



AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Functional pangenome analysis reveals high virulence plasticity of Aliarcobacter butzleri and affinity to human mucus

This is the author's manuscript Original Citation: Availability: This version is available http://hdl.handle.net/2318/1795647 since 2021-08-02T12:13:18Z Published version: DOI:10.1016/j.ygeno.2021.05.001 Terms of use: Open Access Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use

of all other works requires consent of the right holder (author or publisher) if not exempted from copyright

(Article begins on next page)

protection by the applicable law.

| 1 | Functional pangenome analysis reveals high virulence plasticity of Aliarcobacter butzleri and |
|--------|---|
| 2 | affinity to human mucus |
| 3 | |
| 4 | Davide Buzzanca ^{1,2} , Cristian Botta ¹ , Ilario Ferrocino ¹ , Valentina Alessandria ¹ , Kurt Houf ² , Kalliopi |
| 5 | Rantsiou ^{1*} |
| 6 | |
| 7 | ¹ Department of Agricultural, Forest and Food Sciences (DISAFA), University of Turin, Italy |
| 8 9 | ² Department of Veterinary Public Health, Faculty of Veterinary Medicine, Ghent University, Belgium |
| 10 | |
| 11 | |
| 12 | *Corresponding author: <u>kalliopi.rantsiou@unito.it</u> , Largo Paolo Braccini 2, 10095 Grugliasco, Italy, |
| 13 | 0039-011-6708870 |
| 14 | |
| 15 | |
| 16 | |
| 17 | Keywords: Aliarcobacter butzleri, virulence genes, mucus, comparative genomics, pangenome. |
| 18 | |
| 19 | |
| 20 | |
| 21 | |
| 22 | |
| 23 | |
| 24 | |
| 25 | |
| 26 | |

27 ABSTRACT

Aliarcobacter butzleri is an emerging pathogen that may cause enteritis in humans, however, the incidence of disease caused by this member of the *Campylobacteriaceae* family is still underestimated. Furthermore, little is known about the precise virulence mechanism and behavior during infection. Therefore, in the present study, through complementary use of comparative genomics and physiological tests on human gut models, we sought to elucidate the genetic background of a set of 32 *A. butzleri* strains of diverse origin and to explore the correlation with the ability to colonize and invade human intestinal cells *in vitro*.

The simulated infection of human intestinal models showed a higher colonization rate in presence of 35 36 mucus-producing cells. For some strains, human mucus significantly improved the resistance to physical removal from the in vitro mucosa, while short time-frame growth was even observed. 37 Pangenome analysis highlighted a hypervariable accessory genome, not strictly correlated to the 38 39 isolation source. Likewise, the strain phylogeny was unrelated to their shared origin, despite a certain degree of segregation was observed among strains isolated from different segments of the intestinal 40 tract of pigs. The putative virulence genes detected in all strains were mostly encompassed in the 41 accessory fraction of the pangenome. The LPS biosynthesis and in particular the chain glycosylation 42 43 of the O-antigen is harbored in a region of high plasticity of the pangenome, which would indicate 44 frequent horizontal gene transfer phenomena, as well as the involvement of this hypervariable structure in the adaptive behavior and sympatric evolution of A. butzleri. 45

Results of the present study deepen the current knowledge on *A. butzleri* pangenome by extending
the pool of genes regarded as virulence markers and provide bases to develop new diagnostic
approaches for the detection of those strains with a higher virulence potential.

- 50
- 51
- 52

53

54 **1. INTRODUCTION**

Aliarcobacter butzleri (Basonym: *Arcobacter butzleri*) is a Gram negative bacterium belonging to the *Campylobacteriaceae* family often isolated from human stool, animal feces, drinking water, and food [1,2]. It is the most widespread species within the genus *Aliarcobacter* and is considered as an emerging pathogen, transmissible from livestock through the food of animal origin [1,3,4]. In this frame, *A. butzleri* has been isolated from healthy pigs, indicating a possible direct and indirect (crosscontamination dynamics) source of infection mediated by pork [5,6]. Spreading of this pathogen along the food chain is favored by its capability to survive in cold environments [7].

62 A. butzleri pathogenesis for humans is recognized but the underlying mechanisms are still largely unknown [1,8]. In vitro tests, using human cell line models have been employed to simulate adhesion 63 and invasion and infer the virulence potential of strains [9]. Although this approach is a simplification 64 of gut systems, it remains fundamental in the phenotypic investigation of host-pathogen interaction 65 [10]. In this context, the intestinal mucus appears to be relevant and may influence the ability of A. 66 67 butzleri to adhere and invade [11]. The mucus is composed mainly of glycoproteins and is present in different organs such as the stomach and gut. The number of mucus layers is variable along the 68 intestinal tract; in the small intestine and in the colon are present one and two mucus layers, 69 70 respectively [12]. The presence of mucus on the gut tissue is an important factor that has been shown to influence the development and behavior of intestinal bacteria [13]. 71

Survey studies have been performed to isolate *Aliarcobacter* spp. from different environments, animals and foods [1,8]. Isolates so far have been mainly genetically characterized for their virulence potential, focusing essentially on the presence of putative virulence genes that have been identified based on sequence similarity to other pathogens but without a biological confirmation of their role in pathogenicity [14,15]. The objective of this study was to characterize 32 *A. butzleri* strains, selected based on their source
origin, by combining Whole Genome Sequencing (WGS) and physiological data of colonization and
invasion of Caco-2 (*Homo sapiens*, Caucasian colon adenocarcinoma), and HT29 MTX (*H. sapiens*,
Caucasian colon adenocarcinoma treated with methotrexate), a mucus producer cell line.

81 2. RESULTS AND DISCUSSION

82 2.1 SIMULATED INTESTINAL COLONIZATION IS ENHANCED BY HUMAN MUCUS

Thirty-two A. butzleri strains previously collected from human and animal feces, pig intestine, animal 83 skin and meat (Table 1) were tested on human gut models to define their capability to colonize (cell 84 association) and invade intestinal cells. More specifically, the mixed culture of Caco-2/HT29-MTX 85 cells and Caco-2 were used as mucus producing (MP) and not-mucus producing (NMP) models, 86 respectively. All A. butzleri strains were able to colonize both models after 90 minutes of co-87 incubation, while strains 7 (isolated from bovine), 26, 28, 31 (isolated from Human feces) presented 88 a not detectable invasion on the MP models; moreover, invasion by strain 28 was not detectable also 89 90 on NMP models (Fig. 1). However, only a minor part of the bacterial cells colonizing the models $(0.64 \pm 0.34 \%$ on average) invaded the human cells, corresponding to an average decrease of $3.7 \pm$ 91 1.5 Log CFU/cm² from the initial inoculum (Supplementary Table 1). The multiplicity of infection 92 93 (MOI) was not the same for all strains. Although this parameter has been shown previously to play a 94 role in the transepithelial resistance of cell lines after 48 hours of contact with A. butzleri [16], it 95 appears not to have an effect during adhesion studies, conducted under short term (1-3 hours) of bacterial contact [17,18]. These Overall, our data are in agreement with previous reports and confirm 96 the ability of A. butzleri to colonize different cell lines with an invasion efficiency similar to the 97 phylogenetically close species of Campylobacter jejuni [19-22]. 98

99 Comparing colonization data (expressed as ΔLog CFU/cm²) from MP and NMP models, *A. butzleri*100 showed an overall greater (*P* < 0.001) colonization capability in presence of human mucus (Fig. 1A).
101 The presence of the human mucus glycoproteins enhanced the colonization capability of all strains,

but significantly (P < 0.001) only for three isolates from pig intestine (strains 16, 17, 19). Other than 102 103 that, no relationship between the two main sources of isolation (human stool and pig intestine in its various sections) and the colonization trend was observed (Fig. 1B). Strains from these two sources 104 highlighted an equal proportion of highly colonizing (positive or close to zero ΔLog) and low 105 colonizing phenotypes (negative ΔLog), regardless of the model used. Positive values observed for 106 some strains suggest bacterial growth during the host-pathogen interaction timeframe, again more 107 108 evident in presence of mucus. Considering each model, all strains have shown comparable colonization abilities. Finally, considering differences between strains, strain 2 colonized statistically 109 more (P < 0.05) than strain 31 in the NMP model. 110

111 The effect of mucus in enhancing colonization has already been observed in Aliarcobacter butzleri 112 [19]. This is not surprising, since it is a hallmark of intestinal pathogens [13], which must overcome the mucus in order to exert the infection in the host [23]. In this frame, an in vivo survey suggested a 113 114 chemoattractant function of the mucus towards Aliarcobacter spp. since it was recovered not only from the inner content but also from the mucus layer of pig intestines [5]. The statistically significant 115 higher mucus-model colonization observed for part of the pig isolates also suggests a rather strain-116 dependent mucus affinity that may result in exploitation of its protective action against intestinal 117 118 peristalsis under in vivo conditions.

119 2.2 FUNCTIONAL CHARACTERIZATION OF PUTATIVE ENCODED PROTEOMES

All *A. butzleri* strains were *de novo* sequenced, assembled and subjected to whole-genome comparative analysis. Genomes obtained display a GC content between 26.74 and 27.11 % and a length ranging from 2.04 Mb to 2.50 Mb. We observed several genes/proteins belonging to incomplete (< 60 % of similarity) and questionable (< 90 %) prophage regions, but no intact known prophage region was found in the 32 genomes. Clustered regularly interspaced short palindromic repeat (CRISPR) sequences are present in 23 genomes, of which only ten CRISPR regions are flanked to CRISPR-associated protein (CAS; general class 1 and 2) sequences. Always concerning mobile
genetic elements and signatures of bacteriophages, at least one transposase gene was found in 27 of
the 32 genomes (Supplementary Table 2). Importantly, the presence of numerous mobile genetic
elements are markers of a former evolution and potentially improved fitness, which is important for
any pathogen [24].

Classes of COG (Clusters of Orthologous Groups) are homogeneously distributed along the 32 131 132 putative encoded proteomes (Fig. 2A, Supplementary figure 1). We observed a remarkable abundance (19.4 % on average) of proteins with unknown function or characterized only for general 133 functions, a fact that highlights the limited characterization of A. butzleri proteome [8]. Following, 134 135 signal transduction mechanisms (average of 10.14%) is the most abundant characterized COG class, 136 suggesting the presence of an extended network of control of A. butzleri functions [14,15,25]. Predicted proteins involved in the metabolism and transport of amino acids are significantly more 137 138 abundant (9.16%) than those related to carbohydrate if compared to the COG distribution observed 139 in other bacteria [26]. This is consistent with the limited or null consumption of carbohydrates shown by A. butzleri and other Aliarcobacter species, which instead utilize organic acids and amino acids as 140 main carbon sources [14,25]. Moreover, the classes of signal transduction mechanism and cell 141 142 motility represent more than 10 % of the predicted proteins, whereas only 14 genomes 143 (Supplementary Figure 1B) harbor genes involved in bacterial cytoskeleton function (COG class Z), which are linked to bacterial motility too [27]. It is noteworthy that proteins involved in motility 144 play a pivotal role in the host/pathogen interaction since the bacterial movement in the gel-like matrix 145 146 such as the shallow mucus layer can allow a faster pathogen infection [28].

