CO-SUPERVISED DOCTORAL RESEARCH THESIS

UNIVERSITÀ DEGLI STUDI DI TORINO
SCUOLA DI DOTTORATO SCIENZE DELLA NATURA E TECNOLOGIE INNOVATIVE
DOTTORATO IN SCIENZE AGRARIE, FORESTALI E ALIMENTARI

GHENT UNIVERSITY
DOCTORAL PROGRAM IN VETERINARY SCIENCE
DEPARTMENT OF VETERINARY AND BIOSCIENCES, FACULTY OF VETERINARY MEDICINE

CYCLE: XXXIV

Study of the *Arcobacter butzleri* genome and transcriptome and evaluation of *Arcobacteraceae* genomic features

Davide Buzzanca

Supervisors:
Prof. Kalliopi Rantsiou
Prof. Kurt Houf

Co-supervisor:
Prof. Valentina Alessandria

Cycle Coordinator:
Prof. Domenico Bosco

YEARS
2018/2019; 2019/2020; 2020/2021

Index

Abbreviations

AAI;	Average Amino-acid Identity
AFLP;	Amplified Fragment Length Polymorphism
ANI;	Percentage of Conserved Proteins
ASB;	Arcobacter selective enrichment broth
ASM;	Arcobacter selective semisolid medium
ASW;	Artificial seawater
ATP;	Adenosine triphosphate
CAT;	Cefoperazone, Amphotericin B, Teicoplanin selective supplement
cDNA;	Complementary DNA
CDS;	Coding DNA sequences
CFU;	Colony forming unit
COGs;	Clusters of Orthologous Genes classes
CRISPR;	Clustered regularly interspaced short palindromic repeat
DEGs;	Differentially expressed genes
DGGE PCR;	Denaturing gradient gel electrophoresis PCR
DMEM;	Dulbecco's Modified Eagle's Medium
dNTPs;	Deoxyribonucleosides triphosphate
EMJH;	Semi-solid Ellinghausen-McCullough-Johnson-Harris
emPCR;	Water-in-oil emulsion PCR
ENA;	European Nucleotide Archive
EP;	Efflux pump
ERIC-PCR;	Enterobacterial Repetitive Intergenic Consensus PCR
FBS;	Fetal bovine serum
FDR;	False discovery rate
FISH;	Fluorescence in situ hybridization

gDNA; Genomic DNA

HGT; Horizontal gene transfer

*is*DDH; *in silico* DNA–DNA hybridization

KEEG; Kyoto Encyclopedia of Genes and Genomes

LogCPM; Log counts per million

LogFC; Log-fold change

LOS; Lipooligosaccharide

LPS; Lipopolysaccharides

MALDI-TOF MS; Matrix-assisted laser desorption ionization mass spectrometry

Mb; Megabases

mCCDA; media modified charcoal cefoperazone deoxycholate agar

mCIN; cefsulodin-irgasan- novobiocin

MEM; Eagle's minimal essential media

MLSA; Multi-locus Sequence Analysis

MLST; Multi-locus sequence typing

MOI; Multiplicity of infection

MP; Mucus producing

m-PCR; Multiplex PCR

NCBI; National Center for Biotechnology Information

NMP; Not-mucus producing

PBS; Phosphate Buffered Saline

PCR; Polymerase Chain Reaction

PE; Paired-end

PFGE; Pulsed-Field Gel Electrophoresis

POCPs; Percentage of Conserved Proteins

PVGs; Putative virulence genes

RAPD;	Random Amplification of Polymorphic DNA
RFLP;	Restriction fragment length polymorphism
RGPs;	Regions of genomic plasticity
RNA-seq;	RNA sequencing
RSCU;	Relative Synonymous Codon Usage
SBS;	Sequencing by synthesis
SMRT;	Single Molecule Real-Time
SNP;	Single nucleotide polymorphism
SRA;	Sequence Read Archive
TEER;	Transepithelial electrical resistance
TSI;	Triple Sugar Iron Agar
WGS;	Whole genome sequencing
ZMWs;	Zero-mode waveguides

1 Literature overview

Table 1.1. Main symptoms related to cases of *Arcobacter* spp. infection. In the table are indicated gastrointestinal and non-gastrointestinal symptoms associated to *Arcobacter* spp. infections [59].

Table 1.2. Information about *A. butzleri* infections collected in the study of Vandamme and colleagues in 1992. These patients lack the most common symptom, diarrhea. The patients were affected by other symptoms [65].

Table 1.3. Presence of nine putative virulence genes in strains studied by Douidah and colleagues. The table shows the strain isolation origin: horse (H), chicken (C), pig (P), sheep (S), human (Hu) and dog (D). The table indicates the genes present (1) and absent (0) in the type strains of the species with reference codes. The number of strains in which the gene has been detected can be seen under the gene name. The column "tot" is the total number of strains tested [91].

Table 1.4. Phenotypical information of nine *Arcobacteraceae* species. The table shows phenotypical information of *A. thereius* (th), *A. butzleri* (bz), *A. cryaerophilus* (cr), *A. skirrowii* (sk), *A. cibarius* (cb), *A. nitrofigilis* (ni), *A. halophilus* (ha), *A. mytili* (my), *A. trophiarum* (tr). The different characteristics are indicated as positive (1, ≥ 95% positive strains) or negative (0, ≤ 11% positive strains) in the different species while the asterisk (*) indicates a range of positive strains of 12-94%. ND = not determined [1,35].

Table 1.5. DNA sequencing technologies. The information in the table regarding some sequencing technologies currently in use have been collected by Bansal and colleagues. From this information it is possible observe the differences between DNA sequencing methods in terms of reads length, error rate and throughput [137].

2 Functional pangenome analysis reveals high virulence plasticity of *Arcobacter butzleri* and affinity to human mucus

Table 2.1. *A. butzleri* strains used in this study. In the table are shown the number of the strains (nr.), C-country of origin (Country), the source of sampling (Source), the specific sampling matrix (Isolation source) and additional information such as official strain codes and information related to the patients from whom the strain was isolated.

Table 2.2. Pangenome partitions estimated by two computational methods.

Figure 2.1. Colonization and invasion capabilities on mucus producer (MP) and not-mucus producer (NMP) models are expressed as Δ Log CFU/cm^2 (medians ± interquartile range; n=3; dots=outliers) and shown for all 32 strains together **(A)** and individually for each strain **(B)**. The red dotted line marks the ΔLog equal to 0: a condition in which all bacterial cells added colonized/invaded the model. 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 box-plots color are displayed in the caption. Significant differences between models and among the strains are reported in the graph (*P*-value) employing Wilcoxon's test.

Figure 2.2. 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.

Figure 2.3. Partitioned pangenome network **(A)** displaying the genomic diversity of the 32 strains. 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 colocalized in the pangenome and their thickness is proportional to the number of genomes sharing that link. Edges are colored as described for nodes, except for edges between partitions (mixed colors). The frame highlights a broad plasticity region of the pangenome (zoomed-in **B**) harboring shell/cloud gene families alternatively present in the 32 genomes (pangenome plasticity region; Supplementary Table 2.2). Input files (nodes.csv and edges.csv) set up for network visualization in Gephi (https://gephi.org/) are provided on Zenodo (http://doi.org/10.5281/zenodo.4301795). Bar-plots (**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 2.2).

Figure 2.4. 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, 28); **III** (strains 12, 19, 20, 21) and **IV (**strains 11, 13, 23**).**

Figure 2.5. 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.

Supplementary Table 2.1. 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 logarithm_{10} (log) and are relative to Mucus producing models (MP) and Not mucus producing models (NMP). Moreover, are indicated 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. ND indicates a not detectable bacterial load (< 100 CFU ml^{-1}).

Supplementary Table 2.2. 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), coverage ((read count * read length)/genome size), 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 classes are indicated with the number of sequences linked to their translation.

Supplementary Table 2.3. 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 **Figure 2.5.**

Supplementary Table 2.4. 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 gene cluster assembling (*) are indicated with asterisks.

Supplementary Table 2.5. genes MLST codes of the strains object of study. In the last column is indicated the nearest sequence type code-(nearest ST). Some genes sequences resulted in new alleles, these genes are indicated with an asterisk.

Supplementary Figure 2.1. Bar-plots (**A**) displaying the distribution of COG classes in the 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 cytoskeleton function.

Supplementary figure 2.2. Singletons distribution along the genomes and composition of the main clusters of singletons (> 10 loci) identified in the accessory genome.

Supplementary figure 2.3. UPGMA phylogenetic analysis of *porA*. The groups of strains from isolated are indicated from the panes, the numbers (I–IV) indicate the different groups of strains.

Supplementary figure 2.4. UPGMA phylogenetic analysis of O-antigen ligase. The strains grouped by source of isolation are indicated from the panes, the numbers (I–IV) indicate the different groups of strains. In the case of the O-antigen ligase dendrogram some strains are repeated, this aspect is linked to the presence of several gene copies. The O-antigen ligase sequence of *Klebsiella pneumoniae* ATCC 700721 (used from Prokka for the functional annotation) has been used as outgroup.

3 Functional pangenome analysis reveals high virulence plasticity of *Arcobacter butzleri* and affinity to human mucus

Figure 3.1. Genomic comparison between *A. butzleri* strains tested (A) and *in vitro* colonization – invasion assay (B). (**A**) The genomes of LMG 11119, LMG 10828^{T} (reference strain) and 31 have been compared with Anvi'o tool. The dendrogram shows gene cluster presence/absence, as well as general information about genomes. The bar chart (**B**) shows the colonization (colonizers bacteria, adhered and internalized bacteria to host

cells) and bacteria localized intracellularly after 30' (light colors) and 90' (dark colors). The values are indicated as Δlog CFU cm^{-2} correspondents to adhered and internalized bacteria (see paragraph 4.3 for calculation specifications). The error bars represent the standard errors while red lines indicate a Δlog of 0 (bacterial load detected equal to initial bacterial inoculum load inoculated into cell models). The red X indicates a bacterial load not detected indicating the absence of bacteria adhering to or entering cells. Furthermore, the figure shows statistical analysis *p*-values with statistical differences between strains indicated near bars.

Figure 3.2. DEGs Venn diagram of *A. butzleri* strains DEGs comparison (A), DEGs pathway classes (B) and virulence-related DEGs heatmap (C) of *A. butzleri* incubated in DMEM after 2 h. The Venn diagram (**A**) shows the presence of DEGs shared between different strains and present in a single strain (hypothetical proteins excluded). The percentage of over-expressed genes (percentages calculated has been calculated on total DEGs including hypothetical proteins) are indicated in brackets next to the strain name. The heatmap **B** shows percentages relating to COG gene classes of total DEGs detected after 2 h of *A. butzleri* incubation in DMEM. The *A. butzleri* differentially expressed genes in DMEM after 2 h of incubation that based on the described genomic functions could play a role in *A. butzleri* virulence are shown in the heatmap **C**.

Figure 3.3. DEGs Venn diagram of *A. butzleri* strains DEGs comparison (A), DEGs pathway classes (B), (C) virulence-related DEGs heatmap of *A. butzleri* in contact with host cells and (D) *tonB* operon organization. The Venn diagram (**A**) shows the presence of DEGs shared between different strains and present in a single strain detected after 30' (in red) and 90' of *A. butzleri* contact with *in vitro* host cells (hypothetical proteins excluded), and the percentage of over-expressed genes (percentages calculated has been calculated on total DEGs, including hypothetical proteins) are indicated in brackets next to the strain name (red = 30'; blue = 90').
The heatmap **B** shows percentages relating to COG gene classes of total DEGs detected in virulence conditions at 30' and 90'. The heatmap **C** shows differentially expressed genes of *A. butzleri* after 30' and 90' of contact with host cells and considered *A. butzleri* putative virulence genes from functional gene annotation analysis. The organization of *tonB* operon is shown in figure **D**, the times shown in the table (30 'and 90') indicate the overexpression (green) or underexpression (red) of the indicated gene. The

promoter detection with BPROM software indicates a putative promoter structure 21 bp upstream of *exbB* gene.

