Lipids* 1862(11) (2017) 1414-1423.

696 [56] M. Tata, A. Konovalova, Improper coordination of BamA and BamD results in Bam complex
697 jamming by a lipoprotein substrate, *MBio* 10(3) (2019).

698 [57] E.M. Hart, M. Gupta, M. Wühr, T.J. Silhavy, The synthetic phenotype of
699 $\Delta bamB \Delta bamE$ double mutants results from a lethal jamming of the Bam complex by the
700 lipoprotein RcsF, *MBio* 10(3) (2019).

701 [58] A. Kuhn, D. Kiefer, Membrane protein insertase YidC in bacteria and archaea, *Mol Microbiol*
702 103(4) (2017) 590-594.

703 [59] N. Noinaj, S.E. Rollauer, S.K. Buchanan, The β -barrel membrane protein insertase machinery
704 from Gram-negative bacteria, *Curr Opin Struct Biol* 31 (2015) 35-42.

705 [60] L. Egea, L. Aguilera, R. Giménez, M.A. Sorolla, J. Aguilar, J. Badía, L. Baldoma, Role of
706 secreted glyceraldehyde-3-phosphate dehydrogenase in the infection mechanism of
707 enterohemorrhagic and enteropathogenic *Escherichia coli*: interaction of the extracellular enzyme
708 with human plasminogen and fibrinogen, *Int J Biochem Cell Biol* 39(6) (2007) 1190-203.

709 [61] L. Aguilera, E. Ferreira, R. Giménez, F.J. Fernández, M. Taulés, J. Aguilar, M.C. Vega, J.
710 Badia, L. Baldomà, Secretion of the housekeeping protein glyceraldehyde-3-phosphate

711 dehydrogenase by the LEE-encoded type III secretion system in enteropathogenic *Escherichia*
712 *coli*, Int J Biochem Cell Biol 44(6) (2012) 955-62.

713 [62] B. Henderson, A. Martin, Bacterial virulence in the moonlight: multitasking bacterial
714 moonlighting proteins are virulence determinants in infectious disease, Infect Immun 79(9) (2011)
715 3476-91.

716 [63] P. Amimanan, R. Tavichakortrakool, K. Fong-Ngern, P. Sribenjalux, A. Lulitanond, V.
717 Prasongwatana, C. Wongkham, P. Boonsiri, J. Umka Welbat, V. Thongboonkerd, Elongation
718 factor Tu on *Escherichia coli* isolated from urine of kidney stone patients promotes calcium
719 oxalate crystal growth and aggregation, Sci Rep 7(1) (2017) 2953.

720 [64] X.X. Xia, M.J. Han, S.Y. Lee, J.S. Yoo, Comparison of the extracellular proteomes of
721 *Escherichia coli* B and K-12 strains during high cell density cultivation, Proteomics 8(10) (2008)
722 2089-103.

723 [65] M.J. Han, Exploring the proteomic characteristics of the *Escherichia coli* B and K-12 strains
724 in different cellular compartments, J Biosci Bioeng 122(1) (2016) 1-9.

725 [66] M.P. Nandakumar, A. Cheung, M.R. Marten, Proteomic analysis of extracellular proteins
726 from *Escherichia coli* W3110, J Proteome Res 5(5) (2006) 1155-61.

727 [67] M.J. Han, S.Y. Lee, S.H. Hong, Comparative analysis of envelope proteomes in *Escherichia*
728 *coli* B and K-12 strains, J Microbiol Biotechnol 22(4) (2012) 470-8.

729 [68] M.P. Molloy, B.R. Herbert, M.B. Slade, T. Rabilloud, A.S. Nouwens, K.L. Williams, A.A.
730 Gooley, Proteomic analysis of the *Escherichia coli* outer membrane, Eur J Biochem 267(10)
731 (2000) 2871-81.

732 [69] A. Boysen, J. Borch, T.J. Krogh, K. Hjerno, J. Moller-Jensen, SILAC-based comparative
733 analysis of pathogenic *Escherichia coli* secretomes, J Microbiol Methods 116 (2015) 66-79.

734 [70] R.S. Nirujogi, B. Muthusamy, M.S. Kim, G.J. Sathe, P.T. Lakshmi, O.N. Kovbasnjuk, T.S.
735 Prasad, M. Wade, R.E. Jabbour, Secretome analysis of diarrhea-inducing strains of *Escherichia*
736 *coli*, *Proteomics* 17(6) (2017).

737 [71] J.L. Hobman, C.W. Penn, M.J. Pallen, Laboratory strains of *Escherichia coli*: model citizens
738 or deceitful delinquents growing old disgracefully?, *Mol Microbiol* 64(4) (2007) 881-5.

739 [72] M. Li, I. Rosenshine, S.L. Tung, X.H. Wang, D. Friedberg, C.L. Hew, K.Y. Leung,
740 Comparative proteomic analysis of extracellular proteins of enterohemorrhagic and
741 enteropathogenic *Escherichia coli* strains and their ihf and ler mutants, *Appl Environ Microbiol*
742 70(9) (2004) 5274-82.