147

148

149 2.3 GENOME-WIDE ANALYSIS SHOWS AN OPEN PANGENOME

150 The core- and accessory-genomes sizes were estimated (**Table 2**) by clustering the predicted 151 aminoacidic sequences of the 32 annotated genomes through two pangenome computing programs

[29,30]. In both cases the accessory genome resulted to represent from 78% to 75% of the pangenome, 152 comprising most of the hypothetical proteins (up to 90 %) and composed of more than 55 % of 153 singletons (gene family exclusively present in one genome). Similar partitioning of A. butzleri 154 pangenome has recently been observed on a set of 49 genomes [15]. This leads us to speculate a wide 155 156 and open pangenome, which reflects a sympatric evolution with frequent episodes of horizontal gene transfer (HGT), like the exchange of genes involved in pathogenesis and antibiotic resistance, that 157 158 can confer an adaptive advantage in changing environments [31]. As done for other foodborne pathogens [32,33], with the increasing number of available genomes a large-scale pangenomic survey 159 will be needed soon to confirm these first observations. 160

161 The number of genomes here investigated is adequate to infer the structural organization of the 162 pangenome with Markov Random Field networks (Fig. 3A), which display the localization of each gene family (nodes) by following a pattern of continuity (edges connect loci that are frequently 163 164 neighbors) regardless from contigs succession [30]. As previously observed during the validation of this approach on a large set of Acinetobacter baumannii genomes [30], the pangenome of our 32 165 strains shows organized clusters of persistent gene families (present in more than 95 % of the strains) 166 either surrounded or, less frequently, interrupted by islands of dispensable genes (shell and cloud 167 genome). It is noteworthy the presence of a large pangenome plasticity island that represents hotspots 168 169 of alternative structural organizations along the genomes analyzed, thus possible sites of HGT (Fig. **3B**). In addition to a predictable presence of hypothetical and not functionally characterized proteins, 170 this island encompasses several accessory gene families generally involved in the COG class of cell 171 172 wall/membrane and envelope biogenesis, besides proteins more specifically associated with pili/flagella glycosylation, LPS glycosylation/assembling and exopolysaccharide (capsule) secretion 173 (Fig. 3C). Additionally, we detect up to 323 regions of genome plasticity (RGP) that can be referred 174 as genomic islands shared by at least two of the genomes [34,35]. As observed for the whole 175 176 pangenome, these RGP overall encompass gene families involved in the cell wall/membrane and 177 envelope biogenesis, which however represents the second class after genes involved in replication,

recombination, and repair of DNA (Fig. 3C). The presence of genes involved in the biosynthesis of 178 179 capsular polysaccharide, lipopolysaccharide (LPS) glycosylation and flagellin/pilin glycosylation within the island of pangenome plasticity and the RGPs is not surprising, since are accounted as 180 dispensable genetic structures that can be acquired or lost, to face host-to-host transition and colonize 181 new ecological niches [36,37]. For instance, loss of the flagellin glycosylation genes may determine 182 phenotypic changes that decrease recognition of strains by the host-immune system [38], while the 183 184 polysaccharide chain of LPS (inserted between hydrophobic lipid and hydrophilic O-antigen) possess hypervariable structures that reflect the specific pathogen serological signature [39]. 185

Other genomic regions were entirely constituted by singletons and in some genomes (strains 18, 17, 186 187 and 3) represent large sections of it (up to 20,000 bp), containing hypothetical proteins, mobile genetic elements (phage proteins, transposases, recombinase) and genes with poorly defined functions. 188 Interestingly, few or no singletons are found in pig duodenum/caecum isolates and 8 of the 9 genomes 189 190 from pig rectum (Supplementary Figure 2). On the contrary, by excluding the singletons we observed that 425 and 140 accessory gene families are significantly (Scoary statistics; P [FDR] < 191 0.05) overrepresented in the genomes of pig rectum and pig duodenum/caecum isolates, respectively. 192 Besides, no overrepresentation of gene families is observed in each of two major ecological sources 193 194 of isolation, i.e. human stool and all pig intestine. Accordingly, the phylogenetic trees (Fig. 4) do not 195 show a clear segregation between these two groups, but regardless of the type of input sequences (i.e. 196 the whole genomes, core genomes, SNPs, and MLST loci) a recurrent clustering pattern that consists of group I (strains 14 and 15, from duodenum and caecum of pig), group II (human strains 1 and 28), 197 198 group III (strains 12, 19, 20, 21 from pig) and group IV (strains 11, 13, 23 from pig rectum) was observed. 199

Considering their genomic plasticity (high level of intra-group shared genes, low or absent singletons)
 and phylogenetic analysis, the isolates from pig duodenum/caecum (group I) and rectum (group III e
 IV) represent three distinct lineages. This aspect also suggests that distinct genotypes of *Aliarcobacter butzleri* may colonize specific segments of pig intestine, as already observed at the species level for

Aliarcobacter spp. [5]. Moreover, the low genomic plasticity of these three groups and the fact that 204 205 pig intestine (particularly the rectum section) can be a favorable niche for this pathogen with limited 206 or absent symptoms in the host [40], lead us to speculate host and/or tissue tropism phenomena for these three groups [41]. On the contrary, the remaining strains seem to have undergone more episodes 207 208 of HGT, likely in reason of frequent host transition events and developed a more host-generalist genotype [42]. In light of our pangenomic observations, A. butzleri may represent a pathogen with 209 210 both host-generalist and host-specialist phenotypes, which can alternately arise in livestock in response to external selective pressures (antibiotics, intensive breeding) and then transmitted to 211 humans, as recently reported at large-scale for *Campylobacter* spp. [38,41]. 212

213

214 2.4 REPERTOIRE OF VIRULENCE GENETIC TRAITS

To detect possible genomic signatures linked to A. butzleri virulent phenotypes we manually curated 215 216 the annotated genes by focusing on those sequences putatively associated with host-pathogen interaction in this and other pathogenic bacteria (Supplementary Table 3). This produced a list of 217 100 genes (of which 39 are accessory genes) putatively involved in functions related to human 218 mucosa adhesion/invasion, interaction with host mucosa/mucus, flagellum and motility, as well as 219 220 proteins more widely correlated to virulence of A. butzleri and other pathogens, such as hemolysis, 221 secretion and regulatory systems (Fig 5). It is noteworthy that the 32 presence/absence profiles (Pearson's correlation-based dendrogram) cluster as previously observed in the whole-genome 222 phylogenetic dendrogram and almost all these genes are included in accessory gene families of the 223 224 pangenome. Accordingly, we might speculate that the biodiversity within A. butzleri populations are partly shaped from the exchangeable virulome as a genomic tracking of the host-to-host transitions 225 undergone by each strain. Nevertheless, strains origin and other genes not directly involved in 226 virulence mechanisms may play an important role in the phylogenetic segregation of the strains. 227

228

229 2.4.1 GENES COMMONLY RECOGNIZED AS VIRULENCE FACTORS

As first step, we focused on ten genes commonly employed as markers to assay the virulence potential of the *Aliarcobacter* genus [20,43]. Genes correlated to adhesion (*cadF*), invasion (*cj1349*, *ciaB*) and hemolysis (*pldA*, *tlyA*, *mviN*) are present in all the genomes but, except for *pldA* and *tlyA*, were initially annotated as different proteins.

234 On the other hand, the gene *hecA* (hemagglutinin alternatively annotated as *shlA* or *hpmA*) and *hecB* (hemolysin activation protein here annotated as *shlB*) are present and adjacent to each other in 31 % 235 236 of the genomes, while *iroE* (encoding a periplasmic enzyme and annotated as *besA*) and the generic virulence factor *irgA* were found in 75 % and 81 % of the genomes, respectively. According to PCR 237 based studies and other genomic surveys [44,45], these latter four genes are less prevalent across the 238 239 whole A. butzleri population. Moreover, the presence of these four genes in our strains do not significantly correlate (Pearson's moment correlation, P [FDR adjusted] > 0.05) with their 240 colonization phenotypes. This is not surprising since they encode for functions useful in following 241 242 infection phases [46]. Regardless of their impact on the colonization, the initial misannotation observed for most of these genes underlines once more their high polymorphism, which often leads 243 to underestimating virulence potential and diffusion of A. butzleri due to false negatives 244 amplifications [14,15]. 245

246

247 2.4.2 GENES RELATED TO ADHESION AND INVASION

An important gene involved in the colonization, specifically in the host mucosa adhesion, is *porA* that encodes for a major outer membrane protein responsible for the hypervirulence of *C. jejuni* [47]. Here it was found in all genomes, properly annotated or indicated as putative gene for *Campylobacter* major outer membrane protein. In *A. butzleri* the high polymorphism of this gene and its flanking regions have been recently proposed as a meaningful signature of pathogenicity, not related to the shared ecological origin and whole genome phylogeny [15]. The UPGMA dendrogram of the 32 *porA* sequences (**Supplementary Figure 3**) partially confirms the previous observations, with a grouping pattern unrelated to the initially shared origin, but instead resembling the phylogenetic segregation
previously described (for instance the recurrent groups I, II, III, IV).

Another ubiquitous gene is the *inlJ*, which encodes in *Listeria monocytogenes* for a protein of the LPXTG-internalin family and is involved in host adhesion and invasion [48]. However, other orthologues of the *Listeria monocytogenes* internalin operon are missing in the genomes of all but strain 29 that encompass the internalin A in a different genomic region. The absence of internalin orthologs seems to correlate well with the aforementioned limited invasiveness of *A. butzleri* when compared to *Listeria monocytogenes* [49].

As here phenotypically confirmed, *A. butzleri* can penetrate and likely move throughout the human mucus (**Fig. 1A**). The mucus, having a protective function towards the intestinal epithelium (in our case the cell model layer), must be overcome to allow colonization, capacity observed in the 32 *A*. *butzleri* strains object of study (**Fig. 1B**) [12]. However, only one gene (encoding an Arylesterase precursor) linked to mucus degradation was detected. Two different Arylesterase forms are present in the genomes, not correlated with greater or lower colonization in presence of mucus (*P* [FDR] > 0.05) [50].