Figure 3.4. Acetic and lactic acid concentration and related DEGs. The heatmap (**A**) shows lactate and acetate-related DEGs of *A. butzleri* in contact with host cells (*p*-value < *0.05*, FDR corrected). The association of *actP* and *yjcH* gene copies is shown in figure **B**, the times shown (30 'and 90') indicate the relative overexpression (green) of the indicated gene. The other component of this operon *acs* gene present in *E. coli*, was found to be overexpressed in the LMG 11119 strain but is not present adjacent to *actP* and *yjcH*. The bar chart (**C**) shows concentrations in DMEM of acetic and lactic acid in µM. In the figure are indicated the strain codes and sampling times 0' (after 2h of adaptation), 30' and 90'. The different conditions are indicated by different colors (color codes below figure) while the lines above the bars of the graph show the analysis of DMEM from host cells (blue; C), bacteria (yellow; B) and bacteria in contact with the host cell (green; B+C). The error bars represent the standard errors, *p*-value with statistical differences between conditions are indicated near bars with different colors of the letters indicate different comparisons.

Supplementary Table 3.1. *A. butzleri in vitro* test log CFU cm^{-2} values. The data in this table represent bacteria load of bacterial inoculum (T0), bacteria detected after PBS washing (T1) and bacteria detected after gentamicin application (T2). T1 and T2 loads have been detected at 30' and 90'. In table are shown the corresponding standard errors while "ND" indicates a not detectable load.

Supplementary Table 3.2. *A. butzleri* DMEM DEGs from the comparison with Arcobacter agar. The table shows logFC (< -1.5, > 1.5) logCPM, *p-value* (< 0.05) and FDR (< 0.05) values of differentially expressed genes linked to currently considered putative virulence genes. The column "gene" shows the protein name of the DEGs and the relative locus tag.

Supplementary Table 3.3. *A. butzleri* DEGs after 30' of contact with host cells. The table shows logFC (< -1.5, > 1.5), *p value* (< 0.05) values of differentially expressed genes linked to the currently considered putative virulence genes. The column under the strains names shows the protein name of the DEGs and the relative locus tag.

Supplementary Table 3.4. *A. butzleri* DEGs after 90' of contact with host cells. The table shows logFC (< -1.5, > 1.5), *p value* (< 0.05) values of differentially expressed genes linked to the currently considered putative virulence genes. The column under the strains names shows the protein name of the DEGs and the relative locus tag.

Supplementary Table 3.5. *A. butzleri* pyruvic acid and glucose related DEGs after 2 h of incubation in DMEM. The table shows logCPM, *p value* (< 0.05) and FDR (< 0.05) values of differentially expressed genes linked to pyruvate and glucose. The column "gene" shows the gene number and his locus tag. The logFC values indicated result higher than 1.5 or lower than 1.5 (except for LMG 11119 locus tag 01408 and 02015, logFC > 1.40).

Supplementary Table 3.6. Annotation statistics of the 3 *A. butzleri* strains. In the table are indicated the genome size, coverage ((read count * read length)/genome size), number of total genes, number of total CDS, tRNA and hypothetical proteins number.

Supplementary figure 3.1. Glucose and pyruvic acid concentration of DMEM inoculated with *A. butzleri* strains. The bar chart shows the concentrations in DMEM of glucose and pyruvic acid (µM). In the figure are indicated the strain codes and control (normal DMEM, C). The different sampling times are indicated as 0' (after acclimation), 30' and 90'. The error bars represent the standard errors (Past3). The figure shows statistical analysis *p*-value. The statistical differences between strains are indicated near bars.

4. *Arcobacteraceae* pangenome analysis demonstrates genomes heterogeneity and reduction in genome size of species isolated from animals and humans

Table 4.1. *Arcobacteraceae* species information. In this table the strains codes of the 20 species belonging to the *Arcobacteraceae* family and two outgroups sequences that were used are also shown. The codes about Sequence Read Archive (SRA) are shown. Accession numbers indicated with "*" have been retrieved from ENA (https://www.ebi.ac.uk/ena/browser/home), the other strains sequenced in this work are available at the NCBI bioproject PRJNA808439. The column "group" reports the group of strains as indicated by Pérez-Cataluña and colleagues [1].

Figure 4.1. Anvi'o Analysis. The figure shows genomes characteristics of the *Arcobacteraceae* species object of study sorted for gene cluster presence and absence. In this figure are observable the different sources of isolation and the different species groups as well as general information regarding the genomes (size, GC content).

Figure 4.2. Dendrograms of *Arcobacteraceae* species studied. The trees in the figure have been computed from different input sequences: amino acid annotated sequences, best bcgTree (**A**), tree inferred from all orthogroups (**B**) and 16S rRNA sequences (**C**). The groups of species from Pérez-Cataluña and colleagues are indicated by branches colors as indicated in the figure [1].

Figure 4.3. Pangenome analysis of the *Arcobacteraceae* species. The figure shows different data about pangenome analysis performed on 21 isolates belonging *Arcobacteraceae* family. The histogram **(A)** shows gene families frequency, shell and persistent genes are indicated respectively by green and orange bar on the right. Figure **B** shows the presence of multiple genes copies (blue = present, red = absent), it's possible to observe that most of the multiple copies are in the lower part of the graph in correspondence with the genomic portion relating to the core and shell genome. The histogram **C** shows the percentage of the different persistent OGs and gene pathway families, below this figure are showed the different gene partitions obtained by different tools **(D)**.

Figure 4.4. Genomes Annotated Pathway. The heatmap (A) shows the pathway modules with at least 50% of completeness in at least one genome. The bar chart (**B**) shows the pathway modules with at least 80% of completeness in at least one genome.

Figure 4.5. PVGs presence-absence dendrogram. The figure shows the dendrogram produced on the presence-absence binary matrix of nine genes currently considered virulence-related. The tree (Jaccard, Neighbor-joining) shows bootstrap 10.000 value while the different colors near species names indicate the species belonging group.

Supplementary table 1. General information about *Arcobacteraceae* genomes. In the table are indicated different information about genomes general information (Quast), coverage (reads nr. * reads length)/genome size) and CRISPR/CAS sequences number. In the first line under the name of the strains the group to which they belong is indicated.

Supplementary table 4.2. Clusters of Orthologous Genes functions number. The table shows the number of annotated orthogroups COGs obtained from EggNOG mapper analysis. At the end of the table are indicated the coding of the letters relating to the different classes.

Supplementary table 4.3. Presence/absence of secondary metabolites related genes. The table shows presence (1) and absence (0) of secondary metabolites related sequences in the 21 genomes object of study.

Supplementary figure 4.1. Orthogroups pathway present in all genomes. The figure shows the number of COGs codes relative to orthogroups present in all *Arcobacteraceae* species object of study.

1. Literature overview

1.1. Introduction

The *Arcobacteraceae* family includes Gram-negative bacteria isolated from different environmental matrices such as sewage, oil production, environments, marine sediments, estuarine and river waters, oysters, snails, tube worms (abyssal annelid), oysters and fish farms [1,2]. However, species belonging to *Arcobacteraceae* have also been isolated from terrestrial animals and in particular from poultry, pigs, cattle, and derived foods. Furthermore, some species have been isolate form stool of diarrheic human patients, as well as from cases of septicemia [2]. Last few years have been characterized by an increase of interest about the *Arcobacteraceae* bacterial family, and in particular to *Arcobacter butzleri*, which has been frequently isolated from different clinical human cases and from animals [3].

The first *Arcobacter* species was isolated in 1983, from the roots of *Spartina alterniflora.* This bacterium was a nitrogen-fixing Gram-negative first included in the *Campylobacter* genus, and called *Campylobacter nitrofigilis*[4]. The discovery of *C. nitrofigilis* was followed by the isolation from animal abortions of the heterogeneous species *Campylobacter cryaerophila* and by the characterization of *Campylobacter butzleri* [5]. The creation of the genus "*Arcobacter"* was proposed by Vandamme and colleagues in 1991 to include *C. nitrofigilis* (*Arcobacter nitrofigilis*), *C. cryaerophila* (*Arcobacter cryaerophilus*) and in the following year, *C. butzleri* (*Arcobacter butzleri*) and *Arcobacter skirrowii* in a single new genus [6,7]. Nowadays, 40 species are included, but their taxonomic subdivision is still under discussion (Taxonomy Browser, NCBI 2022) [8].

1.1.1 Taxonomy

The taxonomy has been affected by some changes in recent years. This family *Arcobacteraceae* was proposed in 2017 to group species belonging to the genus *Arcobacter*, previously included in *Campylobacteraceae* family. The family *Arcobacteraceae* has been proposed together with the inclusion of the *Epsilonproteobacteria* in the novel phylum Epsilonbacteraeota (recently repurposed as Campylobacterota) [8]. A further division of the *Arcobacter* genus has been proposed by Pérez-Cataluña and colleagues after analysis by Multilocus Sequence Analysis (MLSA, 13 housekeeping genes) and the analysis of 286 core genes to obtain different genomic indexes. These indexes were *in silico* DNA–DNA hybridization (*is*DDH), Average Amino-acid Identity (AAI), Percentage of Conserved Proteins (POCPs), Average Nucleotide Identity (ANI), and Relative Synonymous Codon Usage (RSCU) [9] (Figure 1.1). The results led to split the former genus *Arcobacter* spp. in six bacterial genera, called *Arcobacter*, *Aliarcobacter*, *Malacobacter*, *Haloarcobacter*., *Poseidonibacter*, *Pseudoarcobacter*, and the candidate genus *Arcomarinus* [9]. The genus *Aliarcobacter* contains the species *Aliarcobacter butzleri*, *Aliarcobacter cibarius*, *Aliarcobacter cryaerophilus*, *Aliarcobacter skirrowii* and *Aliarcobacter thereius*. These *Aliarcobacter* species are considered to be of great clinical and veterinary interest, are associated with mammals, and considered food-borne pathogens [2].

In the species *A. cryaerophilus*, a phylogenetic peculiarity from ANI, MLPA and *is*DDH analyses has been observed. This species has been originally divided into two subspecies, subgroup 1A and 1B, though without any biological or clinical relevance, and now the division into four

subgroups has been proposed, named *A. cryaerophilus gv. pseudocryaerophilus* (Cluster I, LMG 10229^T), *A. cryaerophilus gv. crypticus* (Cluster II, LMG 9065^T), *A. cryaerophilus gv. cryaerophilus* (Cluster III, LMG 24291^T) and *A. cryaerophilus gv. occultus* (Cluster IV, LMG 29976^T) [10].

The division of the genus *Arcobacter* into several new bacterial genera is not fully accepted, and strongly contested by some authors which propose the maintenance of a single genus [6,11,12]. The existence of only one bacterial genus within this bacterial family is also supported by the data obtained and presented in part of this thesis. For these reason, the term "*Arcobacter*" used in the thesis will refer to all species belonging to the family *Arcobacteraceae*. In any case, bacteria belonging to the *Arcobacteraceae* family appear to be heterogeneous in physiological and genomic terms. These characteristics will be discussed in the next paragraphs.

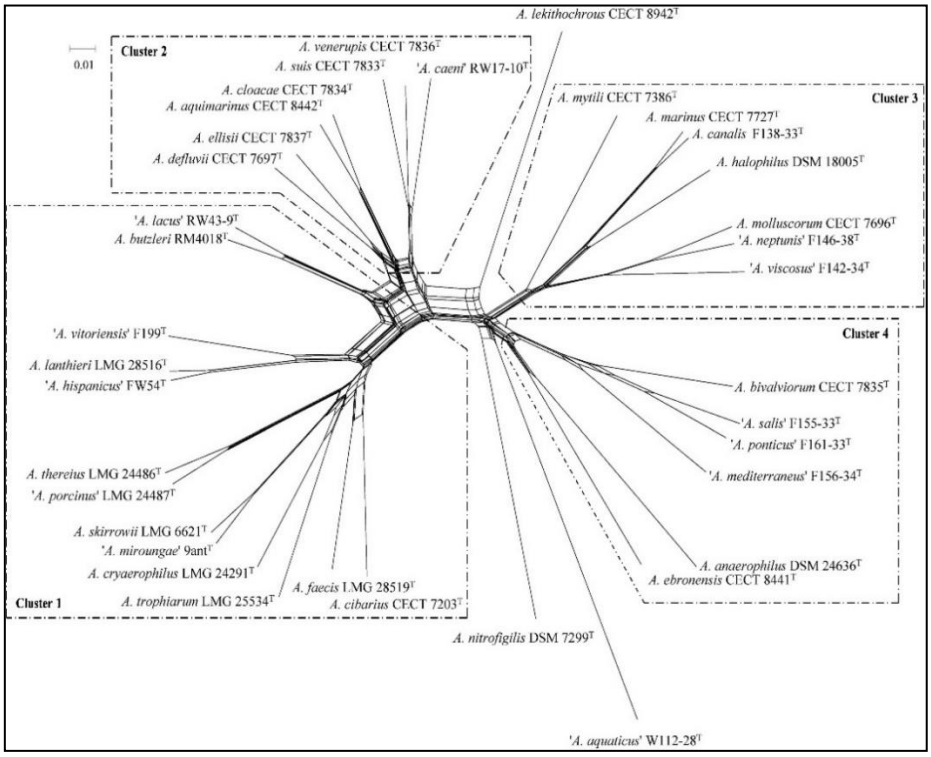


Figure 1.1. Network constructed with 284 core genes concatenated sequences. The core genes analyzed come from the genomes of 36 type strains studied object of study in the work of Pérez-Cataluña and colleagues in which has been proposed the division of *Arcobacter* spp. in different genera has been proposed (scale bar, base substitutions per site) [9].