743 [73] R.E. Jabbour, S.V. Deshpande, P.E. McCubbin, J.D. Wright, M.M. Wade, A.P. Snyder,
744 Extracellular protein biomarkers for the characterization of enterohemorrhagic and
745 enteroaggregative *Escherichia coli* strains, *J Microbiol Methods* 98 (2014) 76-83.

746 [74] N. Islam, A. Nagy, W.M. Garrett, D. Shelton, B. Cooper, X. Nou, Different cellular origins
747 and functions of extracellular proteins from *Escherichia coli* O157:H7 and O104:H4 as determined
748 by comparative proteomic analysis, *Appl Environ Microbiol* 82(14) (2016) 4371-4378.

749 [75] P.F. Popp, T. Mascher, Coordinated cell death in isogenic bacterial populations: Sacrificing
750 some for the benefit of many?, *J Mol Biol* (2019).

751 [76] M. Desvaux, Contribution of holins to protein trafficking: secretion, leakage or lysis?, *Trends*
752 *Microbiol* 20(6) (2012) 259-61.

753 [77] B. Kenny, A. Abe, M. Stein, B.B. Finlay, Enteropathogenic *Escherichia coli* protein secretion
754 is induced in response to conditions similar to those in the gastrointestinal tract, *Infect Immun*
755 65(7) (1997) 2606-12.

756 [78] I. Rosenshine, S. Ruschkowski, B.B. Finlay, Expression of attaching/effacing activity by
757 enteropathogenic *Escherichia coli* depends on growth phase, temperature, and protein synthesis
758 upon contact with epithelial cells, *Infect Immun* 64(3) (1996) 966-73.

759 [79] M. Skurnik, I. Bolin, H. Heikkinen, S. Piha, H. Wolf-Watz, Virulence plasmid-associated
760 autoagglutination in *Yersinia* spp, *J Bacteriol* 158(3) (1984) 1033-6.

761 [80] T. Michiels, J.C. Vanooteghem, C. Lambert de Rouvroit, B. China, A. Gustin, P. Boudry,
762 G.R. Cornelis, Analysis of virC, an operon involved in the secretion of Yop proteins by *Yersinia*
763 *enterocolitica*, *J Bacteriol* 173(16) (1991) 4994-5009.

764 [81] N.P. Hoe, J.D. Goguen, Temperature sensing in *Yersinia pestis*: translation of the LcrF
765 activator protein is thermally regulated, *J Bacteriol* 175(24) (1993) 7901-9.

766 [82] M. Falconi, B. Colonna, G. Prosseda, G. Micheli, C.O. Gualerzi, Thermoregulation of
767 *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition
768 of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS, *EMBO J*
769 17(23) (1998) 7033-43.

770 [83] F. Ebel, C. Deibel, A.U. Kresse, C.A. Guzman, T. Chakraborty, Temperature- and medium-
771 dependent secretion of proteins by shigatoxin-producing *Escherichia coli*, *Infect Immun* 64(11)
772 (1996) 4472-9.

773 [84] G.S. Abu-Ali, L.M. Ouellette, S.T. Henderson, T.S. Whittam, S.D. Manning, Differences in
774 adherence and virulence gene expression between two outbreak strains of enterohaemorrhagic
775 *Escherichia coli* O157 : H7, *Microbiology* 156(Pt 2) (2010) 408-19.

776 [85] B. House, J.V. Kus, N. Prayitno, R. Mair, L. Que, F. Chingcuanco, V. Gannon, D.G.
777 Cvitkovitch, D. Barnett Foster, Acid-stress-induced changes in enterohaemorrhagic *Escherichia*
778 *coli* O157 : H7 virulence, *Microbiology* 155(Pt 9) (2009) 2907-18.

779 [86] H. Hirakawa, T. Kodama, A. Takumi-Kobayashi, T. Honda, A. Yamaguchi, Secreted indole
780 serves as a signal for expression of type III secretion system translocators in enterohaemorrhagic
781 *Escherichia coli* O157:H7, *Microbiology* 155(Pt 2) (2009) 541-50.

782 [87] N. Nakanishi, K. Tashiro, S. Kuhara, T. Hayashi, N. Sugimoto, T. Tobe, Regulation of
783 virulence by butyrate sensing in enterohaemorrhagic *Escherichia coli*, *Microbiology* 155(Pt 2)
784 (2009) 521-30.

785 [88] S.K. Anand, M.W. Griffiths, Quorum sensing and expression of virulence in *Escherichia coli*
786 O157:H7, *Int J Food Microbiol* 85(1-2) (2003) 1-9.

787 [89] F. Girard, F. Dziva, P. van Diemen, A.D. Phillips, M.P. Stevens, G. Frankel, Adherence of
788 enterohemorrhagic *Escherichia coli* O157, O26, and O111 strains to bovine intestinal explants *ex*
789 *vivo*, *Appl Environ Microbiol* 73(9) (2007) 3084-90.