- 270
- 271

272 2.4.3 SECRETION SYSTEMS INVOLVED IN PATHOGENICITY

Several genes of our proposed virulome are part of secretion systems and can play a role in the hostpathogen interaction. Among these, the operon encompassing genes *epsE/epsF* and the *xcpO* gene
are part of a type II secretion pathway fundamental in the infection mechanism of *Vibrio cholerae*[51], albeit numerous components of the original operon are missing in *A. butzleri* genomes.
Similarly, genes (*epsD*, *epsH epsM*, *epsN*, *epsJ*) responsible for exopolysaccharide secretion and
biofilm-forming capability in *E. coli* are present, but not organized in a single operon [52]. Some
molecules linked to biofilm production can promote bacterial adhesion on human intestinal cells [53].

In this light and considering that *A. butzleri* is proven to form biofilm [54], further investigation to define its *eps* genes role and regulation is now needed.

Moreover, six genes (*virB10*, *virB8*, *virB6*, *virB4*, *virB3*, *virB*), encoding a rare T4SS structure (type IV secretion system), recently described in *A. butzleri* [15], were annotated in strain 18. Differently from the previous observation, these genes are not comprised in a single genomic region but are instead spread along with the several islands of singletons found in this genome. The T4SSs are an important virulence mediator in different pathogens, including *C. jejuni*, since are connected to host cell apoptosis, cytotoxicity, bacterial cell survival, adhesion, and invasion to host cell [55–57]. Anyhow, this peculiarity did not result in a greater colonization or invasion activity for this strain.

289

290 2.4.4 GENOMIC SIGNATURES RECOGNIZED BY HOST IMMUNE RESPONSE

A consistent fraction of putative virulence genes are involved in the flagellar assembling/motility 291 292 (flagellins), chemotaxis and urease activity (indirectly responsible for increase of external pH), mostly organized in clusters or anyway located in the same genomic regions [14,15]. In particular, 293 the flagellar proteins are important virulence factors related to human pathogens motility in the 294 proximal mucus layer and recognition by host immune response [13]. Thus, it is intriguing that gene 295 296 families encoding flagellins are included in the core genome of A. butzleri, while we found genes 297 responsible for their glycosylation in the accessory and plastic genomic regions. This suggests the 298 heterogeneous glycans compositions of flagellum may lead to a strain-specific antigenic fraction of this bacterial component [58]. 299

LPS O-antigen plays a pivotal role in the pathogen survival on the human mucosa, modulating host
immune response and counteracting its defense mechanisms [11].

All 32 genomes contain at least one copy of the O-antigen ligase gene, of which polymorphism
follows the whole-genome phylogeny (formerly groups I, II, III, IV) and goes hand by hand with the
structural organization of the surrounding genes (Supplementary Figure 4, Supplementary Table
Interestingly, in 32 genomes we observed up to 25 different genomic structures flanked to O-

antigen ligase that encompass genes involved in LPS O-antigen assembling [59], such as lipid A 306 307 biosynthesis protein (msbB), LPS transferases (rfaC),sugars/glycosides transferases/epimerase/reductase (rfaF, sunS, pglJ, lacA, epsJ, rfbB, rfbC, rmlD, kfoC). The lack of a 308 single component of the O-antigen genes cluster (ABC transporters, glycotransferase, 309 glycosyltransferase) can dramatically affect the infectiveness of Gram-negative pathogens [60,61]. 310 Therefore, the role of such variability regarding the genes flanking the O-antigen ligase genes in 311 312 pathogenicity deserves further investigation. Indeed, the intraspecific complexity of the O-antigen pathway, already noticed in four A. butzleri genomes [62], and here confirmed by a large scale 313 genomic comparison, highlights this region of the plastic virulome as one of the most useful to define 314 315 strain-specific virulence signatures.

316

317 2.4.5 GENES INVOLVED IN MULTIPLE VIRULENCE MECHANISMS AND REGULATION

Other meaningful elements of the *A. butzleri* virulome (**Fig. 5**) are represented by membrane components, like TonB transport protein (different protein forms and domains) and the transport complex ExbB/ExbD, which are required for *Shigella dysenteriae* and *E. coli* invasion/spread in human cells [63,64]. Invasion ability shown here and, even more, the capability of *A. butzleri* to cause septicemia by spreading in human cells may suggest similar functions of these genes in *A. butzleri* [1]. Moreover, TonB is involved in the iron uptake as *irgA* [65], by suggesting its possible role in hemolysis [66].

Particularly relevant is the presence of *phoP* and *phoQ* genes (respectively encoding the transcriptional regulatory protein PhoP and sensor protein PhoQ), which constitute a two-component signal transduction system able to regulate intracellular virulence, cell envelope composition, and the within-host lifestyle in Gram negative bacteria [67,68]. Twenty-two genomes contained at least one form of *phoP*, while *phoQ* was only found in eight genomes and not flanking the *phoP* gene. However, several genes encoding for proteins with potential homologous function to *phoP* or *phoQ* were found flanking or nearby the gene encoding for the respective complementary protein. For

instance, in all eight genomes, the phoQ gene is located next to the gene mprA encoding a 332 333 transcriptional factor. Interestingly, when this transcriptional factor is joined by the mprB gene (regulatory system *mprB/mprA*) the *Mycobacterium tuberculosis* infection increase in its persistence 334 [69]. Considering that *mprA* exerts a transcription regulation activity comparable to *phoP*, the 335 336 genomic continuity with *phoQ* suggests a possible homologous function. On the other hand, several sensor proteins with potential histidine kinase function homologous to phoQ are flanking the gene 337 338 phoP, such as the genes fixL, zraS, pdtaS (strain 25), and ttr (strain 29). Moreover, several strains harbor genes encoding phosphorelay sensor kinase activity that regulates PhoP-PhoQ in other 339 bacteria, such as the virulence sensor bvgS [70] or the couple of genes dsbA/dsbB, here annotated as 340 341 DSBA-like thioredoxin domain protein and thiol-disulfide oxidoreductase resA, respectively. 342 Particularly, this latter two-component system activates the *phoP* gene in E. coli [71].

In terms of virulence phenotypes, we did not observe a significant correlation (P [FDR] > 0.05) between the colonization/invasion data and the genomic occurrences of *phoP/phoQ* or the alternative two-component system above described. Nevertheless, the impact of this signal transduction system on the pathogen phenotype is dependent also on upstream regulators/activators and downstream triggered virulence genes [72], which in *A. butzleri* need to be characterized with future transcriptomic investigations.

349 **3. CONCLUSION**

The attention of the scientific community towards *A. butzleri* is significantly rising in the last years, with a parallel increase of concerns about its genomic flexibility, virulence predisposition in humans, adaptability to different hosts. In this frame, we focused our efforts on the first two issues by exploiting the pangenomic approach as an advanced comparative tool, integrating the genomic data with physiological tests on a set of strains tested with human gut models with and without mucus.

In summary, *A. butzleri* strains have shown a similar capability to colonize *in vitro* the human mucosa by adhering and even proliferating within human mucus, without showing marked invasiveness. Notwithstanding, it is not clear if a commensal lifestyle within mucus is conceivable in humans. In

pigs, asymptomatic infections suggest that it may have developed a host specialist lifestyle and such 358 359 hypothesis is supported by the genomic data of this study. In this context, the open pangenome and the interchangeability of potential virulome have been recently demonstrated and proposed as key 360 genomic features for the host adaptation of this pathogen. Here, also, to confirm these first findings, 361 we link the variable virulome to strains phenotypes, by identifying in the LPS assembling pathway 362 one potential strain-specific signature. Despite the intrinsic limit of pangenomic based comparison 363 364 that does not necessarily permit to exhaustively explain the multifaceted virulence mechanism of A. butzleri, we pointed out and described the presence of putative virulence promoters and antigen 365 recognition markers, such as master regulators. 366

367 Therefore, these outcomes will provide concrete guidelines for more comprehensive omics
368 investigation of the *A. butzleri* lifestyle in human mucosa.

369

370

4. MATERIALS AND METHODS

372 *4.1 BACTERIAL STRAINS*

The *A. butzleri* strains (**Table 1**) were obtained from the Belgian Coordinated Collection of Microorganisms (BCCM; Laboratory for Microbiology, Ghent University, Belgium) isolated from different sources, and stored in Laked Horse Blood (Oxoid, Basingstoke, Hampshire, UK) at -80 °C. Cultivation was performed in microaerophilic conditions at 30°C on agarized Arcobacter broth (CM0965, Oxoid) supplemented with C.A.T supplement (SR0174, Oxoid) [73].

Before each experiment, a single fresh colony was resuspended in Arcobacter broth and incubated at 30 °C for 48 hours. Afterward, 0.5 ml of culture was inoculated on Arcobacter plates supplemented with C.A.T supplement, grown for 48 hours in microaerobic conditions, collected with 1 ml of Ringer's solution (1.15525, Millipore, Burlington, Massachusetts, U.S.A) and thus used as working suspension in the interaction experiments. The bacterial load (Log CFU mL⁻¹) of each working suspension was determined by measuring OD at 630 nm with ELx880 microtiter plate reader
(Savatec, Turin, Italy) and set to the same initial count by an internal standard curve.

385

386 *4.2 CELL LINES AND HUMAN GUT MODELS*

Human colon carcinoma cell lines Caco-2 (86010202, ECACC, European Collection of 387 Authenticated Cell Cultures, Public Health England), HT29 (ATCC® HTB38, ECACC) and HT29 388 MTX (12040401, ECACC) were cultured in Dulbecco's Modified Eagle's Medium (DMEM 6429; 389 Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% of fetal bovine serum (FBS; 390 F7524 Sigma-Aldrich) and EmbryoMax Penicillin-Streptomycin Solution, 100X (TMS-AB2-C, 391 Sigma-Aldrich). The cell lines were grown in 25 and 75 cm² culture flasks (Corning, New York, New 392 393 York USA) at 37 °C in a humidified atmosphere containing 5% CO₂ and 95% air and sub passaged 394 every 3-4 days (Eppendorf, Galaxy 170 S, Hamburg, Germany) [74].