1.1.2 General metabolism and morphology characteristics

Arcobacter spp. are Gram-negative bacteria with a single flagellum and grow on blood agar medium as small white/clear colonies (2-4 mm in diameter). The cellular morphology is curved rods- generally 0.2–0.9 μm wide and 1–3 μm long [2].

Unlike *Campylobacter* spp., *Arcobacter* spp. can tolerate oxygen. For the first isolation of *Arcobacter* spp., a microaerophilic condition at a temperature between 30 and 37°C is generally necessary. *A. butzleri* and

A. cryaerophilus can tolerate a pH range of 5.5 to 9.5 (optimal 6.8–8.0) [13]. The metabolism of *Arcobacter* spp., exploits various amino acids and organic acids as a source of energy. Most of the bacteria of this genus test positive to the catalase test and the indoxyl acetate hydrolysis and nitrate reduction. Generally *Arcobacter* spp. strains are negative to urease test, hippuric acid hydrolysis and the production of hydrogen sulphide from TSI (Triple Sugar Iron Agar) [2].

1.1.3 Genome characteristics

Whole-genome sequences of several *Arcobacter* species have become available in the last years. The complete genomes of *A. butzleri* (RM4018) and *A. nitrofigilis* (DSM 7299^{T}) are frequently used for the various comparative genomic analyzes. The *A. butzleri* genome has a size of 2.34 Megabases (Mb), with 2259 coding sequences (CDS) where about half of them are with an unknown function. The study of *A. butzleri* genome leads to the hypothesis that this bacterium does not require a host to develop, and several genes related to sulfur metabolism have been identified. These genes are typical of bacteria that live free in the environment [14]. The genome of *A. nitrofigilis* is of 3.19 Mb with four 16S rRNA operons with differences of more than two nucleotides. This indicates that *Arcobacter* spp. shows microheterogeneity at the level of this genomic portion [15]. As previously stated, genome data and raw sequences of *Arcobacter* spp. are available for other species, although comparative analyzes between whole genomes of different species has not been performed so far. These analyzes are subject of part of this thesis.

1.1.4 Prevalence in animals

Arcobacter spp. species have been isolated from farm and domestic animals, such as cattle, pigs, sheep, horses, dogs, chickens, and from reptiles such as lizards, and snakes [2]. Pérez-Cataluña and colleagues have collected the information of 25 *Arcobacteraceae* species isolated from land and marine animals [9]. *Arcobacter* spp. usually don't cause disease in animals [2,16]. However, some cases of symptoms associated with *Arcobacter* spp. infections have been reported.

A. cryaerophilus is the predominant species of *Arcobacter* spp. isolated from cases of pig abortions [17]. The isolation of *A. cryaerophilus* from milk produced by a herd affected by mastitis has been reported by Logan and colleagues [18]. These *A. cryaerophilus* strain were used to infect four dairy cows that consequently developed mastitis, but healed spontaneously after five days [18]. Kerkhof and colleagues isolated *Arcobacter* spp. from 76 fecal samples of wild boars (*Sus scrofa*) [19]. The species isolated were *A. cryaerophilus*, *A. butzleri* and *A. skirrowii*, respectively in 9.21 %, 1.32 % and 1.32 % of the samples.

A. butzleri has been associated with enteritis, with symptoms of diarrhea in cattle, pigs, and horses [7,16]. *A. butzleri* has been isolated from fecal samples of chickens, turkeys, ducks and domestic geese [20,21]. In a research performed in Turkey, *A. butzleri* was the species most frequently isolated from chickens, geese, ducks, turkeys, and quails followed by *A. cryaerophilus, A. skirrowii* and *A. cibarius* [22]. The species *Arcobacter thereius* has been isolated from pigs and ducks in Belgium [23]. *A. skirrowii* has been associated with diarrhea and hemorrhagic colitis in cattle and sheep [7,16]

In two studies, *A. butzleri* was the only species of the genus *Arcobacter* isolated from primates [24,25]. In one of these studies, histological analysis of enteric tissues revealed chronic active colitis [25].

Arcobacter spp. strains have been detected in cats and dogs [26]. The isolation index of *Arcobacter* spp. from dogs and cats ranges from 78.8 % of the samples tested to no isolation in cats [26,27]. *Arcobacter* spp. was isolated up to a maximum of 54.4 % of fecal dog samples analyzed [28,29]. *A. butzleri* is the most frequently isolated species in cat and dogs followed by *A. cryaerophilus* [26].

The isolation of *Arcobacter* species has been obtained also from aquatic animals. *A. cryaerophilus* has been isolated from rainbow trout, *Oncorhynchus mykiss,* and its pathogenicity has been demonstrated in healthy trout artificially infected. The fish died after *A. cryaerophilus* infection with damage to their liver, lungs and intestines [30]. *A. butzleri* and *A. cryaerophilus* have been isolated from mussels and clams leading to speculate that their derived foods are a source of *Arcobacter* spp. infection for humans [31]. The species *Arcobacter bivalviorum, Arcobacter mytili, Arcobacter molluscorum* and *Arcobacter venerupis* have been detected in shellfish [32–34].

These information shows that *Arcobacter* spp. is present in animals from different ecological niches.

1.1.5 *Arcobacter* species in food and water

Arcobacter spp. have been isolated from different foods of animal origin. This aspect has led to consider some *Arcobacter* species as food-borne pathogens [2], with meat as one of the most contaminated [35,36]. The percentages of contaminated food vary according to the area of origin [2].

Arcobacter spp., and in particular of *A. butzleri*, is able to survive at temperatures of 4-10°C, typical meat storage temperatures [35,37]. This survival ability of *Arcobacter* spp. suggests possible risks for meat consumers.

A study carried out in Ireland investigated the presence of *Arcobacter* spp. in food of animal origin and applied both phenotypic and genotypic identification (genus-specific and species-specific PCR). In that study,404 samples of raw milk (n = 101), raw meat of chicken (n = 94), bovine (n = 108) and pork (n = 101) were analyzed. The most contaminated samples were raw chicken meat (62%) followed by raw milk (46%), raw pork (35%), and raw beef (34%) [38]. Shah and colleagues in 2012 detected *A. butzleri* and *A. cryaerophilus* in bovine milk, while in goat's milk *Arcobacter* spp. was not detected [39]. A study carried out in an Italian artisanal dairy plant led to isolate *A. butzleri* [40]. Subsequent studies have shown that *A. butzleri* can survive at 4-10 °C for six days in the water of the artisan buffalo *mozzarella* cheese [41,42]. These aspects show how contamination by *Arcobacter* spp. represent a problem for the meat and dairy supply chain and products. Giacometti and colleagues in 2015 performed a study focused on *Arcobacter* spp. contamination in samples from dairy animals, production environment (e.g. water, feed) and milk [43]. *Arcobacter* spp. were found in the 22.6% of the samples, with a wide presence of *A. butzleri*. Subsequently, the different ways of contamination of final products were investigated by Pulsed-Field Gel Electrophoresis (PFGE). Strains with the same PFGE profile were found in different samples among which environment, animal gut, water, milk processing plant and milk. The authors speculated that fecal contamination was not the only source of contamination. The water, animal stable environment,

and processing plants can contribute to the milk contamination (cross-contamination) [43].

Regarding vegetables, *Arcobacter* spp. have been detected on lettuces [44], rocket [45], napa cabbage, water parsley [46] and ready-to-eat salad [47]. The species most isolated from vegetables is *A. butzleri* while *A. cryaerophilus* has been isolated from leafy green vegetables [3,46]. *A. butzleri* is able to survive to apple and pear purees production process but with a significant bacterial load reduction [48].

Species belonging to *Arcobacter* spp. have also been associated with marine environments. Furthermore, the presence of these species in water environments has been observed in some water animals and in the respective derived foods. *Arcobacter* species have been isolated from seafood like oysters, fish, shellfish, and clams [2,31]. *A. butzleri* has been isolated from mussels, razor clams, clams, shrimps, squids, octopuses and fish [3]. The genome analysis of *A. butzleri* isolates from clams (*Tapes philippinarum*) and mussels (*Mytilus galloprovincialis*) showed the presence of putative virulence genes and antibiotic resistance genes. The presence of these genes suggests that these food contaminations represent a public health risk [49].

An interesting case of food outbreak related to *A. butzleri* and *A. cryaerophilus* has been observed in patients with diarrhea after the consumption of fried chicken. The detection of *Arcobacter* spp. was carried out by PCR analysis of the 16S/23S rDNA and the *rpoB*/*C* sequences [50]. This study demonstrated the relevance of *Arcobacter* spp. infections caused by contaminated food. Moreover, this case of study indicates that PCR analyzes can be an important tool for *Arcobacter* spp. detection.

Water is one of the main sources of *Arcobacter* spp.. *A. butzleri* and *A. cryaerophilus* have been isolated from wastewater treatment plants [51], while the obligate halophile *Arcobacter halophilus* has been isolated from saline water [52]. The species *Arcobacter marinus* and *Arcobacter pacificus* have been isolated for the first time in seawater [53,54]. A study about Mediterranean seawater allowed the detection of *A. butzleri*, *A. cryaerophilus* and *A. skirrowii*. The detection of these species was carried out through 16S and 23S rRNA genes PCR amplification. In this study, it was possible to isolate *A. butzleri* from plankton and seawater [55].

A. butzleri showed the ability to adhere to water distribution pipe surfaces [56]. The presence of *Arcobacter* spp. in water and water environments indicates the significance of this environmental element for *Arcobacter* spp. and for its diffusion [2].

Due to the high presence of *Arcobacter* spp. in food, prevention procedures are essential to reduce the risk of *Arcobacter* spp. infections. For this purpose a proper cooking together with suitable preservation techniques are important to reduce the risk of infections [2,35]. *A. butzleri* is more resistant to radiation and pasteurization than *C. jejuni* [57]. The growth of *A, butzleri* is inhibited by citric and lactic acid (1-2%) and sodium lactate (2%). The nisin has been shown to inhibit the *A. butzleri* growth at a concentration of 500 IU ml^{-1} after 5 h. [57,58].

The ability of *Arcobacter* spp. to survive in the environment and therefore on the surfaces of food production plants requires suitable hygienic procedures to reduce the risk of contamination [42,56].

1.1.6 Human infections

The first *Arcobacter* species associated with human disease was *A. cryaerophilus*, soon followed by *A. butzleri*.

A. butzleri and *A. cryaerophilus* are mainly associated to gastrointestinal disorders [59], though cases of septicemia have also been reported. However, these two species are not the only ones associated with clinical cases: *A. lanthieri* has been recently isolated from a patient with persistent abdominal bloating and cramps. This *A. lanthieri* strain was characterized by the presence of five putative virulence genes (*cadF*, *ciaB*, *mviN*, *hecA* and *iroE*) indicating a possible pathogenic potential [60]. *Arcobacter* was simultaneously present with *H. pylori* and *Campylobacter* spp. in South African patients [61]. *A. butzleri* was the third prevalent pathogen in fecal samples from South African patients (6.2%), preceded by *Helicobacter pylori* (50.6%) and *C. jejuni* (10.2%). The species *A. cryaerophilus* (2.8%) and *A. skirrowii* (1.9%) were detected in a lower percentage compared with the *A. butzleri* percentage. *A. butzleri* was found to be the fourth most frequent species isolated from patients with diarrhea in Belgium and France [62,63]. *Arcobacter butzleri* was found to be present in 8% of US and European travelers with diarrhea from Guatemala, India and Mexico [64]. As previously explained, *A. butzleri* and *A. cryaerophilus* are considered the most relevant *Arcobacter* species in terms of pathogenicity. Some clinical cases information reviewed by Figueras and colleagues are shown in table 1.1 [59]. Furthermore, the study described a severe case of diarrhea in a 26 year old male associated with *A. cryaerophilus*, identified by MALDI-TOF and *rpoB* gene sequencing analyses. The *ciaB* invasion-related gene was present in this *A. cryaerophilus* strain[59].