790 [90] J. Xicohtencatl-Cortes, V. Monteiro-Neto, M.A. Ledesma, D.M. Jordan, O. Francetic, J.B.
791 Kaper, J.L. Puente, J.A. Giron, Intestinal adherence associated with type 4 pili of
792 enterohemorrhagic *Escherichia coli* O157:H7, *J Clin Invest* 117(11) (2007) 3519-29.

793 [91] K. Kovárová-Kovar, T. Egli, Growth kinetics of suspended microbial cells: from single-
794 substrate-controlled growth to mixed-substrate kinetics, *Microbiol Mol Biol Rev* 62(3) (1998)
795 646-66.

796 [92] P.A. Hoskisson, G. Hobbs, Continuous culture--making a comeback?, *Microbiology* 151(Pt
797 10) (2005) 3153-9.

798 [93] M. Rojas-Lopez, R. Monterio, M. Pizza, M. Desvaux, R. Rosini, Intestinal pathogenic
799 *Escherichia coli*: Insights for vaccine development, *Front Microbiol* 9 (2018) 440.

800 [94] M. Rojas-Lopez, M. Martinelli, V. Brandi, G. Jubelin, F. Polticelli, M. Soriani, M. Pizza, M.
801 Desvaux, R. Rosini, Identification of lipid A deacylase as a novel, highly conserved and protective
802 antigen against enterohemorrhagic *Escherichia coli*, *Sci Rep* 9(1) (2019) 17014.

PRE-PRINT

803 **Figures legends**

804 **Figure 1 Decision tree for predicting the subcellular localisation of proteins in a LPS-diderm** 805 **bacterium based on the biology of protein secretion.**

806 The protein sequences are sequentially analysed in a workflow for (i) signal peptide (SP), (ii) the
807 type of SP, (iii) the presence of transmembrane (TM) α -helices responsible for IM retention, (iv)
808 the delivery of lipoprotein to the Lol system, (v) the presence of the presence of TM β -barrel
809 responsible for OM retention, and (vi) the systems involved in translocation into or across the OM.
810 In the end, proteins are discriminated first into different protein categories and then for their final
811 subcellular localisation (SCL).

812 SP I: SP of type I, SP II: SP of type II, SP IV: SP of type IV, Tat-SP: twin-arginine translocation
813 SP, Lol: localisation of lipoproteins, Bam: β -barrel assembly machinery, T2SS: Type II secretion
814 system, T3aSS: Type III, subtype a, secretion system, T3bSS: Type III, subtype b, secretion
815 system, T5aSS: Type V, subtype a, secretion system, T5eSS: Type V, subtype e, secretion system,
816 T7SS: Type VII secretion system, NCS: non-classical secretion, CP: cytoplasmic protein, IMP: inner
817 membrane protein, IML: inner membrane lipoprotein, PP: periplasmic protein, OML: outer
818 membrane lipoprotein, OMP: outer membrane protein, SOML: surface outer membrane
819 lipoprotein, SCSA: subunit of cell surface appendage, EP: exoprotein, ECP: extracytoplasmic
820 cytoplasmic protein, CY: cytoplasm, IM: inner membrane, PE: periplasm, OM: outer membrane, CS: cell
821 surface, EM: extracellular milieu.

822 **Figure 2 Secretome based analysis of the proteins identified in the different extracytoplasmic** 823 **subproteomes of *E. coli* O157:H7 EDL933**

824 Schematic representation of the secretion routes for the proteins identified in the
825 proteosurfaceomes, proteovesiculomes and exoproteomes (Figure 1 and supplementary material

826 Table S1). The different categories of proteins are depicted in red and the number of proteins
827 identified in the proteosurfaceome, proteovesiculome and exoproteome is indicated in black, grey
828 and white, respectively. Components of the different protein transport systems were identified in
829 *E. coli* O157:H7 EDL933 and reported with Uniprot number; export systems are indicated in
830 violet, insertases in yellow, posttranslocational maturation pathway in green, and protein secretion
831 systems in blue.

832 Sec: secretion, SPase I: signal peptidase of type I, SPase II: lipoprotein maturation pathway
833 involving the signal peptidase of type II, Lol: localisation of lipoproteins, Bam: β -barrel assembly
834 machinery, T2SS: Type II secretion system, T3aSS: Type III, subtype a, secretion system, T3bSS:
835 Type III, subtype b, secretion system, T5aSS: Type V, subtype a, secretion system, T5eSS: Type
836 V, subtype e, secretion system, T7SS: Type VII secretion system, NCS: non-classical secretion,
837 CPs: cytoproteins, IMPs: inner membrane proteins, IMLs: inner membrane lipoproteins, PPs:
838 periplasmic proteins, OMLs: outer membrane lipoproteins, OMPs: outer membrane proteins,
839 SOMLs: surface outer membrane lipoproteins, SCSAs: subunits of cell surface appendages, EP:
840 exoproteins, ECPs: extracytoplasmic cytoproteins, CY: cytoplasm, IM: inner membrane, PE:
841 periplasm, OM: outer membrane, CS: cell surface, EM: extracellular milieu.