Two in vitro monolayer human epithelial structures were prepared: a mucus-secreting (MP) co-395 culture of differentiated Caco-2 and HT29-MTX cells in a 9/1 ratio; and two non-mucus-secreting 396 397 cell models (NMP) represented by a single culture of differentiated Caco-2 cells and a mixed model of Caco-2 and HT29 cells with the same ratio of MP model [75]. Briefly, the cells were seeded at a 398 density of 35,000 cells cm⁻² and grown in complete culture medium under the same conditions 399 described above, for 14–15 days with regular changes of the media, until functional polarization was 400 reached and models could be considered differentiated and ready for the experiments [76]. Before (3-401 402 4 days) the assessment of strains colonization and invasion capability, the MP and NMP models were washed twice with PBS 1X and the complete culture media was replaced with media without 403 antibiotics to allow the pathogen growth. 404

405

406 4.3 ASSESSMENT OF COLONIZATION AND INVASION CAPABILITY

The working suspensions of the strains were inoculated on MP and NMP cell models. Depending on 407 408 the growth capacity of the individual strains, different inoculum levels could be experimentally achieved; in the majority of cases the density of bacterial suspensions was 7-8 Log₁₀ CFU mL⁻¹ 409 (Supplementary Table 1). Due to this experimental limitation, the multiplicity of infection (MOI) 410 was not the same for all strains tested. Colonization-invasion assays were performed on two different 411 model wells for each biological replicate. After 90 minutes of co-incubation at 37 °C in a normal 412 413 atmosphere, the not adherent bacteria were removed by two washing steps with PBS 1X. Colonization and invasion capabilities were evaluated in parallel on MP and NMP models on at least three 414 biological replicates. 415

To quantify the colonization capability (also defined as cell association), which represents the pathogen ability to adhere and enter the human cells, *A. butzleri* cells were recovered from one duplicate of the cells model by incubating for 30 minutes with 1 mL cm² of 0.25 % Triton X-100 (v/v; in PBS 1X). Counts of the resulting suspension were performed employing the CFU method, plating the dilutions on solidified Arcobacter broth supplemented with C.A.T supplement for 48 h at 30 °C in microaerobic conditions.

In parallel, to define the invasion capability (number of bacterial cells that penetrate in the human cells excluding those adherents) the culture media supplemented with 300 μ g ml⁻¹ of gentamicin sulfate (G1914, Sigma-Aldrich) was added in the cell models for 120' at 37 °C to kill all the extracellular bacteria. After two washing steps with PBS, the internalized viable cells of *A. butzleri* were recovered and enumerated as described for total colonization [20,77,78].

427 Raw counts data were expressed as Log CFU cm⁻² of bacteria inoculated (T0), bacteria colonizing 428 the model after washing steps (Tc) and after gentamicin treatment (Ti). Colonization was expressed 429 as Δ Log CFU cm⁻², by following the formula: Log CFU mL⁻¹_{Tc}– Log CFU mL⁻¹_{T0}. Invasion 430 capability was expressed following the formula: Log CFU mL⁻¹_{Ti}– Log CFU mL⁻¹_{T0}.

432 *4.4 GENOME SEQUENCING, ANNOTATION AND BIOINFORMATIC ANALYSIS*

433 Genomic DNA (gDNA) extraction of A. butzleri strains was performed by the beads-beating, phenolchloroform DNA extraction method followed by a RNAse A (5 µg µl⁻¹, MRNA092 Epicenter, 434 Madison, Wisconsin, U.S.A) treatment to digest RNA in the DNA samples, with an incubation of 30 435 min at 37 °C. The DNA quantification was performed with the employment of Nanodrop (ND 1000, 436 Thermo SCIENTIFIC). The gDNA quality check, to confirm the absence of degradation and 437 438 impurity, was performed through an electrophoretic run (100 V for 30') on agarose gel 0.8% (w v⁻¹, 0710 VWR) in TAE 1X (Tris - Acetic acid - EDTA, K915 VWR), gelRed (41005, Biotium) was 439 used as DNA intercalating. 440

441 Whole genome sequencing (2X150bp, coverage 100X) was performed on Illumina Novaseq 6000 machine by the Novogene company (Cambridge, United Kingdom). Briefly, after a Qubit 2.0 442 quantification 1 µg of gDNA was used for the library preparation using NEBNext® library prep Kit, 443 444 randomly fragmented (350 pb) by shearing and then the samples were polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and PCR enriched by P5 and indexed P7 oligos. 445 The PCR product purification was performed with the use of the AMPure XP system, afterwards, the 446 libraries were analyzed by Agilent 2100 Bioanalyzer (size distribution) and quantified using real-time 447 PCR. 448

449 Sequencing reads were quality filtered with Solexa QA++ software, and sequences less than 60 bp450 and dereplicated sequences were removed by Prinseq.

451 Reads were *de novo* assembled with SPAdes (version 3.11.0) [79] and the quality of the contigs was 452 checked with QUAST software to obtain statistics related to the genomes assembly process and data 453 quality, such as coverage, total genome bp length and the number of contigs [80] (**Supplementary** 454 **Table 2**).

Genomes were annotated using the Prokka (version 1.11) suite [81] and putative encoded proteins have been manually checked through on UniProt (<u>https://www.uniprot.org/</u>), Pfam 457 (https://pfam.xfam.org/), and CDD database 458 (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to understand their functional role [82]. The CRISPR-CAS sequences have been detected with the software CRISPRCasFinder 1.1.2 459 (https://crisprcas.i2bc.paris-saclay.fr/), while phage sequences were retrieved with Phaster 460 (https://phaster.ca/) [83,84]. Additional analysis on the metabolic pathway was performed on the 461 putative predicted proteome with the software RPS-BLAST 2.2.15 on WebMga (http://weizhong-462 463 lab.ucsd.edu/webMGA/), to obtain the related COGs (Clusters of Orthologous Groups) codes and classes [85-87]. 464

Proteins inferred by Prokka were then processed with the parallel use of Roary (version 3.13.0) and 465 466 PPanGGOlin (version 1.1.85) with default parameters to generate the presence-absence binary matrices of core and accessory genes [29,30]. The structural settlement of the loci (gene families) in 467 the pangenome was inferred through the matrix generated by PPanGGOlin and visualized using the 468 469 program Gephi 0.9.2-beta (https://gephi.org). The presence of regions of genome plasticity (RGP) has been detected from PPanGGOlin output through the script ppanggolin rgp -p pangenome.h5 [35]. 470 Associations between binary matrices (presence/absence) of accessory gene families or RGP 471 (singletons excluded) and the main sources of isolation (human stool, pig intestine, and its main 472 473 sections) were assessed with Scoary scripts [88] and considered significant for P-value [FDR 474 adjusted] < 0.05.

Moreover, with the purpose to explore all possible virulence-associated genes present in the genomes, we constructed a repertoire of genes linked to host/pathogen interactions (mucus interaction, adhesion, invasion, modulation of host genes), chemotaxis, motility and general factors related to virulence mechanisms (**Supplementary Table 4**). Genomes, and genes detected using the software described above, were manually curated, and the presence of sequences of interest has been confirmed by BLAST alignment towards reference sequences (<u>https://blast.ncbi.nlm.nih.gov/Blast.cgi</u>) [89].

481

482 *4.5 PHYLOGENETIC ANALYSES*

Phylogenetic UPGMA trees were computed for whole and core genomes of the strains object of the
present study using the software ND tree (version 1.2) with the *Campylobacter jejuni* NCTC 11168
(NC_002163.1) genome as outgroup. The MLST sequences were analyzed with the software clustalX
(Multiple aligned modes, version 2.0) [90].

An *in silico* MLST analysis has been performed employing the on-line suite MLST 2.0 (https://cge.cbs.dtu.dk/services/MLST/) for all strains [91], by using the sequences used for *Aliarcobacter* spp., specifically *aspA*, *atpA*, *glnA*, *gltA*, *glyA*, *pgm*, *tkt* [92]. After the obtainment of the MLST numeric codes (**Supplementary Table 5**), the MLST sequences of all strains were stored in FASTA format for phylogenetic analysis.

The Approximately-maximum-likelihood phylogenetic tree of SNPs present in the 32 genomes was produced with the type genome of *A. butzleri* RM4018 synonymous of LMG10828^T as reference (NC_009850.1), using the CSI Phylogeny pipeline (Call SNPs & Infer Phylogeny, CGE, version 1.4) with default options. SNPs detected by the software CSI Phylogeny have been checked with BWA (version 0.7.17) and Samtools software (version 0.1.19) [93].

Phylogenetic trees were visualized with iTOL (version 5.5.1) to obtain the image format choice [94],
while the software Morpheus (https://software.broadinstitute.org/morpheus) was used in the heatmap
production [95].

500

501 *4.6 STATISTICAL ANALYSIS*

502 Correlation between presence/absence of virulence-associated genes and colonization/invasion rates 503 was computed by Pearson's product-moment correlation (considered significant for *P*-value [FDR 504 adjusted] < 0.05) in R environment.

505 Normality and homogeneity of the data from colonization and invasion assays were checked using 506 Shapiro-Wilk's W and Levene's tests, respectively. Kruskal–Wallis (K-W) and ANOVA were used 507 to assess the overall variation and differences between the multiple groups, for nonparametric and 508 parametric data respectively. Pairwise Wilcoxon's test and Duncan's test were used as post hoc

analyses for nonparametric and parametric data respectively. Data were presented in boxplots graph 509 510 (median, range interquartile, min/max and outliers). Statistics and data plotting were performed with the R program for Statistical Computing 3.6.0 (http://www.r-project.org) unless otherwise stated. 511 512

5. AVAILABILITY OF DATA AND MATERIAL 513

Raw sequence reads were deposited at the Sequence Read Archive of the National Center for 514 515 Biotechnology Information (Bioproject accession number: PRJNA660594). The genomes assembled sequences, the sequences of the predicted transcripts and amino acidic sequences (.faa, .fna, .gff, .gbf, 516 .sqn, .tbl, .ffn), and the files used to construct the pangenome network (edges.csv, nodes.csv) are 517 518 available on Zenodo (https://zenodo.org/) at http://doi.org/10.5281/zenodo.4301795.

519

6. AUTHOR CONTRIBUTIONS 520

Investigations: DB. Formal analysis: DB, CB, IF. Writing- Original draft preparation: DB, CB. 521 Methodology: CB, VA, KH, KR. Writing – Review & Editing: IF, VA, KH, KR. Conceptualization: 522

523 VA, KR. Supervision: VA, KH, KR. Resource: KH. Funding acquisition: KR.

524

525

526

527 REFERENCES

- [1] D. Chieffi, F. Fanelli, V. Fusco, Arcobacter butzleri : Up-to-date taxonomy, ecology, and 528 pathogenicity of an emerging pathogen, Compr. Rev. Food Sci. Food Saf. (2020) 1541-529
- 4337.12577. https://doi.org/10.1111/1541-4337.12577. 530
- A. Oren, G.M. Garrity, List of new names and new combinations previously effectively, but 531 [2] not validly, published, Int. J. Syst. Evol. Microbiol. 70 (2020) 1-5. 532

533 https://doi.org/10.1099/ijsem.0.003881.