A. butzleri represents the most isolated species in clinical cases related to *Arcobacter* spp. infections [2,59]. A study regarding the pathogenicity of *A. butzleri* was conducted by Vandamme and colleagues where different symptoms were observed with a prevalence of abdominal pain in ten Italian children [65]. The children didn't show diarrhea, a symptom most frequently observed in cases of *A. butzleri* infections (Table 1.2). In a study in Denmark, a similar clinical condition related to *A. butzleri* in a patient that showed fever was observed, with general malaise and erythema without diarrhea [66]. These reports indicate that the onset of various symptoms is possible, although *Arcobacter* spp. has been frequently associated with gastrointestinal disorders [59].

Table 1.1. Main symptoms related to cases of *Arcobacter* spp. infection. In the table are indicated gastrointestinal and non-gastrointestinal symptoms associated to *Arcobacter* spp. infections [59].

sex/age	Country	Symptoms	Species	ref
m/35	Australia	Chronic diarrhea (6 months)	*A. cryaerophilus*	[67]
1. m/48 2. f/52	Germany	1. Acute liquid diarrhea (15 days), abdominal cramps 2. Chronic diarrhea (3 weeks), abdominal cramps	*A. butzleri*	[68]
m/2 f/1	Chile	1. Acute diarrhea and vomiting 2. Chronic diarrhea with cramps and abdominal pain	*A. butzleri*	[69]
m/73	Belgium	Chronic diarrhea (2 months)	*A. skirrowii*	[70]

m/30	Turkey	Acute liquid diarrhea. abdominal pain, nausea, abnormal sweating	*A. butzleri*	[71]
m/26	Spain	Watery bloody diarrhea (3 months)	*A. cryaerophilus*	[59]
neonate	UK	Bacteriosis with hypertension, hypothermia, and hypoglycemia	*A. butzleri*	[72]
m/72	Taiwan	Bacteriosis, hematogenous pneumonia	*A. cryaerophilus*	[73]
Between 60-70	Denmark	Fever, general malaise and erythema of	*A. butzleri*	[66]
m/85		Fever (39.3 °), hypotension, diffuse maculopapular skin rash, and diarrhea.	*A. butzleri*	[74]
f/69	Hong Kong	Bacteriosis with fever and diffuse mild pains	*A. butzleri*	[75]
f/63	China	Peritonitis after catheter repositioning with fever and abdominal pain	*Arcobacter* spp.	[76]

Table 1.2. Information about *A. butzleri* infections collected in the study of Vandamme and colleagues in 1992. These patients lack the most common symptom, diarrhea. The patients were affected by other symptoms [65].

Sex/ age	Duration of symptoms (days)	Abdominal pain	Vomit	Fever	Hospitalization (days)
m/4	7	+	+	-	
f/7	10	+	-	-	8
m/7	8	+	-	-	
m/3	10	+	+	+	
f/7	8	+	-	-	8
f/7	5	+	-	-	

m/7	6	+	-	-	5
f/4	8	+	+	-	
f/4	7	+	-	-	
f/3	7	+	-	-	

1.1.7 Antibiotic resistance

The reported clinical cases linked to *Arcobacter* spp. were principally solved by antibiotic treatments. In the last years, the *Arcobacter* spp. antibiotic resistance has been reported. Several studies related to *A. butzleri* antibiotic resistance were performed [2,77].

A. butzleri and *A. cryaerophilus* strains isolated from humans in Belgium were found to be ampicillin resistant [78]. These strains were susceptible to gentamicin (99% of the strains) and tetracycline (89% of the strains), while erythromycin, ciprofloxacin and doxycycline were moderately effective [78]. A study in Thailand on 84 *A. butzleri* strains showed the absence of resistance to gentamicin, kanamycin, streptomycin, and tetracycline. Of the *A. butzleri* strains, 40.5% were resistant to nalidixic acid [79]. In 2016, Shirzad Aski and colleagues showed that *A. butzleri* and *A. cryaerophilus* isolates from cattle and sheep were generally resistant to rifampicin, vancomycin, ceftriaxone, trimethoprim and cephalothin. These isolates were highly susceptible to tetracycline, oxytetracycline, erythromycin, ciprofloxacin, kanamycin, amikacin, gentamicin and enrofloxacin [80]. A study on *A. butzleri* strains isolated from chicken carcasses, minced beef, quail carcass, turkey meat, cattle rectal swabs, cattle gallbladder, broiler cloacal swab, and from human showed antibiotic resistance to amoxicillin/clavulanic acid (20% of the strains), nalidixic acid (44.28%) and ampicillin (78.57%) while they were

susceptible to gentamycin [81]. The susceptibility to ciprofloxacin was observed in *A. butzleri*, *A. cryaerophilus* and *A. lanthieri* strains in Germany. The authors of this study suggested the use of ciprofloxacin in clinical cases [82]. Other authors suggested the use of tetracycline and aminoglycosides for the treatment of *Arcobacter* spp. infections in humans [2,81]. Recently, a high percentage of strains resistant to tetracycline and ciprofloxacin has been found [49,77,83,84]. Isidro and colleagues in 2020 carried out a study about *A. butzleri* 49 genomes allowed the detection of new sequences related to antibiotic resistance (Table 1.3) [77]. The resistance of *A. butzleri* to macrolides has been linked to the inactivation of a TetR repressor, while fluoroquinolones resistance has been correlated to a Thr-85-Ile substitution in GyrA protein. In this study the resistance of *A. butzleri* to ampicillin was linked to an OXA-15-like β-lactamase [77].
The different results reported in literature about the antibiotic resistance of *Arcobacter* species leads to an absence of a shared protocol related to antibiotic treatment however, the antibiotic treatment are generally not necessary in resolving infections related to *Arcobacter* spp. [2].

1.1.8 Virulence Factors

The pathogenesis and transmission of *A. butzleri* and *A. cryaerophilus* has been evaluated in several studies (Figure 1.2) [2]. The pathogenesis and transmission of these bacteria has been hypothesized by their entry through the digestive tract followed by gut colonization, following several mechanisms such as the production of cytotoxin, affecting the expression of tight junction proteins and the release of pro-inflammatory cytokines and epithelial barrier dysfunction leading to cell death.

Cell dysfunctions were observed in human cellular epithelial monolayers produced with the HT-29 and B6 cell lines in contact with *Arcobacter* spp. For these epithelial barriers, the most significant protein in the formation of the tissue are claudins. The claudins 1, 5 and 8 are characterized by a high sealing capacity. These claudins are not expressed in cells infected by *A. butzleri,* this phenomenon occurs also in T84 cells infected with *E. coli* [85]. Tsang and colleagues observed that the properties of hemagglutinin were associated with the interaction between *Arcobacter* spp. and host cells, promoting the adhesion of *Arcobacter* spp. [86]. The contact of host cells with *A. butzleri, A. cibarius, A. cryaerophilus* and *A. skirrowii* caused the release of interleukin-8 (pro-inflammatory signal molecule) by host cells. This aspect was also observed in response to other pathogen infection such as *Helicobacter* spp. and *Campylobacter* spp. [1,87]. Bruegge and colleagues demonstrated that pro-inflammatory response of infected macrophages towards *A. butzleri* was linked to the induction of interleukins IL-1, IL-1ß, IL-6, IL-8, IL-12ß and TNF [88]. A study in mice showed that *A. butzleri* induced extra-intestinal and systemic immune responses with increased TNF, IFN-γ, IL-12p70 and IL-6 levels in infected mice serum [89].

The presence of a flagellum is another feature linked to the pathogenesis of *Arcobacter*. Flagellins are considered virulence factors and a primary target for the host immune system [90]. The lipopolysaccharides (LPS) are considered possible virulence factor of *Arcobacter* spp. Moreover, LPS are important for the *A. halophilus* survival in marine environments [2].

Several sequences have been identified after genome sequencing of *A. butzleri* and considered as putative virulence genes [14]. Nine genes are currently considered linked to *Arcobacter* spp. virulence. These genes are

cadF (fibronectin binding protein encoding gene), *cj1349* (fibronectin binding protein encoding gene), *hecA* (member of filamentous haemagglutinin encoding gene), *hecB* (encoding gene of protein activation of hemolysin), *ciaB* (invasin encoding gene), *irgA* (iron-regulated external membrane protein encoding gene), *mviN* (virulence factor encoding gene), *pldA* (phospholipase encoding gene) and *tlyA* (hemolysin encoding gene) [91]. These genes are often used to determine the presence of pathogenic *Arcobacter* spp. strains (Table 1.3). In addition to these nine genes, the *iroE* sequence is also analyzed in some studies [60,92]. The genes mentioned are linked to different functions during pathogenesis. Genes *cadF* and *cj1349* encode membrane proteins that facilitate the connection of intestinal epithelial tissue with fibronectin by promoting bacterial adhesion. The gene *pldA* encodes a phospholipase present on the outer membrane while *ciaB* participates in the invasion of host cells. The gene *tlyA* encodes a hemolysin which hydrolyzes acyl ester bonds. The *tlyA, hecA* and *hecB* genes are considered involved in hemolysis. The *Arcobacter* spp. iron intake is attributed to *irgA* and *iroE* [79].

The study of the genomes made possible the detection of sequences considered putative virulence genes. A comparative study of 49 *A. butzleri* strains showed the presence of a vast arsenal of potential virulence factors [77]. The putative virulence genes detected were related to flagellar assembly/ function. Part of these genes (*flgD, flgL, flgK, flgE2, flgG2* and *flg*) were detected in a polymorphic region of the core genome of *A. butzleri* [77]. Among the putative virulence genes of *A. butzleri* were identified the urease cluster enrolling six genes (*ureD(AB)CEFG*),

chemotaxis related genes (*cheA-cheY*) and the complete type IV secretion system (T4SS) [77].

Fanelli and colleagues detected orthologues of the genes *waaC* (Lipopolysaccharide heptosyltransferase) and *waaF* (ADP-heptose–LPS heptosyltransferase) in *A. butzleri* [49]. These genes were described as putative virulence genes in *A. thereius* [93]. Gene *waaF* is involved in the biosynthesis of lipooligosaccharide (LOS), while *waaC* in lipopolysaccharide heptosyltransferase. The genes *waaC* catalyzes the transfer of sugar moieties from activated donor molecules to acceptor molecules, forming glycosidic bonds [49]. Fanelli and colleagues detected genes with different functions related to virulence among which flagellar assembly/function, type III secretion system (T3SS), hemolysis [49]. Moreover, is interesting the presence of virulence sensor protein BvgS precursor gene and sensor histidine kinase PhoQ gene detected by these authors in *A. butzleri* [49]. The whole-genome analysis allows the detection of 9 putative virulence genes (*cadF, cj1349, hecA, hecB, ciaB, irgA, mviN* and *tlyA*) frequently employed in PCR analysis in *Arcobacter* spp. [10,49,77,91,93]. The biofilm production is a factor linked to bacterial virulence and adhesion [94]. Martinez-Malaxetxebarria and colleagues confirmed the production of biofilm in *Arcobacter* spp. strains isolated from different foods [95]. The biofilm production was mainly observed in *A. butzleri* and *A. cryaerophilus* and has been linked to increased bacterial survival and adhesion to surfaces, increasing the risk of transmission. Moreover, these authors observed the presence of at least one putative virulence genes (PCR; *cadF, cj1349, hecA, hecB, ciaB, irgA, mviN* and *tlyA*) in *A. butzleri, A. cryaerophilus, A. skirrowii, A. thereius* and *A. vitoriensis* [95].