842 **Figure 3 Heat maps for differential proteins in the proteosurfaceomes, proteovesiculomes**
843 **and exoproteomes of *E. coli* O157:H7 EDL933 grown in M9, DMEM and BHI**

844 Shades of blue represent the relative abundance of variant proteins between the distinct growth
845 media and for each subproteome. Dark blue indicates higher protein abundance in the given
846 conditions compare to the other two, and *vice versa* for light blue.

847 **Figure 4 Distribution of common and differential proteins identified in the different**
848 **extracytoplasmic subproteomes of *E. coli* O157:H7 EDL933 respective to the predicted SCL**
849 **and biological process**

850 **A.** Percentage of common and differential proteins identified in the proteosurfaceomes,
851 proteovesiculomes and exoproteomes respective to predicted SCL according the gene ontology
852 (GO) terms for cellular component. Pink shades indicate proteins predicted as extracytoplasmic
853 (ECY), *i.e.* are either located at the cell envelope (GO: 0030313), namely at the IM
854 (GO: 0005886), in the periplasm (PE; GO: 0042597), at the OM (GO: 0019867), or at the cell
855 surface (CS; GO: 0009986), or in the extracellular milieu (EM; GO: 0005576); it further includes
856 proteins primarily predicted as cytoplasmic but also as present in the extracellular milieu
857 (EM/CY). Violet colour stands for proteins solely predicted as cytoplasmic (CY; GO: 0005737).

858 **B.** Percentage of common and differential proteins identified in the proteosurfaceomes,
859 proteovesiculomes and exoproteomes respective to the GO terms for molecular function. Identified
860 molecular functions terms included (i) GO: 0005488 related to binding with the selective, non-
861 covalent, often stoichiometric, interaction of a molecule with one or more specific sites on another
862 molecule, (ii) GO: 0005215 related to transport with the directed movement of substances (such
863 as macromolecules, small molecules, ions) into, out of or within a cell, or between cells, (iii) GO:
864 0003824 related to catalytic activity with catalysis of a biochemical reaction at physiological
865 temperatures, and (iv) GO: 0005198 related to structural activity with the action of a molecule that
866 contributes to the structural integrity of a complex or its assembly within or outside a cell.

867 **C.** Percentage of common and differential proteins identified in the proteosurfaceomes,
868 proteovesiculomes and exoproteomes respective to the GO terms for biological process. Identified
869 biological process terms included (i) GO: 0071840 related to cellular organisation with

870 biosynthesis of constituent macromolecules, assembly, arrangement of constituent parts, or
871 disassembly of a cellular component, (ii) GO:0051704 related to organism interaction with the
872 involvement of another organism of the same or different species, (iii) GO:0022610 related to
873 biological adhesion with the attachment of a cell or organism to a substrate, another cell, or other
874 organism, (iv) GO: 0065007 related to regulation process with any process that modulates a
875 measurable attribute of any biological process, quality or function, (v) GO: 0008152 related to
876 metabolic process with the chemical reactions and pathways, including anabolism and catabolism,
877 by which living organisms transform chemical substances, (vi) GO: 0050896 related to response
878 to stimulus with any process that results in a change in state or activity of a cell, *e.g.* in terms of
879 gene expression, protein synthesis, secretion or movement, as a result of a detected stimulus, (vii)
880 GO: 0051179 related to molecular localisation with any process in which a substance or molecule
881 is transported, tethered to or otherwise maintained in a specific location and (viii) GO: 0009987
882 related to cellular process with any process that is carried out at the cellular level. Green bars stand
883 for proteosurfaceome, blue bars for proteovesiculome and orange bars for exoproteome.
884 Enrichment analysis was performed with Panther to the ancestor level 2 of GO [51].