- A.-M. Van den Abeele, D. Vogelaers, J. Van Hende, K. Houf, Prevalence of *Arcobacter*Species among Humans, Belgium, 2008–2013, Emerg. Infect. Dis. 20 (2014) 1746–1749.
 https://doi.org/10.3201/eid2010.140433.
- 537 [4] G. Gölz, T. Alter, S. Bereswill, M.M. Heimesaat, The immunopathogenic potential of
- Arcobacter butzleri Lessons from a meta-analysis of murine infection studies, PLoS One.
 11 (2016) 1–18. https://doi.org/10.1371/journal.pone.0159685.
- 540 [5] S. De Smet, L. De Zutter, K. Houf, Spatial Distribution of the Emerging Foodborne Pathogen
- 541 *Arcobacter* in the Gastrointestinal Tract of Pigs, Foodborne Pathog. Dis. 9 (2012) 1097–
- 542 1103. https://doi.org/10.1089/fpd.2012.1184.
- [6] N. Shange, P. Gouws, L.C. Hoffman, *Campylobacter* and *Arcobacter* species in foodproducing animals: prevalence at primary production and during slaughter, World J.
 Microbiol. Biotechnol. 35 (2019) 146. https://doi.org/10.1007/s11274-019-2722-x.
- 546 [7] C.L. Hilton, B.M. Mackey, A.J. Hargreaves, S.J. Forsythe, The recovery of *Arcobacter*
- 547 *butzleri* NCTC 12481 from various temperature treatments, J. Appl. Microbiol. 91 (2001)

548 929–932. https://doi.org/10.1046/j.1365-2672.2001.01457.x.

- 549 [8] T.P. Ramees, K. Dhama, K. Karthik, R.S. Rathore, A. Kumar, M. Saminathan, R. Tiwari,
- 550 Y.S. Malik, R.K. Singh, *Arcobacter* : an emerging food-borne zoonotic pathogen, its public
- 551 health concerns and advances in diagnosis and control a comprehensive review, Vet. Q. 37
- 552 (2017) 136–161. https://doi.org/10.1080/01652176.2017.1323355.
- 553 [9] S.C. Pearce, H.G. Coia, J.P. Karl, I.G. Pantoja-feliciano, N.C. Zachos, K. Racicot, Intestinal
- in vitro and ex vivo Models to Study Host-Microbiome Interactions and Acute Stressors,
- 555 Front. Physiol. 9 (2018). https://doi.org/10.3389/fphys.2018.01584.

| 556 | [10] | G.P. Donaldson, S.M. Lee, S.K. Mazmanian, Gut biogeography of the bacterial microbiota, |
|-----|------|--|
| 557 | | Nat. Rev. Microbiol. 14 (2016) 20-32. https://doi.org/10.1038/nrmicro3552. |
| 558 | [11] | A.P. Moran, A. Gupta, L. Joshi, Sweet-talk: role of host glycosylation in bacterial |
| 559 | | pathogenesis of the gastrointestinal tract, Gut. 60 (2011) 1412–1425. |
| 560 | | https://doi.org/10.1136/gut.2010.212704. |
| 561 | [12] | G.C. Hansson, Role of mucus layers in gut infection and inflammation, Curr. Opin. |
| 562 | | Microbiol. 15 (2012) 57-62. https://doi.org/10.1016/j.mib.2011.11.002. |
| 563 | [13] | JF. Sicard, G. Le Bihan, P. Vogeleer, M. Jacques, J. Harel, Interactions of Intestinal |
| 564 | | Bacteria with Components of the Intestinal Mucus, Front. Cell. Infect. Microbiol. 7 (2017). |
| 565 | | https://doi.org/10.3389/fcimb.2017.00387. |
| 566 | [14] | F. Fanelli, A. Di Pinto, A. Mottola, G. Mule, D. Chieffi, F. Baruzzi, G. Tantillo, V. Fusco, |
| 567 | | Genomic Characterization of Arcobacter butzleri Isolated From Shellfish: Novel Insight Into |
| 568 | | Antibiotic Resistance and Virulence Determinants, Front. Microbiol. 10 (2019) 1–17. |
| 569 | | https://doi.org/10.3389/fmicb.2019.00670. |

J. Isidro, S. Ferreira, M. Pinto, F. Domingues, M. Oleastro, J.P. Gomes, V. Borges, Virulence
and antibiotic resistance plasticity of *Arcobacter butzleri*: Insights on the genomic diversity

of an emerging human pathogen, Infect. Genet. Evol. 80 (2020) 104213.

573 https://doi.org/10.1016/j.meegid.2020.104213.

[16] R. Bücker, H. Troeger, J. Kleer, M. Fromm, J.-D. Schulzke, *Arcobacter butzleri* Induces
Barrier Dysfunction in Intestinal HT-29/B6 Cells, J. Infect. Dis. 200 (2009) 756–764.
https://doi.org/10.1086/600868.

577 [17] M.A. Bianchi, D. Del Rio, N. Pellegrini, G. Sansebastiano, E. Neviani, F. Brighenti, A
578 fluorescence-based method for the detection of adhesive properties of lactic acid bacteria to

579 Caco-2 cells, Lett. Appl. Microbiol. 39 (2004) 301–305. https://doi.org/10.1111/j.1472-

580 765X.2004.01589.x.

J.M. Laparra, Y. Sanz, Comparison of in vitro models to study bacterial adhesion to the
intestinal epithelium, Lett. Appl. Microbiol. 49 (2009) 695–701.

583 https://doi.org/10.1111/j.1472-765X.2009.02729.x.

- [19] G. Karadas, R. Bücker, S. Sharbati, J.-D. Schulzke, T. Alter, G. Gölz, *Arcobacter butzleri*isolates exhibit pathogenic potential in intestinal epithelial cell models, J. Appl. Microbiol.
 120 (2016) 218–225. https://doi.org/10.1111/jam.12979.
- [20] G. Karadas, S. Sharbati, I. Hänel, U. Messelhäußer, E. Glocker, T. Alter, G. Gölz, Presence
 of virulence genes, adhesion and invasion of *Arcobacter butzleri*, J. Appl. Microbiol. 115
 (2013) 583–590. https://doi.org/10.1111/jam.12245.
- 590 [21] A. Klančnik, I. Gobin, B. Jeršek, S. Smole Možina, D. Vučković, M. Tušek Žnidarič, M.
 591 Abram, Adhesion of *Campylobacter jejuni* Is Increased in Association with Foodborne
 592 Bacteria, Microorganisms. 8 (2020) 201. https://doi.org/10.3390/microorganisms8020201.
- 593 [22] P.H. Everst, H. Goossens, J.P. Butzler, D. Lloyd, S. Knutton, J.M. Ketley, P.H. Williams,
- 594 Differentiated Caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *C*.

595 *coli*, J. Med. Microbiol. 37 (1992) 319–325. https://doi.org/10.1099/00222615-37-5-319.

- A. Alemka, N. Corcionivoschi, B. Bourke, Defense and Adaptation: The Complex InterRelationship between *Campylobacter jejuni* and Mucus, Front. Cell. Infect. Microbiol. 2
 (2012) 1–6. https://doi.org/10.3389/fcimb.2012.00015.
- 599 [24] L. Douidah, L. De Zutter, F. Van Nieuwerburgh, D. Deforce, H. Ingmer, O. Vandenberg,
- A.M. Van Den Abeele, K. Houf, Presence and analysis of plasmids in human and animal
- associated *Arcobacter* species, PLoS One. 9 (2014) 1–8.
- 602 https://doi.org/10.1371/journal.pone.0085487.

| 603 | [25] | F. Rovetto, A. Carlier, A.M. Van Den Abeele, K. Illeghems, F. Van Nieuwerburgh, L. |
|-----|------|--|
| 604 | | Cocolin, K. Houf, Characterization of the emerging zoonotic pathogen Arcobacter thereius |
| 605 | | by whole genome sequencing and comparative genomics, PLoS One. 12 (2017) 1–27. |
| 606 | | https://doi.org/10.1371/journal. pone.0180493. |
| 607 | [26] | C. Botta, A. Acquadro, A. Greppi, L. Barchi, M. Bertolino, K. Rantsiou, Genomic |

- assessment in *Lactobacillus plantarum* links the butyrogenic pathway with glutamine
 metabolism, Sci. Rep. (2017) 1–13. https://doi.org/10.1038/s41598-017-16186-8.
- 610 [27] M.T. Cabeen, C. Jacobs-wagner, The Bacterial Cytoskeleton, Annu. Rev. Of Genetics.
 611 (2010). https://doi.org/10.1146/annurev-genet-102108-134845.
- [28] M. Furter, M.E. Sellin, G.C. Hansson, W.-D. Hardt, Mucus Architecture and Near-Surface
 Swimming Affect Distinct *Salmonella* Typhimurium Infection Patterns along the Murine
 Intestinal Tract, Cell Rep. 27 (2019) 2665-2678.e3.
- 615 https://doi.org/10.1016/j.celrep.2019.04.106.
- 616 [29] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G.G. Holden, M. Fookes, D.

617Falush, J.A. Keane, J. Parkhill, Roary: Rapid large-scale prokaryote pan genome analysis,

618 Bioinformatics. 31 (2015) 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

619 [30] G. Gautreau, A. Bazin, M. Gachet, R. Planel, L. Burlot, M. Dubois, A. Perrin, C. Médigue,

A. Calteau, S. Cruveiller, C. Matias, C. Ambroise, E.P.C. Rocha, D. Vallenet,

621 PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph, PLoS

622 Comput. Biol. 16 (2020) 1–27. https://doi.org/10.1371/journal.pcbi.1007732.

623 [31] A.A. Golicz, P.E. Bayer, P.L. Bhalla, J. Batley, D. Edwards, Pangenomics Comes of Age:

From Bacteria to Plant and Animal Applications, Trends Genet. 36 (2020) 132–145.

625 https://doi.org/10.1016/j.tig.2019.11.006.