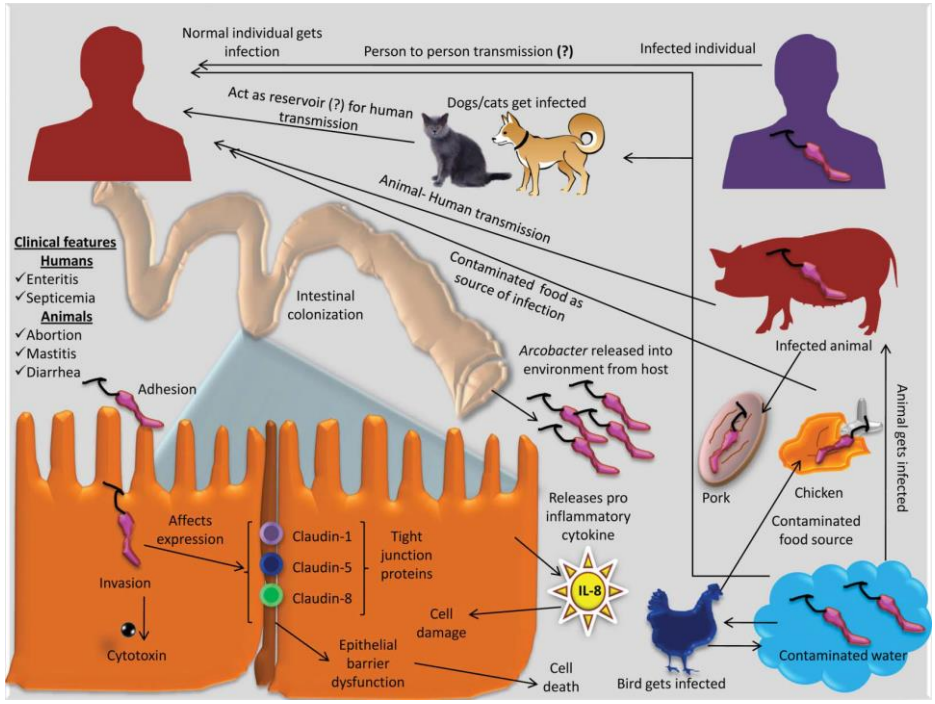


Figure 1.2. Pathogenesis and transmission of *Arcobacter* spp. The figure shows a pathogenesis and transmission model proposed by Rameese and colleagues [2].

Table 1.3. Presence of nine putative virulence genes in strains studied by Douidah and colleagues. The table shows the strain isolation origin: horse (H), chicken (C), pig (P), sheep (S), human (Hu) and dog (D). The table indicates the genes present (1) and absent (0) in the type strains of the species with reference codes. The number of strains in which the gene has been detected can be seen under the gene name. The column "tot" is the total number of strains tested [91].

	tot	*cadF*	*ciaB*	*cjl349*	*irgA*	*hecA*	*hecB*	*mviN*	*pldA*	*tlyA*
A. butzleri										
LMG 10828^T	1	1	1	1	1	1	1	1	1	1
Hu	78	78	78	78	27	16	53	78	78	78
C	36	36	36	36	9	8	16	36	36	36
P	33	33	33	33	4	6	14	33	33	33
C	29	29	29	29	14	17	25	29	29	29
S	2	2	2	2	0	0	2	2	2	2
H	2	2	2	2	0	0	1	2	2	2
D	2	2	2	2	0	0	0	2	2	2
A. cryaerophilus										
LMG 9904^T	1	1	1	0	0	1	0	0	1	1
Hu	22	14	20	16	1	2	8	22	13	12
C	34	5	33	11	1	1	0	32	1	5
P	23	6	21	15	1	0	1	21	6	14
C	9	4	8	3	0	1	0	7	5	4
S	6	0	5	3	0	0	0	5	0	0
H	1	1	1	0	0	0	0	1	1	0
D	4	4	4	3	0	0	1	4	0	0
A. skirrowii										
LMG 6621^T	1	1	1	0	1	0	1	1	1	1
Hu	1	1	1	1	0	0	0	1	1	1
P	13	2	12	2	0	1	3	1	2	5
C	23	4	23	2	0	0	5	8	3	1
S	1	1	1	1	0	0	0	0	1	1

1.1.9 *Arcobacteraceae* isolation

The first media used for the isolation of *Arcobacter* spp. was the semi-solid Ellinghausen-McCullough-Johnson-Harris (EMJH) media supplemented with 5-fluorouracil [2,96]. The media modified charcoal cefoperazone deoxycholate agar (mCCDA) with C.A.T supplement was used for the isolation of *Arcobacter* spp. from poultry [97]. The medium cefsulodin-irgasan- novobiocin (mCIN) agar with C.A.T supplement has been used for the isolation of *Arcobacter* spp. from poultry [98]. Houf and colleagues used five antimicrobials (5-Fluorouracil, novobiocin, trimethoprim, and teicoplanin or vancomycin) to be added to *Arcobacter* media (Oxoid, United Kingdom) which were found to be sufficiently selective for the isolation of *Arcobacter* species [99]. Other media are Arcobacter selective enrichment broth (ASB) and Arcobacter selective semisolid medium (ASM), formulated for the recovery of *Arcobacter* spp. from food. [100]. Johnson and Murano developed a medium for the isolation of *Arcobacter* spp., to distinguish from species related to *Campylobacter* spp. The *Arcobacter* spp. isolates assume a deep red color around the colonies [101].

Merga and colleagues performed a comparison between different *Arcobacter* spp. isolation methods. The results showed that the use of Houf broth followed by mCCDA-C.A.T plates was the method with a higher sensitivity (71%) and specificity (64%) for the isolation of *Arcobacter* spp. [96]. New isolation methods have been recently tested. These protocols include the use of salts during the isolation of *Arcobacter* spp. Sodium chloride and artificial seawater (ASW) were used to obtain a greater growth and specificity during the *Arcobacter* spp. isolation from marine environments [102,103]. The shellfish-associated *Arcobacter* species

growth was higher in *Arcobacter* broth with 50% of ASW or with lyophilized oysters [103].

The incubation conditions for *Arcobacter* isolation are normally characterized by microaerophilic atmosphere at 30°C for 48-72 h. In the media, horse or sheep blood at a percentage of 5% v/v is often added as supplement to promote *Arcobacter* spp. growth [2,22,96,97,99,103,104]. For the isolation of *Arcobacter* spp. the membrane filtration method has been used, showing a higher number of positive samples compared to methods without filtration [2,105,106].

1.1.10 *Arcobacteraceae* identification and genotyping

The phenotypical characteristics of *Arcobacter* spp. vary according to the species but also at strain level within the same species (Table 1.4) [1,35]. The limits of phenotypic identification of *Arcobacter* spp. due to the difficult growth of the isolates and their phenotypic variability led to the design of molecular based detection techniques such as PCR, multiplex PCR (m-PCR), real-time PCR, restriction fragment length polymorphism (RFLP), matrix-assisted laser desorption ionization mass spectrometry (MALDI-TOF MS), denaturing gradient gel electrophoresis PCR (DGGE PCR) and fluorescence *in situ* hybridization (FISH) [2].

Table 1.4. Phenotypical information of nine *Arcobacteraceae* species. The table shows phenotypical information of *A. thereius* (th), *A. butzleri* (bz), *A. cryaerophilus* (cr), *A. skirrowii* (sk), *A. cibarius* (cb), *A. nitrofigilis* (ni), *A. halophilus* (ha), *A. mytili* (my), *A. trophiarum* (tr). The different characteristics are indicated as positive (1, ≥ 95% positive strains) or negative (0, ≤ 11% positive strains) in the different species while the asterisk (*) indicates a range of positive strains of 12-94%. ND = not determined [1,35].

	th	**bz**	**cr**	**sk**	**cb**	**ni**	**ha**	**my**	**tr**
Growth 25°C	1	1	1	1	*	1	1	1	*
Growth 37°C	0	1	*	1	0	*	1	ND	0
Growth microa. 37°C	0	1	*	1	1	0	1	*	0
Growth NaCl 2% w/v	*	*	*	1	0	1	1	ND	*
Growth NaCl 4% w/v	0	0	0	1	0	1	1	1	0
Catalase	1	*	1	1	*	1	0	1	1
Oxidase	1	1	1	1	1	1	1	1	*
Urease	0	0	0	0	0	1	0	0	0
Nitrate reduc.	1	1	1	1	0	1	1	0	0
Indole acetate hydrolysis	1	1	1	1	1	1	1	0	1
Resistant to Ceforperazone (64 mg/l)	1	1	1	1	1	0	0	0	1

The MALDI-TOF MS technique made the identification possible of bacterial species by analyzing ribosomal proteins [107].

The molecular-based methods made the identification of different *Arcobacter* spp. species possible [2]. Houf and colleagues designed a PCR method based on 16S and 23S rRNA regions able to identify *A. butzleri, A. cryaerophilus* and *A. skirrowii* [108]. The *gyrA* gene is another target for the species-specific identification of *A. butzleri, A. cibarius* and *A. cryaerophilus, A. skirrowii*[109].

Quantitative PCR (Real-Time PCR) is another tool used for the detection of *Arcobacter* spp. The target sequences for real-time PCR analysis were 23s rRNA [110] and *rpoB/C* [111]. Wang and colleagues showed that the loop-mediated isothermal amplification (LAMP) technique was more sensitive than conventional and multiplex PCR [112].

Although there are several protocols for the identification of *Arcobacter* spp., the protocol designed by Houf and colleagues in 2000 is the most common [2,108,113].

The genotyping is pivotal to characterize different strains of *Arcobacter* spp. during ecological and epidemiological studies. Different molecular techniques have been designed to differentiate isolates at strain level. The most relevant techniques used for this purpose are RFLP (Restriction Fragment Length Polymorphism), AFLP (Amplified Fragment Length Polymorphism), RAPD (Random Amplification of Polymorphic DNA), ERIC-PCR (Enterobacterial Repetitive Intergenic Consensus PCR) and PFGE (Pulsed-field gel electrophoresis) [2]. The RAPD and ERIC-PCR are the most used techniques. RAPD is notoriously not highly repeatable while results from ERIC-PCR are considered more reliable [114]. Merga and colleagues consider the ERIC-PCR questionable [115], while it is

considered a convenient genomic fingerprinting solution for a large number of isolates by other authors [116]. González and colleagues found that PFGE was effective in the detection of epidemiological relationships among closely related *Arcobacter* spp. strains [117].

The increase of the use of DNA sequencing, leads to the use of multi-locus sequence typing (MLST) for the determination of the isolates sequence types (STs) [2]. This technique allows the *Arcobacter* strains genotyping through the sequencing of seven housekeeping genes (*aspA*, *atpA*, *glnA*, *gltA*, *glyA*, *pgm*, *tkt*). The sequences compared in database allow to determine the most similar ST (http://pubmlst.org/arcobacter/) [118]. In a study of De Cesare and colleagues, the results obtained from MLST and PFGE analysis of about 133 *A. butzleri* isolates led to reach similar conclusions suggesting the use of more techniques to obtain reliable results [119].

1.2 Nucleic acids sequencing

The nucleic acids sequencing has become common in recent years although highly specialized personnel is required. In the case of *Arcobacter* spp. these techniques have been used to characterize new species, to obtain information about specific sequences (e.g. 16S rRNA encoding genes) and to design new molecular techniques. The sequencing techniques are also employed to study the whole genome bacterial content [15,32,49,77,115,120].

The studies of this thesis relied on the use of Illumina sequencing. For this reason, the next paragraphs will illustrate the second generation sequencing methods and their use in the study of bacteria.

1.2.1 First-generation DNA sequencing

From the characterization of the DNA structure performed on the data produced by Rosalind Franklin and Maurice Wilkins by Watson and Crick in 1953, various DNA sequencing techniques have been developed. The DNA sequencings methods were employed to obtain the sequence relative to DNA fragments up to entire genomes [121,122]. Sanger and colleagues in 1972 [123] developed the first technique that allowed the obtainment of a complete protein-coding gene [124] and, subsequently, the first whole genome of a bacteriophage [125]. Sanger's method is based on the detection of radio-labeled partial-digestion fragments after two-dimensional fractionation to identify the DNA sequences[126]. In 1975-77 an improvement of this technique was carried out using single separation by polynucleotide length via electrophoresis through polyacrylamide gels [127,128]. In 1977 came the turning point that allowed the Sanger's method to be used efficiently and precisely with the

development of the ‘chain-termination’ or dideoxy technique [129]. The Maxam and Gilbert technique is different from Sanger’s technique, unlike Sanger’s method this sequencing technique doesn’t follow a sequencing-by-synthesis (SBS) principle [128]. Maxam and Gilbert technique takes advantage of the lack of the 3′ hydroxyl group that is required for the extension of DNA chains together with labeled nucleotides. The extension of the DNA strand will be interrupted allowing us to understand, which nucleotide is present in a certain position through the electrophoretic run on a polyacrylamide gel [129]. Sanger’s sequencing technique was then developed further to make it more efficient. These improvements were obtained with the use of fluorescent molecules to reduce the number of electrophoretic runs required, and the use of PCR technique [126].