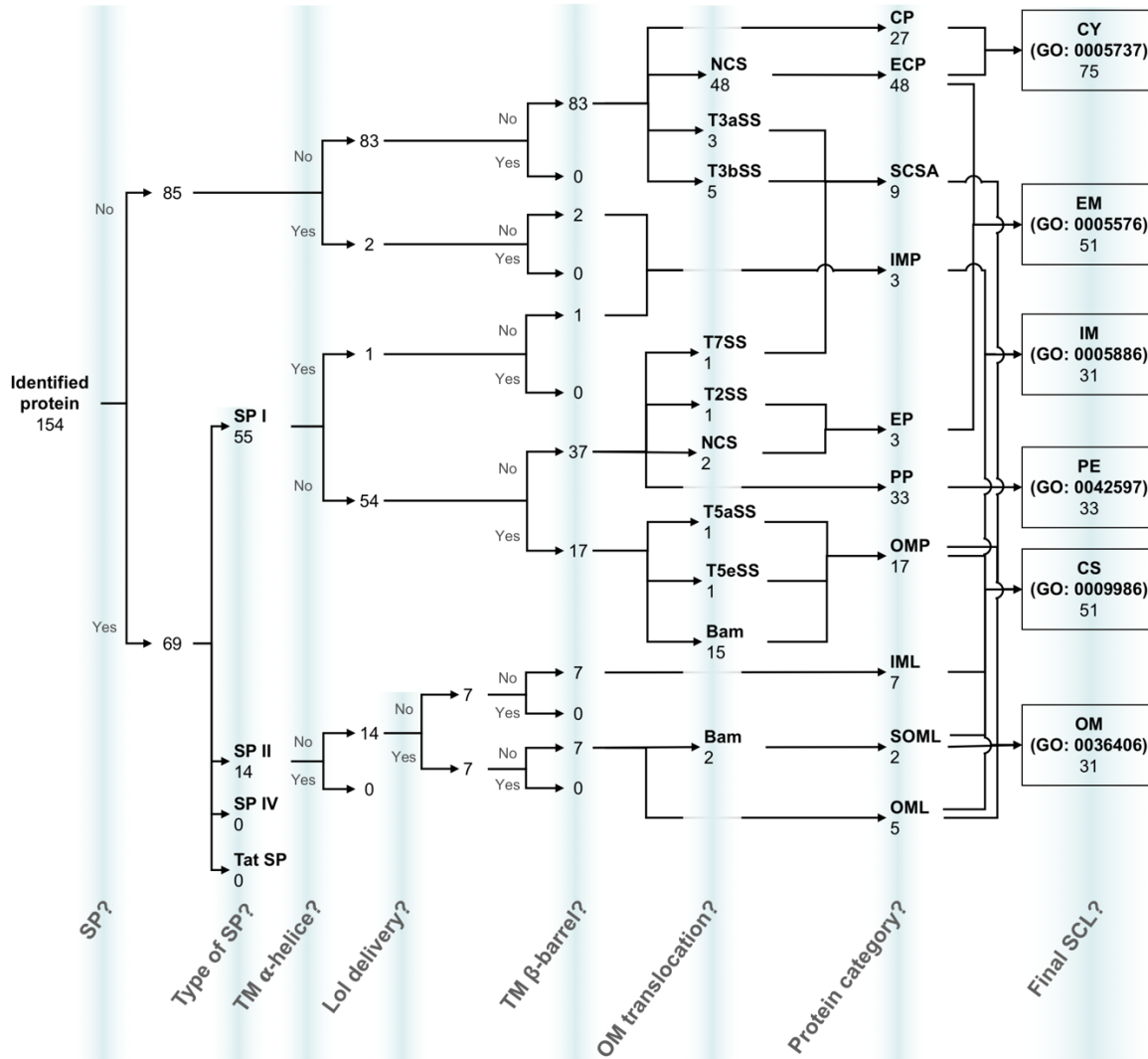
885 **Figure 5: Protein-protein interaction network of differential proteins and common proteins**
886 **identified in the proteosurfaceomes, proteovesiculomes and exoproteomes of *E. coli* O157:H7**
887 **EDL933 grown in M9, DMEM and BHI**

888 Interaction network for the variant subproteomes (A) and core subproteomes (B) with colour
889 shades indicating higher protein abundance in DMEM (red), BHI (green) and M9 (grey) for the
890 proteosurfaceome, proteovesiculome and exoproteome. Each network node (sphere) represents a
891 protein produced by a single protein-coding gene locus. Edges (connecting lines) represent protein-
892 protein associations, which are meant to be specific and meaningful, *i.e.* proteins that jointly

893 contribute to a shared function without necessarily meaning they are physically binding one with
894 another.