626 [32] T. Gouliouris, K.E. Raven, C. Ludden, B. Blane, J. Corander, C.S. Horner, J. Hernandez-

| 627 | | Garcia, P. Wood, N.F. Hadjirin, M. Radakovic, M.A. Holmes, M. de Goffau, N.M. Brown, J. |
|------|------|--|
| 628 | | Parkhill, S.J. Peacock, Genomic surveillance of enterococcus faecium reveals limited sharing |
| 629 | | of strains and resistance genes between livestock and humans in the United Kingdom, MBio. |
| 630 | | 9 (2018) 1–15. https://doi.org/10.1128/mBio.01780-18. |
| 631 | [33] | M.R. Davies, L. McIntyre, A. Mutreja, J.A. Lacey, J.A. Lees, R.J. Towers, S. Duchêne, P.R. |
| 632 | | Smeesters, H.R. Frost, D.J. Price, M.T.G. Holden, S. David, P.M. Giffard, K.A. Worthing, |
| 633 | | A.C. Seale, J.A. Berkley, S.R. Harris, T. Rivera-Hernandez, O. Berking, A.J. Cork, R.S.L.A. |
| 634 | | Torres, T. Lithgow, R.A. Strugnell, R. Bergmann, P. Nitsche-Schmitz, G.S. Chhatwal, S.D. |
| 635 | | Bentley, J.D. Fraser, N.J. Moreland, J.R. Carapetis, A.C. Steer, J. Parkhill, A. Saul, D.A. |
| 636 | | Williamson, B.J. Currie, S.Y.C. Tong, G. Dougan, M.J. Walker, Atlas of group A |
| 637 | | streptococcal vaccine candidates compiled using large-scale comparative genomics, Nat. |
| 638 | | Genet. 51 (2019) 1035–1043. https://doi.org/10.1038/s41588-019-0417-8. |
| 639 | [34] | M. Juhas, J.R. Van Der Meer, M. Gaillard, R.M. Harding, D.W. Hood, D.W. Crook, |
| 640 | | Genomic islands: Tools of bacterial horizontal gene transfer and evolution, FEMS Microbiol. |
| 641 | | Rev. 33 (2009) 376–393. https://doi.org/10.1111/j.1574-6976.2008.00136.x. |
| 642 | [35] | A. Bazin, G. Gautreau, C. Médigue, D. Vallenet, A. Calteau, panRGP: a pangenome-based |
| 643 | | method to predict genomic islands and explore their diversity, BioRxiv. 1 (2020) |
| 644 | | 2020.03.26.007484. https://doi.org/10.1101/2020.03.26.007484. |
| 6.45 | [27] | |
| 645 | [36] | U. Sood, P. Hira, R. Kumar, A. Bajaj, D.L.N. Rao, R. Lal, M. Shakarad, Comparative |
| 646 | | genomic analyses reveal core-genome-wide genes under positive selection and major |
| 647 | | regulatory hubs in outlier strains of <i>Pseudomonas aeruginosa</i> , Front. Microbiol. 10 (2019). |
| 648 | | https://doi.org/10.3389/fmicb.2019.00053. |
| | | |

649 [37] U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, Genomic islands in pathogenic and
650 environmental microorganisms, Nat. Rev. Microbiol. 2 (2004) 414–424.

651 https://doi.org/10.1038/nrmicro884.

[38] E. Mourkasa, A.J. Taylorb, G. Mérica, S.C. Baylissa, B. Pascoea, L. Mageirosa, J.K.

- 653 Callanda, M.D. Hitchingse, A. Ridleyf, A. Vidalf, K.J. Forbesg, N.J.C.S. Strachanh, C.T.
- 654 Parkeri, J. Parkhillj, A.K. Jolley, A.J. Codyk, M.C.J. Maiden, D.J. Kelly, S.K. Sheppard,
- Agricultural intensification and the evolution of host specialism in the enteric pathogen
- 656 *Campylobacter jejuni*, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 11018–11028.

657 https://doi.org/10.1073/pnas.1917168117.

- [39] B. Liu, Y.A. Knirel, L. Feng, A. V. Perepelov, S.N. Senchenkova, P.R. Reeves, L. Wang,
 Structural diversity in *Salmonella* O antigens and its genetic basis, FEMS Microbiol. Rev. 38
 (2014) 56–89. https://doi.org/10.1111/1574-6976.12034.
- [40] I. V. Wesley, A.L. Baetz, D.J. Larson, Infection of cesarean-derived colostrum-deprived 1 day-old piglets with *Arcobacter butzleri*, *Arcobacter cryaerophilus*, and *Arcobacter skirrowii*, Infect. Immun. 64 (1996) 2295–2299.
- 664 [41] D.J. Woodcock, P. Krusche, N.J.C. Strachan, K.J. Forbes, F.M. Cohan, G. Méric, S.K.
- 665 Sheppard, Genomic plasticity and rapid host switching can promote the evolution of
- generalism: A case study in the zoonotic pathogen *Campylobacter*, Sci. Rep. 7 (2017) 1–13.
- 667 https://doi.org/10.1038/s41598-017-09483-9.
- [42] H.H. De Fine Licht, Does pathogen plasticity facilitate host shifts?, PLoS Pathog. 14 (2018)
 1–9. https://doi.org/10.1371/journal.ppat.1006961.
- [43] L. Douidah, L. de Zutter, J. Bare, P. De Vos, P. Vandamme, O. Vandenberg, A.-M. Van den
- 671 Abeele, K. Houf, Occurrence of Putative Virulence Genes in *Arcobacter* Species Isolated
- from Humans and Animals, J. Clin. Microbiol. 50 (2012) 735–741.
- 673 https://doi.org/10.1128/JCM.05872-11.
- 674 [44] A. Parisi, L. Capozzi, A. Bianco, M. Caruso, L. Latorre, A. Costa, A. Giannico, D. Ridolfi,

- C. Bulzacchelli, G. Santagada, Identification of virulence and antibiotic resistance factors in
 Arcobacter butzleri isolated from bovine milk by whole genome sequencing, Ital. J. Food
- 677 Saf. 8 (2019). https://doi.org/10.4081/ijfs.2019.7840.
- [45] C. Girbau, C. Guerra, I. Martínez-Malaxetxebarria, R. Alonso, A. Fernández-Astorga,
- 679 Prevalence of ten putative virulence genes in the emerging foodborne pathogen *Arcobacter*680 isolated from food products, Food Microbiol. 52 (2015) 146–149.

681 https://doi.org/10.1016/j.fm.2015.07.015.

[46] S. Ferreira, J.A. Queiroz, M. Oleastro, F.C. Domingues, Genotypic and phenotypic features
of *Arcobacter butzleri* pathogenicity, Microb. Pathog. 76 (2014) 19–25.

684 https://doi.org/10.1016/j.micpath.2014.09.004.

[47] Z. Wu, B. Periaswamy, O. Sahin, M. Yaeger, P. Plummer, W. Zhai, Z. Shen, L. Dai, S.L.

686 Chen, Q. Zhang, Point mutations in the major outer membrane protein drive hypervirulence

687of a rapidly expanding clone of *Campylobacter jejuni*, Proc. Natl. Acad. Sci. 113 (2016)

688 10690–10695. https://doi.org/10.1073/pnas.1605869113.

- [48] C. Sabet, A. Toledo-Arana, N. Personnic, M. Lecuit, S. Dubrac, O. Poupel, E. Gouin, M.-
- 690 A.M. Nahori, P. Cossart, H.H.H. Bierne, The *Listeria monocytogenes* Virulence Factor InlJ
- Is Specifically Expressed In Vivo and Behaves as an Adhesin, Infect. Immun. 76 (2008)
- 692 1368–1378. https://doi.org/10.1128/IAI.01519-07.

[49] M.W. Gilmour, M. Graham, G. Van Domselaar, S. Tyler, H. Kent, K.M. Trout-Yakel, O.

- 694 Larios, V. Allen, B. Lee, C. Nadon, High-throughput genome sequencing of two *Listeria*
- *monocytogenes* clinical isolates during a large foodborne outbreak, BMC Genomics. 11
- 696 (2010) 120. https://doi.org/10.1186/1471-2164-11-120.
- 697 [50] A.P. Corfield, S.A. Wagner, J.R. Clamp, M.S. Kriaris, L.C. Hoskins, Mucin degradation in
 698 the human colon: production of sialidase, sialate O-acetylesterase, N-acetylneuraminate

- lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria., Infect. Immun. 699 700 60 (1992) 3971-3978. https://doi.org/10.1128/iai.60.10.3971-3978.1992. 701 M. Sandkvist, Type II Secretion and Pathogenesis, Infect. Immun. 69 (2001) 3523–3535. [51] 702 https://doi.org/10.1128/IAI.69.6.3523. 703 [52] T. Yoshida, Y. Ayabe, M. Yasunaga, Y. Usami, H. Habe, H. Nojiri, T. Omori, Genes 704 involved in the synthesis of the exopolysaccharide methanolan by the obligate methylotroph 705 Methylobacillus sp. strain 12S, Microbiology. 149 (2003) 431-444. https://doi.org/10.1099/mic.0.25913-0. 706 707 [53] S. Lebeer, T.L.A. Verhoeven, G. Francius, G. Schoofs, I. Lambrichts, Y. Dufrêne, J. Vanderleyden, S.C.J. De Keersmaecker, Identification of a gene cluster for the biosynthesis 708 of a long, galactose-rich exopolysaccharide in Lactobacillus rhamnosus GG and functional 709
 - analysis of the priming glycosyltransferase, Appl. Environ. Microbiol. 75 (2009) 3554–3563.
 https://doi.org/10.1128/AEM.02919-08.
 - 712 [54] S. Ferreira, M.J. Fraqueza, J.A. Queiroz, F.C. Domingues, M. Oleastro, Genetic diversity,
- antibiotic resistance and biofilm-forming ability of *Arcobacter butzleri* isolated from poultry
- and environment from a Portuguese slaughterhouse, Int. J. Food Microbiol. 162 (2013) 82–

715 88. https://doi.org/10.1016/j.ijfoodmicro.2013.01.003.

- 716 [55] G.G. Sgro, G.U. Oka, D.P. Souza, W. Cenens, E. Bayer-Santos, B.Y. Matsuyama, N.F.
- Bueno, T.R. dos Santos, C.E. Alvarez-Martinez, R.K. Salinas, C.S. Farah, Bacteria-Killing
 Type IV Secretion Systems, Front. Microbiol. 10 (2019) 1–20.
- 719 https://doi.org/10.3389/fmicb.2019.01078.
- 720 [56] R.A. Batchelor, Nucleotide sequences and comparison of two large conjugative plasmids
- from different *Campylobacter* species, Microbiology. 150 (2004) 3507–3517.
- 722 https://doi.org/10.1099/mic.0.27112-0.