1.2.2 Second generation DNA sequencing

The pyrosequencing is one second generation sequencing technique that exploits the luminescence produced by a process that uses two enzymes. The pyrophosphate released during DNA synthesis is converted in ATP (Adenosine triphosphate) using ATP sulfurylase during the sequencing. The ATP is then used as luciferase substrate producing light [130] (Figure 1.3). The first machine conceived with pyrosequencing is the 454-Roche. The use in this technique of a water-in-oil emulsion PCR (emPCR) allows the production of fragments of 400 - 500 bp [126]. In 454 sequencing the DNA attached to beads placed onto a slide, is amplified by PCR. Deoxyribonucleosides triphosphate (dNTPs) and DNA polymerase are then washed over. If the dNTP is complementary, there is a pyrophosphate release through the DNA polymerase action detected through enzymatic conversion to ATP by luciferase with the emission of light. After the light

measurement, the sequencing cycle is repeated [131]. The pyrosequencing is a sequencing by synthesis (SBS) technique, it does not require the use of particular nucleotides (labeled in the previous techniques) and allows to observe a "real-time" sequencing [132,133]. From this method, other methods have been developed.

The Solexa machine uses flowcells, to bind DNA through adapters. In this method, the reversible-terminators dNTPs marked with fluorescence are used to obtain the DNA sequence. The sequencing is performed by the temporary interruption of the fragment which allows the identification of the nucleotide in a certain position. After this step the fluorophore in 3′ hydroxyl position that blocks the polymerization is cleaved away to allow the sequencing continuation. The DNA replication is performed by "bridge amplification" method in which the DNA strands must arch to initiate the next step of polymerization. This technique made possible the obtainment of paired-end (PE) sequences increasing the accuracy of sequencing [134]. Illumina Miseq and Hiseq technologies are among the most used. Hiseq allows a greater read length and depth, while Miseq is a lower-throughput machine with faster turnaround, longer read lengths and lower cost [126,135–137]. In the Illumina methods a part of the dNTPs are dyed to interrupt the PCR process (3′ blocked labelled nucleotides) and to detect the nucleotide present in a specific position. The DNA is fixed and amplified by PCR. Each spot contains multiple DNA fragment copies, after this step the DNA is made single-stranded and deoxynucleotides. The single DNA molecules cannot be PCR extended for the addition of a dye that blocks the fragment extension. This dye is washed over the slide, along with DNA polymerase, while nucleotides that bind the complement fragment are detected visualizing the dye. During sequencing cycles, the

dye and DNA polymerase will be washed over, and fresh-labelled (dNTPs) are added to continue with the extension of the DNA fragment [131].

Other second-generation sequencing technologies are SOLiD (sequencing by ligation) [138] and Ion Torrent which performs the DNA sequencing measuring the ions H^+ release instead of the emission of light measurement of 454 machines [139]. In the Ion Torrent sequencing method, the DNA is attached to a well base and amplified as per other sequencing techniques. If the dNTP is complementary, a hydrogen ion is released, and the pH change is detected. After the pH detection, a new cycle follows [131].

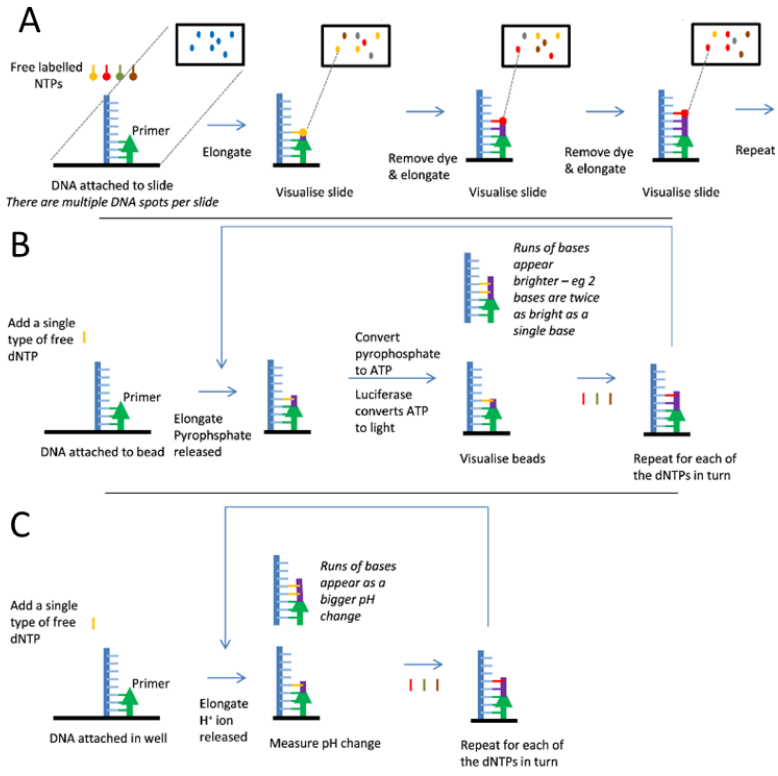


Figure 1.3. **Second generation DNA Sequencing methods mechanism**. This figure adapted from Robert and colleagues shows the different sequencing methods mentioned in the text: (**A**) Illumina; (**B**) 454 sequencing; (**C**) Ion Torrent [131].

1.2.3 Third generation DNA sequencing

One of the most used third-generation sequencing techniques is based on single-molecule sequencing (SMS) in real-time [140,141]. The method called Single Molecule Real-Time (SMRT) sequencing is produced by PacBio in which DNA polymerization is performed in microfabricated nanostructures arrays called zero-mode waveguides (ZMWs). ZMWs are tiny holes in a metallic film that cover a chip. The holes exploit the light properties passing through small apertures, which causes to decay exponentially illuminating the bottom of the well. The light allows the visualization of single fluorophore molecules near the bottom of the ZMW (the laser excitation zone is small, even against the background of neighboring molecules in solution). The deposition of single DNA polymerase molecules in the ZMWs puts them in the illuminated region. By washing the extension of the DNA chains by single nucleotides the sequencing can be monitored in real-time Only the incorporated fluorescent nucleotide will provide a detectable fluorescence, at this point the dye is eliminated, also eliminating the signal for that position [126,141,142]. This type of sequencing is useful for the *de novo* assembly of genomes as it can produce reads longer than 10 kb in length [126]. Another third-generation sequencing technique is Nanopore sequencing [141]. This method uses genetically modified bacterial nanopores into an artificial lipid bilayer, in turn, inserted in individual microwells (tens of micrometers wide) and arrayed on a sensor chip. The electric current running through the pore is interrupted by each single DNA strand that travels through a channel. The electric change is measured by a semiconductor sensor. Since each base disrupts the electric field in a slightly different way the recorded current changes can be translated into

a DNA sequence, this because each nucleotide base changes the electric field in a different way. The reads length is of 10–20 kbp, while was possible to obtain a molecules with a length greater than 2 Mbp long [141,143]. ONT reads are the longest DNA reads generated so far, allowing to sequence repeats where even PacBio may fail.

The SMRT and Nanopore sequencing return similar results, with long reads of low base-calling quality (Table 1.5), the different techniques have advantages and disadvantages linked to the methodology itself.

Table 1.5. DNA sequencing technologies. The information in the table regarding some sequencing technologies currently in use have been collected by Bansal and colleagues. From this information it is possible observe the differences between DNA sequencing methods in terms of reads length, error rate and throughput [137].

	method	**Read length**	**Error Rate**	**Throughput (Gigabases/run)**
Oxford Nanopore MinION	Nanopore	up to 1,000 kb (variable)	5-20 %	5-10
Pacific Biosciences SMRT	synthesis	10-100 kb	5-20 %	10-20
Illumina	synthesis	100-300	0.1 %	200-600

1.2.4 Bacterial DNA and cDNA sequencing

A wide range of DNA sequencing technologies is used in several scientific fields including microbiology. The first whole genomes obtained are bacterial (*Mycoplasma genitalium* and *Haemophilus influenzae*) [144]. The increase in computing power, the design of new bioinformatics

software, and the decrease in the sequencing cost, allowed the common use of DNA sequencing in several laboratories. The DNA sequencing has been employed for the analysis of specific sequences (e.g., MLST, 16S/23S cDNA sequences) from single and from multiple microorganisms present in ecological studies. Moreover, the sequencing of whole genomes allows the obtainment of information related to bacterial genes [145,146]. In the recent year studies about *Arcobacter* spp. allowed the obtainment of the whole genomes of bacteria isolated from different environments and hosts. The *A. butzleri* genome is the most studied of the genus *Arcobacter* in terms of functional annotation [49,77].

The raw sequences obtained from bacterial DNA sequencing are analyzed with different bioinformatics tools. The first step of these analyses is the sequence reads quality evaluation. This step is followed by reads "cleaning" through programs (e.g. fastP) that allows the elimination of sequencing adapters and low-quality sequences [147,148]. The clean sequences are assembled with programs such as Velvet, SPAdes and Shovill [149,150]. The assembled genome is analyzed by software to evaluate its quality (e.g. number of contigs, total length of contigs) [151]. The functional gene annotation allows the obtainment of CDSs (coding DNA sequences) and their function. There are different bioinformatics tools used for the functional annotation and among them RAST and Prokka are the most used [152,153]. These software tools allow to obtain the amino acidic and nucleotides sequences of the CDSs as well as information related to their function [152]. The determination of the Clusters of Orthologous Genes classes (COGs) allows the detection and evaluation of particular gene families [154–158].

The use of software for the pangenome study is pivotal to obtain information about multiple microorganisms and their peculiar or shared characteristics. The detection of genes present in a certain number of isolates (genome partitions) and the detection of sequences linked to certain characteristics are often performed to evaluate bacterial groups [159–162].

The data obtained from functional annotations allow the gene expression evaluation through alignment with the whole transcriptome. The RNA sequencing (RNA-seq) is performed after the DNA elimination and production of cDNA [163]. The transcriptomic study allows the obtainment of information on the gene under different conditions with a level of accuracy higher than microarrays [163,164]. Moreover, the RNA-seq allows an evaluation of the entire transcriptome, unlike the real-time PCR which evaluates the expression of specific genes selected previously. Briefly, once the transcriptome sequences have been obtained, they are aligned on annotated sequences to obtain the number of transcripts corresponding to a given gene. The information obtained from the number of transcripts per gene are normalized and elaborated with statistical packages to evaluate the differentially expressed genes (DEGs), either overexpressed or under-expressed under various experimental conditions. This method can be used to study the gene expression of an organism in contact with another, allowing the transcriptome evaluation of a single or multiple organisms [165].

The different methods discussed allow to determine the bacterial gene content and their role in specific conditions.

1.3 *in vitro* cell models assay

The sequencing methods described above allow the detection of sequences present in the genomes. These data need to be linked to phenotypical information to understand and/or confirm their function in specific organism. A method frequently used to obtain information on the bacterial virulence mechanism *in vitro* is the use of cell models. Such information could be linked to genomic and transcriptomic features to understand the function of specific genes. In the following paragraphs some cell lines are described with the *in vitro* assay methods. Moreover, some examples will be reported about the use of *in vitro* cell models in the study of *Arcobacter* spp.

1.3.1 Human cell lines

The study of the bacterial virulence mechanism has been performed on animal models such as murine models. These models allow the evaluation of the virulence mechanisms and the evaluation of therapies and vaccines [166]. However, the *in vivo* studies require careful experimental design. For this reason the *in vitro* cell models represent a valid alternative for first host-pathogen interaction studies on a large strain number [167,168].

The types of *in vitro* cell models are numerous. These models can consist in a single cell line or several cell lines with different characteristics [168,169]. The monolayer cell models can be used in the study of bacterial adhesion and invasion and cytotoxic effects [170,171] (Figure 1.4).