895

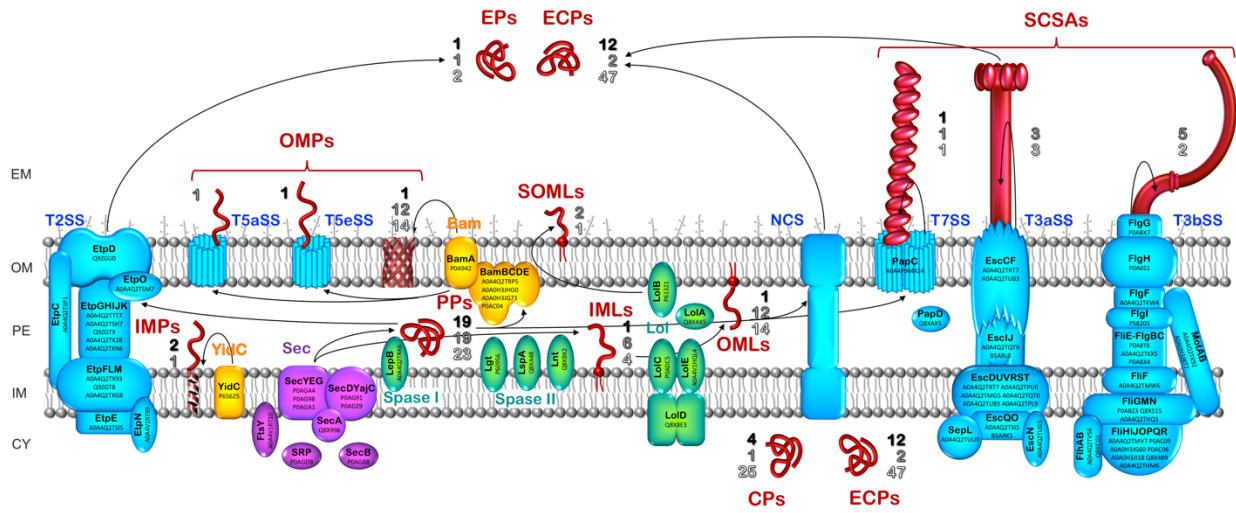
PRE-PRINT



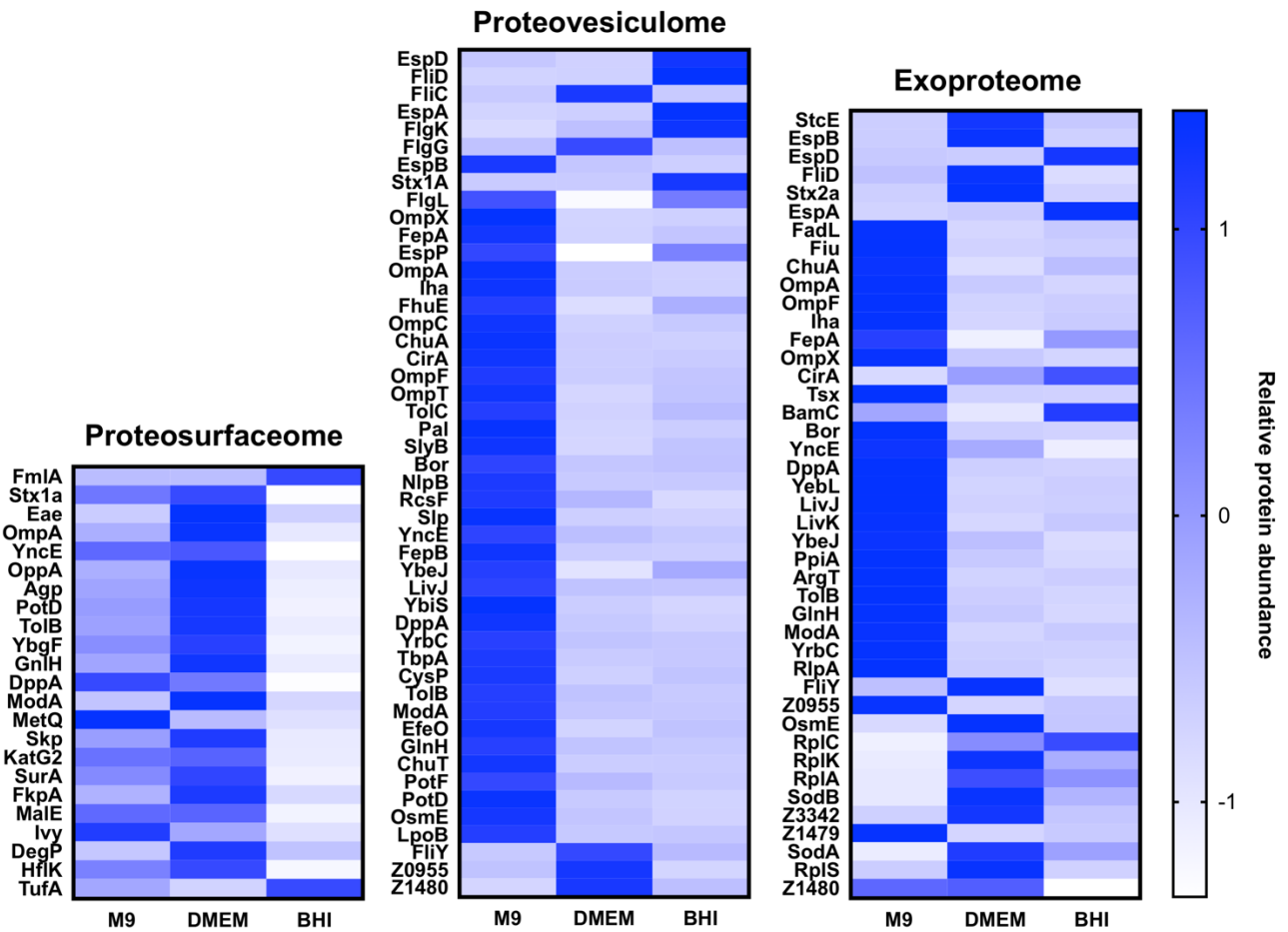
896

897 **Figure 1**