- [57] C.E. Alvarez-Martinez, P.J. Christie, Biological Diversity of Prokaryotic Type IV Secretion
- 724 Systems, Microbiol. Mol. Biol. Rev. 73 (2009) 775–808.
- 725 https://doi.org/10.1128/mmbr.00023-09.
- [58] S. Merino, J.M. Tomás, Gram-negative flagella glycosylation, Int. J. Mol. Sci. 15 (2014)
- 727 2840–2857. https://doi.org/10.3390/ijms15022840.
- J.M.D. Roberts, L.L. Graham, B. Quinn, D.A. Pink, Modeling the surface of *Campylobacter fetus*: Protein surface layer stability and resistance to cationic antimicrobial peptides,
- 730 Biochim. Biophys. Acta Biomembr. 1828 (2013) 1143–1152.
- 731 https://doi.org/10.1016/j.bbamem.2012.10.025.
- [60] S. Kalynych, R. Morona, M. Cygler, Progress in understanding the assembly process of
- bacterial O-antigen, FEMS Microbiol. Rev. 38 (2014) 1048–1065.
- 734 https://doi.org/10.1111/1574-6976.12070.
- 735 [61] R.F. Maldonado, I. Sá-Correia, M.A. Valvano, Lipopolysaccharide modification in Gram-
- negative bacteria during chronic infection, FEMS Microbiol. Rev. 40 (2016) 480–493.
- 737 https://doi.org/10.1093/femsre/fuw007.
- F. Fanelli, D. Chieffi, A. Di Pinto, A. Mottola, F. Baruzzi, V. Fusco, Phenotype and genomic
 background of *Arcobacter butzleri* strains and taxogenomic assessment of the species, Food
 Microbiol. 89 (2020) 103416. https://doi.org/10.1016/j.fm.2020.103416.
- [63] S.A. Reeves, A.G. Torres, S.M. Payne, TonB Is Required for Intracellular Growth and
- 742 Virulence of *Shigella dysenteriae*, Infect. Immun. 68 (2000) 6329–6336.
- 743 https://doi.org/10.1128/IAI.68.11.6329-6336.2000.
- [64] A.G. Torres, P. Redford, R.A. Welch, S.M. Payne, TonB-Dependent Systems of
- 745 Uropathogenic *Escherichia coli*: Aerobactin and Heme Transport and TonB Are Required for
- 746 Virulence in the Mouse, Infect. Immun. 69 (2001) 6179–6185.

747

https://doi.org/10.1128/IAI.69.10.6179-6185.2001.

- [65] W. Simpson, T. Olczak, C.A. Genco, Characterization and Expression of HmuR, a TonBDependent Hemoglobin Receptor of *Porphyromonas gingivalis*, J. Bacteriol. 182 (2000)
 5737–5748. https://doi.org/10.1128/JB.182.20.5737-5748.2000.
- [66] D.J. Morton, R.J. Hempel, T.W. Seale, P.W. Whitby, T.L. Stull, A functional *tonB* gene is
 required for both virulence and competitive fitness in a chinchilla model of *Haemophilus influenzae* otitis media, BMC Res. Notes. 5 (2012) 327. https://doi.org/10.1186/1756-0500-5327.
- 755 [67] X. Zhuge, Y. Sun, F. Xue, F. Tang, J. Ren, D. Li, J. Wang, M. Jiang, J. Dai, A Novel
- PhoP/PhoQ Regulation Pathway Modulates the Survival of Extraintestinal Pathogenic
 Escherichia coli in Macrophages, Front. Immunol. 9 (2018) 1–23.
- 758 https://doi.org/10.3389/fimmu.2018.00788.
- [68] I. Zwir, D. Shin, A. Kato, K. Nishino, T. Latifi, F. Solomon, J.M. Hare, H. Huang, E.A.
- 760 Groisman, Dissecting the PhoP regulatory network of *Escherichia coli* and *Salmonella*
- 761 *enterica*, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2862–2867.
- 762 https://doi.org/10.1073/pnas.0408238102.
- [69] H. He, T.C. Zahrt, Identification and Characterization of a Regulatory Sequence Recognized
 by *Mycobacterium tuberculosis* Persistence Regulator MprA, J. Bacteriol. 187 (2004) 202–
- 765 212. https://doi.org/10.1128/JB.187.1.202.
- 766 [70] Y. Eguchi, T. Okada, S. Minagawa, T. Oshima, H. Mori, K. Yamamoto, A. Ishihama, R.
- 767 Utsumi, Signal Transduction Cascade between EvgA / EvgS and PhoP / PhoQ Two-
- Component Systems of *Escherichia coli*, J. Bacteriol. 186 (2004) 3006–3014.
- 769 https://doi.org/10.1128/JB.186.10.3006.
- 770 [71] A.M. Lippa, M. Goulian, Perturbation of the Oxidizing Environment of the Periplasm

- Stimulates the PhoQ/PhoP System in *Escherichia coli*, J. Bacteriol. 194 (2012) 1457–1463.
 https://doi.org/10.1128/JB.06055-11.
- [72] L.N. Schulte, M. Schweinlin, A.J. Westermann, H. Janga, S.C. Santos, S. Appenzeller, H.
- Walles, J. Vogel, M. Metzger, An Advanced Human Intestinal Coculture Model Reveals
- 775 Compartmentalized Host and Pathogen Strategies during *Salmonella* Infection, MBio. 11
- 776 (2020). https://doi.org/10.1128/mBio.03348-19.
- [73] K. Houf, R. Stephan, Isolation and characterization of the emerging foodborn pathogen
- Arcobacter from human stool, J. Microbiol. Methods. 68 (2007) 408–413.
- 779 https://doi.org/10.1016/j.mimet.2006.09.020.
- [74] C. Botta, T. Langerholc, A. Cencič, L. Cocolin, *In vitro* selection and characterization of new
 probiotic candidates from table olive microbiota, PLoS One. 9 (2014).
- 782 https://doi.org/10.1371/journal.pone.0094457.
- [75] X.M. Chen, I. Elisia, D.D. Kitts, Defining conditions for the co-culture of Caco-2 and HT29 MTX cells using Taguchi design, J. Pharmacol. Toxicol. Methods. 61 (2010) 334–342.
 https://doi.org/10.1016/j.vascn.2010.02.004.
- [76] M. Pinto, S. Robine-Leon, M.-D. Appay, M. Kedinger, N. Triadou, E. Dussaulx, B. Lacroix,
 P. Simon-Assmann, K. Haffen, J. Fogh, A. Zweibaum, Enterocyte-like differentiation and
 polarization of the human colon carcinoma cell line Caco-2 in culture, Biol. Cell. 47 (1983)
 323–330.
- 790 [77] S. Backert, D. Hofreuter, Molecular methods to investigate adhesion, transmigration,
- invasion and intracellular survival of the foodborne pathogen *Ccampylobacter jejuni*, J.
- 792 Microbiol. Methods. 95 (2013) 8–23. https://doi.org/10.1016/j.mimet.2013.06.031.
- 793 [78] A. Levican, A. Alkeskas, C. Gunter, S.J. Forsythe, M.J. Figueras, Adherence to and Invasion
 794 of Human Intestinal Cells by *Arcobacter* Species and Their Virulence Genotypes, Appl.

795 Environ. Microbiol. 79 (2013) 4951–4957. https://doi.org/10.1128/AEM.01073-13.

- 796 [79] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin,
- 797 S.I. Nikolenko, S. Pham, A.D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G.
- 798 Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: A new genome assembly algorithm and its
- applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.
- 800 https://doi.org/10.1089/cmb.2012.0021.
- [80] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome
 assemblies, Bioinformatics. 29 (2013) 1072–1075.
- 803 https://doi.org/10.1093/bioinformatics/btt086.
- [81] T. Seemann, Prokka: Rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–
 2069. https://doi.org/10.1093/bioinformatics/btu153.
- 806 [82] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C.
- 807 Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N.
- 808 Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's
- conserved domain database, Nucleic Acids Res. 43 (2015) D222–D226.
- 810 https://doi.org/10.1093/nar/gku1221.
- [83] I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly
 interspaced short palindromic repeats, Nucleic Acids Res. 35 (2007) W52–W57.
- 813 https://doi.org/10.1093/nar/gkm360.
- 814 [84] D. Arndt, J.R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, D.S. Wishart, PHASTER: a
- better, faster version of the PHAST phage search tool, Nucleic Acids Res. 44 (2016) W16–
- 816 W21. https://doi.org/10.1093/nar/gkw387.
- 817 [85] A. Marchler-Bauer, CDD: a Conserved Domain Database for protein classification, Nucleic
 818 Acids Res. 33 (2004) D192–D196. https://doi.org/10.1093/nar/gki069.

- [86] M. Yang, M.K. Derbyshire, R.A. Yamashita, A. Marchler-Bauer, NCBI's Conserved
 Domain Database and Tools for Protein Domain Analysis, Curr. Protoc. Bioinforma. 69
 (2020) 1–25. https://doi.org/10.1002/cpbi.90.
- 822 [87] S. Wu, Z. Zhu, L. Fu, B. Niu, W. Li, WebMGA: a customizable web server for fast
- metagenomic sequence analysis, BMC Genomics. 12 (2011) 444.
- 824 https://doi.org/10.1186/1471-2164-12-444.
- [88] O. Brynildsrud, J. Bohlin, L. Scheffer, V. Eldholm, Rapid scoring of genes in microbial pangenome-wide association studies with Scoary, Genome Biol. 17 (2016) 1–9.
- 827 https://doi.org/10.1186/s13059-016-1108-8.
- 828 [89] S. McGinnis, T.L. Madden, BLAST: At the core of a powerful and diverse set of sequence
 829 analysis tools, Nucleic Acids Res. 32 (2004) 20–25. https://doi.org/10.1093/nar/gkh435.
- 830 [90] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F.
- 831 Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins,
- Clustal W and Clustal X version 2.0, Bioinformatics. 23 (2007) 2947–2948.
- https://doi.org/10.1093/bioinformatics/btm404.
- 834 [91] M. V Larsen, S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R.L. Marvig, L. Jelsbak, T.
- 835 Sicheritz-Ponten, D.W. Ussery, F.M. Aarestrup, O. Lund, Multilocus Sequence Typing of
- Total-Genome-Sequenced Bacteria, J. Clin. Microbiol. 50 (2012) 1355–1361.
- 837 https://doi.org/10.1128/JCM.06094-11.
- 838 [92] W.G. Miller, I. V. Wesley, S.L.W. On, K. Houf, F. Mégraud, G. Wang, E. Yee, A. Srijan,
- 839 C.J. Mason, First multi-locus sequence typing scheme for *Arcobacter* spp., BMC Microbiol.
- 840 9 (2009) 196. https://doi.org/10.1186/1471-2180-9-196.
- [93] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2 Approximately Maximum-Likelihood Trees
 for Large Alignments, PLoS One. 5 (2010) e9490.