A particular variant of *in vitro* cell model allows the observation of bacterial dynamics over time related to translocation ability (mimicking the crossing of the layer of intestinal enterocytes and the movement from the intestinal lumen to the lamina propria) and host membrane proteins

degradation. The translocation bacterial ability is evaluated by a bacterial count of the media collected in the model basal compartment while the host membrane proteins degradation is evaluated with the measurement of transepithelial electrical resistance (TEER) (Figure 1.4) [172].

After the first use of HeLa cell line in the production of *in vitro* cell models [173], different cell models have been used in the study of pathogens [168,174,175]. The cell models used in intestinal pathogen studies are often produced with cells of tumor origin, as they are more easily cultured. Some of these cell lines are HT-29 (colorectal cancer) [176], T84 (colorectal cancer) [177], TC-7, PD7, PF11 (sub-Caco-2 clones) [168,178,179]. The Caco-2 and HT29-MTX-E12 cell lines have been used in the production of mixed models characterized by some significant intestinal characteristics. The Caco-2 (human colon adenocarcinoma cells) cell line takes the function and morphology of mature enterocytes (epithelial cells of the intestinal villi) [178,180,181]. This cell line allows the obtainment of models with a dense layer, ideal for permeability tests. Caco-2 models have been used for the evaluation of host colonization by bacterial pathogens such as *C. jejuni*, *Campylobacter coli*, *S. enterica* and *E. coli* [175,182,183].

The HT29-MTX-E12 cell line (sub-homogeneous human intestinal mucus-producing cells) is derived from the parental line HT29. These cells were obtained from parental cells treated with methotrexate to induce mucus production exhibiting a phenotype like globose cells with a predominant expression of mucins MUC5AC [184,185]. The mucus represents an element produced by epithelial tissues which affects the behavior of pathogenic and non-pathogenic bacteria [186].

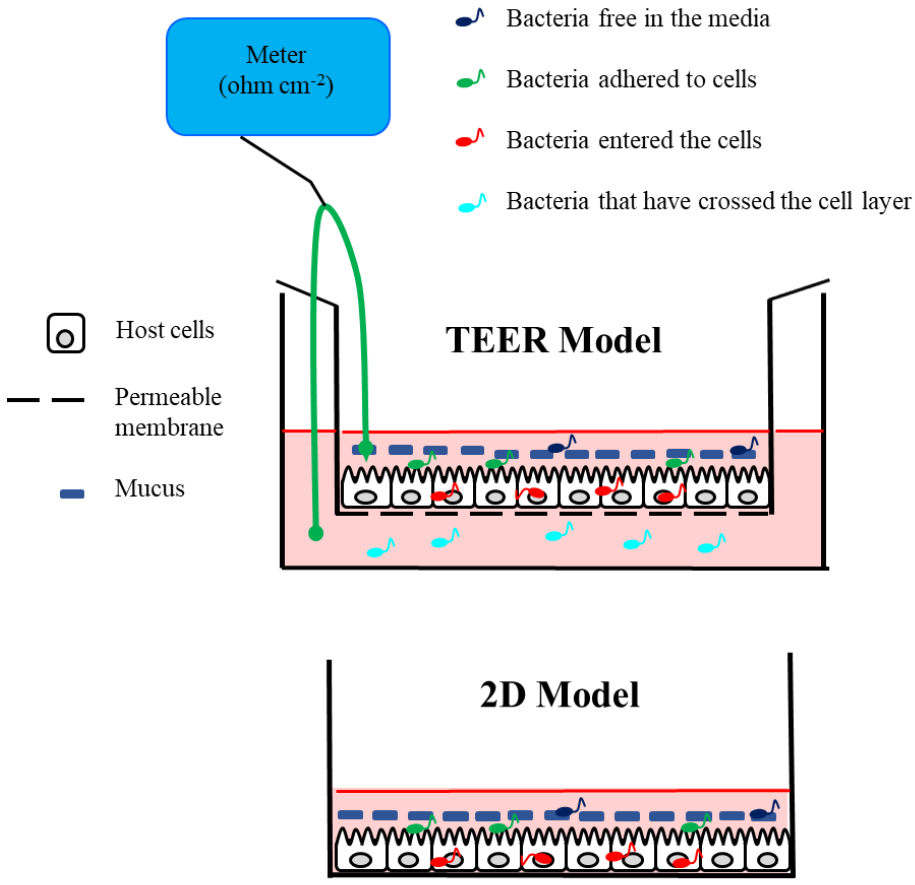


Figure 1.4. 2D and TEER cell models. The figure shows monolayer (2D model) and TEER cell models commonly used in studies of bacterial pathogens. The bacteria can adhere and invade cells until they pass the cell layer. The overcoming of the cellular layer can be detected in TEER models by measuring the transepithelial electrical resistance and by sampling bacteria recovered under the layer. The variations in the stability of the cellular layer can be detected, in the case of transepithelial electrical resistance. In case of damage by the pathogen a greater electrical passage will be observed. In the monolayer models, the bacteria adhered to the cells are detected after washing which allows the elimination of bacterial cells not adhered to host cells. Antibiotics are applied to detect internalized bacteria within host cells. The antibiotics eliminate bacteria unprotected by host cells and present externally. The detection of these bacteria is performed by microbiological load counts.

1.3.2 The role of mucus during bacterial colonization of cells

The mucus is a viscoelastic glycoproteic matrix that protects various surfaces, especially epithelial tissues with living cells on their surfaces [187]. The influence of mucus has been evaluated concerning probiotics and in relation to the pathogenic potential of some pathogens. An example is a decrease in the virulence of *C. jejuni* after its interaction with probiotics bacteria in host's mucus [188,189]. The importance of mucus is linked to its glycoproteic nature which creates a layer on the cell surface. The mucus layer can work as a barrier against pathogenic bacteria colonization. Part of these bacteria developed mechanisms to overcome the mucus barrier [187,190].

The number of goblet cells is directly related to the number of bacteria [191]. The percentage of goblet cells is approximately of 4% in the duodenum, 6% in the jejunum, 12% in the ileum, and 16% in the distal colon [191]. Considering the difficulty of determining the mucus layer thickness in humans there are few studies in literature. In rats, the mucus thickness is considered approximately of 170 µm in the duodenum, 124 µm in the jejunum and 480 µm in the ileum. The mucus thickness is of 500 µm in the duodenum, 250 µm in the jejunum and 200 µm in the ileum of mice. Furthermore, the production of mucins by goblet cells varies in the different gastrointestinal tracts. In humans, this difference is observed for mucin MUC2 that is produced in the different intestinal tracts while the mucin MUC6 and MUC5B are respectively produced in the duodenum and colon [191].

As previously reported, there are cell lines capable of mucus production allowing the evaluation of mucus role during *in vitro* tests (Figure 1.5) [184]. The HT29-MTX-E12 cell line has been used to determine the

influence of mucus on the virulence mechanism of pathogenic species such as *S. enterica, E. coli* and *C. jejuni* [184,192]. The characteristics of bacterial interaction with mucus varies between species and bacterial strains [186]. Studies related to TEER measurement have shown a greater invasiveness of *S. enterica* in HT29-MTX-E2 cells than in non-mucus-producer cell lines [184]. Vieira and colleagues observed that *E. coli* in contact with HT29 MTX cells increase the production of mucins MUC3, MUC4, MUC2 and MUC5AC [192]. In *C. jejuni* and *H. pylori* the mucus produced by HT29-MTX-E12 cells promoted higher levels of bacterial infection in comparison with non-mucus-producer parental cell lines [193]. The pathogens are not the only ones able to adhere to mucus. The commensal bacteria such as *Bifidobacterium bifidum* and *Bifidobacterium longum* subsp. *infantis* have demonstrated their ability to adhere to mucins respectively with the use of an extracellular transaldolase and of a family-1 solute binding proteins [194,195]. Different probiotic bacteria have been tested on HT29 cell line leading to an observation of an immunomodulatory activity on cells interleukin IL-8 production with consequent protection from an acute inflammatory response [179,196]. The bacteria use different mechanisms to colonize the mucus layer. The mucus adhesion of *E. coli* has been linked to fimbriae and flagella, while in *C. jejuni* this association has been attributed to the interaction of carbohydrate-lectin, flagellin A and major outer membrane protein with the mucin MucA. In *S. enterica* serotype Typhimurium the mucus adhesion has been linked to fimbrial adhesins [186]. Moreover, it was observed that the variation of the inflammatory response in the host can modulate the production of mucus [186]. The lipopolysaccharides (LPS) of *E. coli* led to a strong pro-inflammatory response in mammalian cells

with a consequent modulation of host mucus production. Different authors suggest that LPS, flagella and bacterial cell wall components led to a greater expression of genes related to the production of host mucus [186,191,197]. The different bacterial mechanisms linked to mucus interaction demonstrate the development of wide evolutionary strategies in the bacterial relationship with the host. This observation leads to speculate about the presence of genes related to mucus interaction.

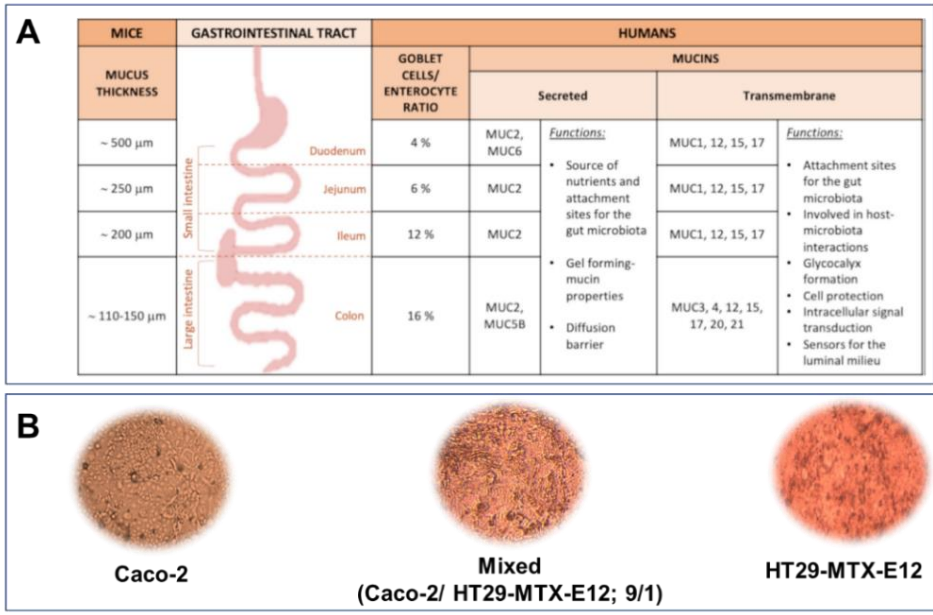


Figure 1.5. Mucins present in different gastrointestinal tracts (A) and *in vitro* intestinal human gut cell layers (B). The figure **A** shows information from Paone and colleagues of mucus present in different gastrointestinal tracts concerning the mucins secreted (in humans), mucus thickness (in mice) and mucus functions [191]. The figure **B** shows the observation under optical microscope of Caco-2, HT29-MTX-E12 cell layers and a mixed layer composed of these two cell lines.

1.3.3 *In vitro* gut models employed on *Arcobacteraceae*

The human intestinal cell lines have been used for the study of different bacterial species including part of *Arcobacteraceae* species [84]. The type

strain of *A. nitrofigilis* isolated from *S. alterniflora* showed the absence of adhesion and invasion on Caco-2 cell lines, unlike strains isolated from mussels [198]. *A. cryaerophilus* strains isolated from humans and different animals (bovine, poultry, sheep, porcine, pork and clams) showed invasive and colonizing abilities on different cell lines such as Caco-2, Hep-2 (epidermoid carcinoma of the larynx), IPI-2I (ileal cell lines), and HeLa (cervical carcinoma) [84,173,179]. Part of these strains isolated from cattle and poultry showed cytotoxicity on HeLa, INT 407 cell lines. Some *A. cryaerophilus* strains from pork, chicken and beef showed cytotoxicity against Vero cells [84,199]. Villarruel-López and colleagues demonstrated the cytotoxicity of 89.5% of *A. skirrowii* strains isolated from meat (pork, chicken, beef) against Vero cells [200]. The cytotoxic effect against Vero cells, was also observed for *A. butzleri* and *A. cryaerophilus* [200]. Levican and colleagues observed that *A. skirrowii* isolated from wastewater treatment plant sludge showed adhesion and invasion ability on Caco-2 cell lines. The same authors observed the absence of invasion for a *A. skirrowii* isolated from pork meat [198]. The Caco-2 cell line was used in the work of Levican and colleagues to study the adhesion and invasion ability of different *Arcobacter* species. Strains of *A. cibarius* (from poultry meat), *A. thereius* (isolated from pork, sewage, porcine abortions, and mussels), *A. trophiarum* (from pig and chicken feaces), *A. defluvi* (isolated from sewage), *A. cloacae* (from mussels and sewage) and *A. ellisi* (from mussels) showed adhesion and invasion ability towards Caco-2 cells. *A. molluscorum* and *A. mytili*, were not able to invade Caco-2 cells while *A. bivalviorum* was not able to adhere to Caco-2 cells [198].