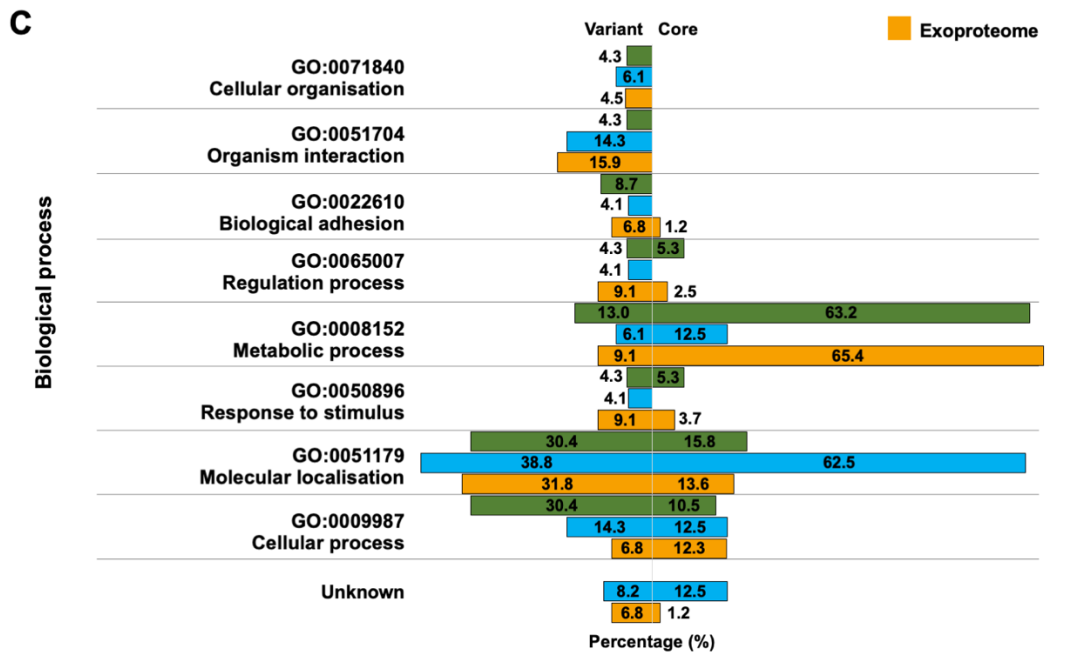
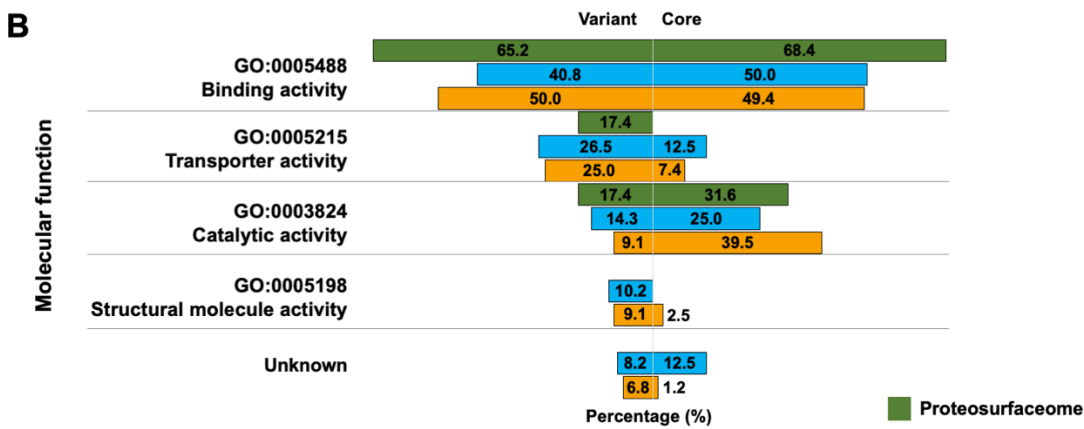
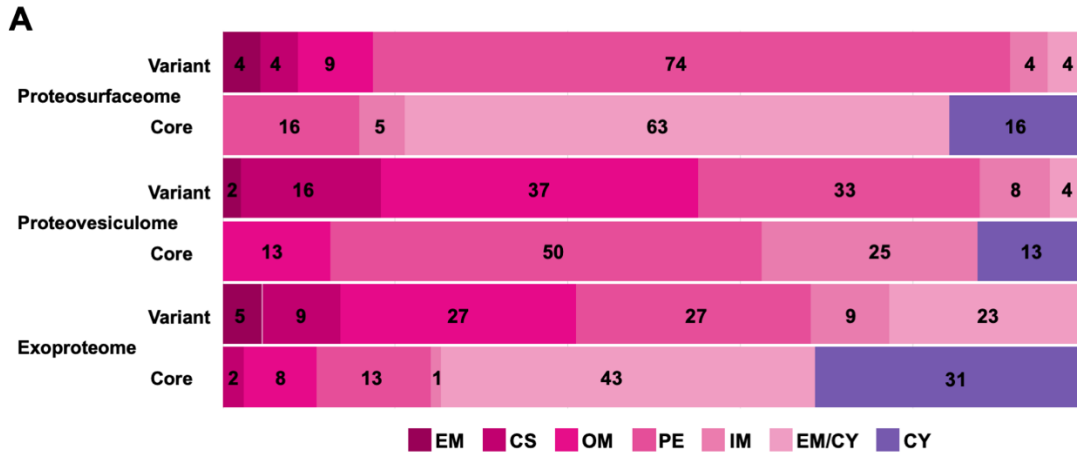
PRL