- 843 https://doi.org/10.1371/journal.pone.0009490.
- 844 [94] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new
- developments, Nucleic Acids Res. 47 (2019) W256–W259.
- 846 https://doi.org/10.1093/nar/gkz239.
- 847 [95] M.C. Ryan, M. Stucky, C. Wakefield, J.M. Melott, R. Akbani, J.N. Weinstein, B.M. Broom,
- 848 Interactive Clustered Heat Map Builder: An easy web-based tool for creating sophisticated
- clustered heat maps, F1000Research. 8 (2020) 1750.
- 850 https://doi.org/10.12688/f1000research.20590.2.

852 Figures captions

853

Figure 1. Colonization and invasion capabilities on mucus producer (MP) and not-mucus producer 854 (NMP) models are expressed as Δ Log CFU/cm² (medians ± interquartile range; n=3; dots=outliers) 855 and shown for all 32 strains together (A) and individually for each strain (B). The red dotted line 856 marks the Δ Log equal to 0: a condition in which all bacterial cells added colonized/invaded the model. 857 858 Positive values indicate the potential growth of added bacteria in the model during the co-incubation, while negative values indicate progressively lower colonization/invasion capability. Coding keys of 859 box-plots color are displayed in the caption. Significant differences between models and among the 860 861 strains are reported in the graph (P-value) employing Wilcoxon's test.

862

Figure 2. Bar-plots displaying the average (± standard deviation) distribution of COG classes in all
32 annotated genomes (% of putative proteins assigned to a class compared to the total putative
proteins). Coding keys of classes colors are shown in the caption.

866

867

Figure 3. Partitioned pangenome network (A) displaying the genomic diversity of the 32 strains. 868 869 Nodes represent the gene families and are colored according to the partition (caption), while their size is proportional to the number of genomes in which are present. Edges connect gene families 870 colocalized in the pangenome and their thickness is proportional to the number of genomes sharing 871 that link. Edges are colored as described for nodes, except for edges between partitions (mixed 872 colors). The frame highlights a broad plasticity region of the pangenome (zoomed-in **B**) harboring 873 874 shell/cloud gene families alternatively present in the 32 genomes (pangenome plasticity region; Supplementary Table 2). Input files (nodes.csv and edges.csv) set up for network visualization in 875 876 Gephi (https://gephi.org) are provided on Zenodo (http://doi.org/10.5281/zenodo.4301795). Barplots (C) showing the functional partitioning of gene families in the pangenome plasticity region
(figure B) and all regions of genomic plasticity (RGPs) along the 32 genomes. Asterisks (*) highlight
groups of gene families of which function is manually assigned (Supplementary Table 2).

880

Figure 4. Phylogenetic trees of whole genomes (A), core genomes (B), MLST sequences (C) and SNPs
(D) of the 32 *A. butzleri* strains. The original source of isolation is indicated and groups of strains that
show recurrent clustering patterns are highlighted with colors and named by roman numbers: I
(strains 14, 15); II (strains 1, 8); III (strains 12, 19, 20, 21) and IV (strains 11, 13, 23).

885

Figure 5. Heatmap representing the absence/presence matrix of putative virulence genes detected in the 32 genomes. Gene names or their annotated product are displayed for each gene considered. Asterix (*) highlight putative virulence genes which annotation was verified by alignment with reference strain LMG 10828^T; original annotation in brackets, while caret symbols (^) indicate the presence of non-unique alleles. The groups of strains are indicated from the panes and the group numbers: I (strains 14, 15 from pig), III (strains 12, 19, 20, 21 from pig) and IV (strains 11, 13, 23 from pig), whereas the group II (strain 1 and 28 from human) results absent.

893

894 Supplementary figures list

895

Supplementary Figure 1. Bar-plots (A) displaying the distribution of COG classes in each of the 32
annotated genomes (% of putative proteins assigned to a class compared to the putative proteins).
Coding keys of colors and class codes are shown in the caption. Heatmap (B) showing the presence
(grey) / absence (white) matrix of genes involved in the accessory pathway and COG classes.
Pathways are linked to sequences identified through UniProt codes from the Prokka annotation.

Supplementary figure 2. Singletons distribution along the genomes and composition of the main
clusters of singletons (> 10 loci) identified in the accessory genome.

904

Supplementary figure 3. UPGMA phylogenetic analysis of *porA*. The original source of isolation of the strains is indicated and roman numbers (I–IV) indicate the different groups of strains that coincide with the groups indicated in Figure 4.

908

909 Supplementary figure 4. UPGMA phylogenetic analysis of O-antigen ligase. The original source 910 of isolation of the strains is indicated and the roman numbers (I–IV) indicate the different groups of 911 strains. In the case of the O-antigen ligase dendrogram some strains are repeated, this aspect is linked 912 to the presence of several gene copies. The O-antigen ligase sequence of *Klebsiella pneumoniae* 913 ATCC 700721 (used from Prokka for the functional annotation) has been used as outgroup.

914

915 Tables captions

916 Table 1. A. *butzleri* strains object of study. In the table are shown the number of the strains (nr.), C-917 country of origin (Country), the source of sampling (Source), the specific sampling matrix (Isolation918 source) and additional information such as official strain codes and information related to the patients919 from whom the strain was isolated.

920

Table 2. Pangenome partitions estimated by two computational methods.

922

923 Supplementary tables list

Supplementary Table 1. Bacterial count about colonization/invasion test of the 32 *A. butzleri* strains.
The single strain bacterial loads of the initial inoculum (T0), bacteria load detected after the cell layer
washing (T1) and after the gentamicin application (T2) are expressed in logarithm₁₀ (log) and are
relative to Mucus producing models (MP) and Not mucus producing models (NMP). Moreover, the

standard deviations (st. dv) and the T0, T1, T2 average of the 32 strains on MP and NMP models with
the relative standard deviations are indicated.

930

931

Supplementary Table 2. Annotation statistics of the 32 *A. butzleri* strains. In the first column are indicated the code of the strains and their source of sampling. In the table are indicated the genome size (Mbp), total genes, number of CDS, number of tRNA, hypothetical proteins, transposase, prophage sequences and CRISPR sequences. The CAS sequences detected in the genomes belong to the general class 1 and general class 2. Sequences putative for the production of protein appertain at the bacteriocins bottromycin, microcin and sactipeptides class are indicated with the number of sequences linked to their translation.

939

Supplementary Table 3. List of genes putatively involved in *A. butzleri* virulence. In the second column is present the locus tag codes on the type strain LMG10828^T (strain 3), unless otherwise reported in brackets. The protein codes are relative to UniProt code and Pfam databases. Part of the genes (*) involved in antibiotics resistance and general chemotaxis are only reported here and not in Supplementary Table 4. List of structures and genes involved in the LPS O-antigen biosynthesis. The presence of O-antigen ligase (**) and genes putatively associated to LPS O-antigen cluster assembling (*) are indicated with asterisks.

Supplementary Table 5. genes MLST codes of the strains object of study. In the last column is
indicated the nearest sequence type code- (nearest ST). Some gene sequences resulted in new alleles,
these genes are indicated with an asterisk.















Fig. 4



971972 Fig. 5973

| Strain code in this study | Source | Additional information | Country |
|------------------------------|----------------------|--|---------|
| 1 | Human faeces | Stool sample, (Strain LMG 147149) | Greece |
| 2 | Human faeces | Stool sample, (Strain LMG 111199) | Italy |
| 3 | Human faeces | Stool sample, (Strain LMG 10828 ^T) | U.S.A |
| 4 | Dog faeces | / | Belgium |
| 5 | Sheep faeces | / | Belgium |
| 6 | Horse faeces | / | Belgium |
| 7 | Cow faeces | / | Belgium |
| 8 | Chicken skin | collected from neck | Belgium |
| 9 | Pig meat | / | Belgium |
| 10 | Pig rectum | Intestinal content, (rc1-13) | Belgium |
| 11 | Pig rectum | Intestinal content, (rc1-14) | Belgium |
| 12 | Pig rectum | Intestinal content, (rc2-10) | Belgium |
| 13 | Pig rectum | Intestinal content, (rc2-20) | Belgium |
| 14 | Pig duodenum | Intestinal content, (dc1-3AAN) | Belgium |
| 15 | Pig caecum | Intestinal content, (cm1-2AAN) | Belgium |
| 16 | Pig colon descendent | Mucus, (cdm1-1AAN) | Belgium |
| 17 | Pig colon descendent | Intestinal content, (cdc2-1AAN) | Belgium |
| 18 | Pig colon descendent | Intestinal content, (cdc2-2AAN) | Belgium |
| 19 | Pig rectum | Intestinal content, (rc1-2kAAN) | Belgium |
| 20 | Pig rectum | Intestinal content, (rc1-3AAN) | Belgium |
| 21 | Pig rectum | Mucus, (rm1-2AAN) | Belgium |
| 22 | Pig rectum | Intestinal content, (rc2-1AAN) | Belgium |
| 23 | Pig rectum | Mucus, (rm2-1AAN) | Belgium |
| 24 | Human faeces | Stool sample (male, 90 y/o, diarrhea) | Belgium |
| 25 | Human faeces | Stool sample (female, 93 y/o, acute gastroenteritis) | Belgium |
| 26 | Human faeces | Stool sample (male, 83 y/o, acute gastroenteritis) | Belgium |
| 27 | Human faeces | Stool sample (male, 4 y/o, acute gastroenteritis) | Belgium |
| 28 | Human faeces | Stool sample (male 59 y/o) | Belgium |
| 29 | Human faeces | Stool sample (male, 51 y/o, diverticulitis) | Belgium |
| 30 | Human faeces | Stool sample (male, 55 y/o, traveler's diarrhea) | Belgium |
| 31 | Human faeces | Stool sample (female, 80 y/o, flair up colitis ulcerosa) | Belgium |
| 32 | Human faeces | Stool sample (female, 79 y/o, recurrent diarrhea episodes) | Belgium |

Table 2

| | Pangenome partitions | | | | | | |
|-------------|----------------------|--|---------------------|------------------------------------|------------------|---|-----------|
| Methods | Core (> 99%) | Soft core (95% <= strains < 99%) | Persistent (> 95 %) | Shell (15% <= strains < 95%) | Cloud (< 15%) | Accessory /of which singletones (< 99%) | Pangenome |
| Roary: | 1587 | 76 | 1663 | 970 | 4703 | 5749 / 3311 | 7336 |
| PPanGGolin: | 1651 | 155 | 1806 | 275 | 4542 | 4972 / 2755 | 6623 |