900 Figure 2

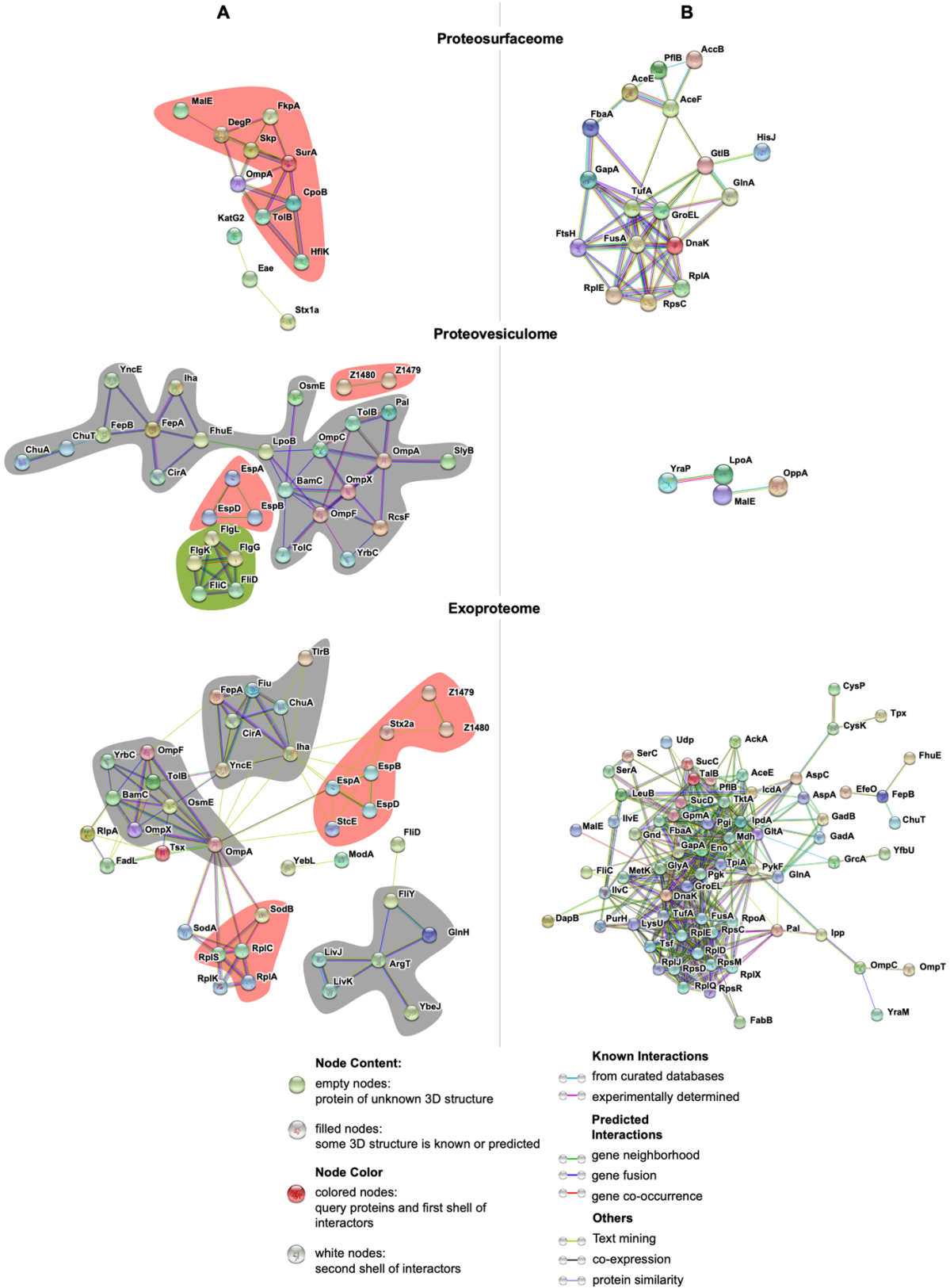


902 Figure 3



903

904 **Figure 4**



905

906 **Figure 5**