A. butzleri is the species belonging to *Arcobacter* genus more frequently isolated from humans, animals and food. For this reason, this species has

been tested on different *in vitro* cell lines [84]. *A. butzleri* strains isolated from human, meat (from chicken, pork, beef), clams and mussel showed invasion and adhesion ability towards HT29 and Caco-2 cell lines [87,198,200,201]. *A. butzleri* demonstrated cytotoxicity in contact with Vero (kidney tissue), Hela and CHO cells, and adhesion ability towards Hep-2 and HeLa cells [3,84]. *A. butzleri* showed greater colonization towards HT29/B6 cell line, compared with results on HT29 cell lines suggesting a favorable mucus role during interaction with host cells [201,202]. In a study of Karadas and colleagues in 2016, *A. butzleri* strains were tested measuring the TEER on layers produced with HT-29/B6 and IPEC-J2 cells. The results showed a decrease in the cell layer integrity of 30-15% and 90-50% respectively for HT-29/B6 and IPEC-J2 layers [202]. The different works in literature show that part of the *Arcobacter* species showed ability to adhere, colonize and cause disfunctions when in contact with host cells. In particular, *A. butzleri* from different isolation source can adhere, invade and cause dysfunctions in different cell lines [3].

The principal symptom of *Arcobacter* spp. infections is diarrhea. Liévin-Le Moal and Servina summarized the structural and functional mechanisms by which enterovirulent bacteria cause diarrhea [203]. The intestinal epithelium consists of a single layer of highly polarized cells. The tight junction is a component of the apical junctional complex and seals the paracellular space between epithelial cells. Specific structural proteins compose the cytoskeleton and microtubule networks, which play pivotal roles in the polarized organization of intestinal cells. The brush border at the apical domain and basolateral cell domain contains proteins and transporters exerting specific intestinal functions. The intestinal epithelial barrier is essential in maintaining immune homeostasis. The

intestinal microbiota is present in the lumen, outside the mucus layer. Goblet cells secrete mucins that, combined with membrane-bound mucins, act as a physicochemical barrier and protecting the epithelial cell surface. The antimicrobial peptides are secreted by Paneth cells and enterocytes localized within the mucus layer, forming a first chemical defense system against enteric pathogens. The lamina propria, located under the membrane, contains immune and dendritic cells. The enterovirulent bacteria use their adhesive factors to interact with the brush border membrane by hijacking membrane-bound molecules as receptors. Structural and functional brush border damage results in adhesin/receptor interaction or T3SS-translocated bacterial effectors. These interactions activate cell signaling pathways that lead to the cytoskeleton-dependent attachment/effacement (A/E) lesion of brush border microvilli or shedding of microvilli. The lesions result in the disappearance of brush border-associated proteins exerting specific intestinal functions. The secreted cytotoxic toxins, by binding to membrane-bound receptors, by endocytosis and retrograde traffic, or by T3SS-translocated bacterial effectors, activate signaling pathways for deregulating membrane-associated proteins controlling nutrient transport or functioning as ions and water channels. Moreover, secreted cytotoxic toxins induce cytoskeleton- or caspase-dependent cell death after endocytosis. In addition, through T3SS-translocated bacterial effectors or secreted toxins, bacterial pathogens also target the junctional domain of polarized epithelial cells, inducing structural and functional lesions at the tight junctions and leading to a fault in the intestinal epithelial barrier. Invasive enterovirulent bacteria cross the epithelial cell membrane through a massive membrane rearrangement. These bacteria penetrate the host cells and pursue sophisticated

intracellular lifestyles within vacuoles containing bacteria. Other enteroinvasive bacteria, after escaping the vacuole, engage in actin-based movements within the cell cytoplasm. These movement is performed to penetrate neighboring cells through bacterium-induced transpodia. Adhering and invading enterovirulent bacteria trigger cellular defense responses, including, the enhanced production/secretion of mucus and the production of proinflammatory cytokines and chemokines activating, in turn, immune cells of lamina propria. The enterovirulent bacteria can cause a loss of the first line of intestinal defenses by modifying the resident microbiota composition or altering the secretory process of mucus from goblet cells [203]. For these reasons, simplified *in vitro* gut models can help to understand the bacterial mechanism of action and the possible adhesive/invasive capacity of the strains.

1.4 References

[1] L. Collado, M.J. Figueras, Taxonomy, Epidemiology, and Clinical Relevance of the Genus *Arcobacter*, Clin. Microbiol. Rev. 24 (2011) 174–192. https://doi.org/10.1128/CMR.00034-10.

[2] T.P. Ramees, K. Dhama, K. Karthik, R.S. Rathore, A. Kumar, M. Saminathan, R. Tiwari, Y.S. Malik, R.K. Singh, *Arcobacter* : an emerging food-borne zoonotic pathogen, its public health concerns and advances in diagnosis and control – a comprehensive review, Vet. Q. 37 (2017) 136–161. https://doi.org/10.1080/01652176.2017.1323355.

[3] D. Chieffi, F. Fanelli, V. Fusco, *Arcobacter butzleri* : Up-to-date taxonomy, ecology, and pathogenicity of an emerging pathogen, Compr. Rev. Food Sci. Food Saf. (2020) 1541-4337.12577. https://doi.org/10.1111/1541-4337.12577.

[4] C.R. McClung, D.G. Patriquin, R.E. Davis, *Campylobacter nitrofigilis* sp. nov., a Nitrogen-Fixing Bacterium Associated with Roots of *Spartina alterniflora* Loisel, Int. J. Syst. Bacteriol. 33 (1983) 605–612. https://doi.org/10.1099/00207713-33-3-605.

[5] J.A. Kiehlbauch, D.J. Brenner, M.A. Nicholson, C.N. Baker, C.M. Patton, A.G. Steigerwalt, I.K. Wachsmuth, *Campylobacter butzleri* sp. nov. isolated from humans and animals with diarrheal illness, J. Clin. Microbiol. 29 (1991) 376–385. https://doi.org/10.1128/jcm.29.2.376-385.1991.

[6] P. Vandamme, E. Falsen, R. Rossau, B. Hoste, P. Segers, R. Tytgat, J. De Ley, Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* Taxonomy: Emendation of Generic Descriptions and Proposal of *Arcobacter* gen. nov., Int. J. Syst. Bacteriol. 41 (1991) 88–103. https://doi.org/10.1099/00207713-41-1-88.

[7] P. Vandamme, M. Vancanneyt, B. Pot, L. Mels, B. Hoste, D. Dewettinck, L. Vlaes, C. van den Borre, R. Higgins, J. Hommez, Polyphasic taxonomic study of the emended genus *Arcobacter* with *Arcobacter butzleri* comb. nov. and *Arcobacter skirrowii* sp. nov., an aerotolerant bacterium isolated from veterinary specimens., Int. J. Syst. Bacteriol. 42 (1992) 344–56. https://doi.org/10.1099/00207713-42-3-344.

[8] D.W. Waite, I. Vanwonterghem, C. Rinke, D.H. Parks, Y. Zhang, K. Takai, S.M. Sievert, J. Simon, B.J. Campbell, T.E. Hanson, T. Woyke, M.G. Klotz, P. Hugenholtz, Comparative Genomic Analysis of the Class *Epsilonproteobacteria* and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.), Front. Microbiol. 8 (2017). https://doi.org/10.3389/fmicb.2017.00682.

[9] A. Pérez-Cataluña, N. Salas-Massó, A.L. Diéguez, S. Balboa, A. Lema,

J.L. Romalde, M.J. Figueras, Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos, Front. Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.02077.

[10] A. Pérez-Cataluña, L. Collado, O. Salgado, V. Lefiñanco, M.J. Figueras, A Polyphasic and Taxogenomic Evaluation Uncovers *Arcobacter cryaerophilus* as a Species Complex That Embraces Four Genomovars, Front. Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.00805.

[11] S.L.W. On, W.G. Miller, P.J. Biggs, A.J. Cornelius, P. Vandamme, *Aliarcobacter*, *Halarcobacter*, *Malaciobacter*, *Pseudarcobacter* and *Poseidonibacter* are later synonyms of *Arcobacter*: transfer of *Poseidonibacter parvus*, *Poseidonibacter antarcticus*, '*Halarcobacter arenosus'*, and '*Aliarcobacter vitoriensis'* to *Arcobacter* as *Arcobacter parvus* comb. nov., *Arcobacter antarcticus* comb. nov., *Arcobacter arenosus* comb. nov. and *Arcobacter vitoriensis* comb. nov, Int. J. Syst. Evol. Microbiol. 71 (2021). https://doi.org/10.1099/ijsem.0.005133.

[12] S.L.W. On, W.G. Miller, P.J. Biggs, A.J. Cornelius, P. Vandamme, A critical rebuttal of the proposed division of the genus *Arcobacter* into six genera using comparative genomic, phylogenetic, and phenotypic criteria, Syst. Appl. Microbiol. 43 (2020) 126108. https://doi.org/10.1016/j.syapm.2020.126108.

[13] E.M. D'sa, M.A. Harrison, Effect of pH, NaCl Content, and Temperature on Growth and Survival of *Arcobacter* spp., J. Food Prot. 68 (2005) 18–25. https://doi.org/10.4315/0362-028X-68.1.18.

[14] W.G. Miller, C.T. Parker, M. Rubenfield, G.L. Mendz, M.M.S.M. Wösten, D.W. Ussery, J.F. Stolz, T.T. Binnewies, P.F. Hallin, G. Wang, J.A. Malek, A. Rogosin, L.H. Stanker, R.E. Mandrell, The complete genome sequence and analysis of the epsilonproteobacterium *Arcobacter butzleri*, PLoS One. 2 (2007) e1358. https://doi.org/10.1371/journal.pone.0001358.

[15] A. Pati, S. Gronow, A. Lapidus, A. Copeland, T. Glavina Del Rio, M. Nolan, S. Lucas, H. Tice, J.-F. Cheng, C. Han, O. Chertkov, D. Bruce, R. Tapia, L. Goodwin, S. Pitluck, K. Liolios, N. Ivanova, K. Mavromatis, A. Chen, K. Palaniappan, M. Land, L. Hauser, Y.-J. Chang, C.D. Jeffries, J.C. Detter, M. Rohde, M. Göker, J. Bristow, J.A. Eisen, V. Markowitz, P. Hugenholtz, H.-P. Klenk, N.C. Kyrpides, Complete genome sequence of *Arcobacter nitrofigilis* type strain (CIT), Stand. Genomic Sci. 2 (2010) 300–308. https://doi.org/10.4056/sigs.912121.

[16] H. Ho, L. Lipman, W. Gaastra, *Arcobacter*, what is known and unknown about a potential foodborne zoonotic agent!, Vet. Microbiol. 115 (2006) 1–13. https://doi.org/10.1016/j.vetmic.2006.03.004.

[17] S.J. de Oliveira, A.L. Baetz, I.V. Wesley, K.M. Harmon, Classification of

Arcobacter species isolated from aborted pig fetuses and sows with reproductive problems in Brazil, Vet. Microbiol. 57 (1997) 347–354. https://doi.org/10.1016/S0378-1135(97)00106-5.

[18] E. Logan, S. Neill, D. Mackie, Mastitis in dairy cows associated with an aerotolerant *Campylobacter*, Vet. Rec. 110 (1982) 229–230. https://doi.org/10.1136/vr.110.10.229.

[19] P.-J. Kerkhof, M.F. Peruzy, N. Murru, K. Houf, Wild boars as reservoir for *Campylobacter* and *Arcobacter*, Vet. Microbiol. 270 (2022) 109462. https://doi.org/10.1016/j.vetmic.2022.109462.

[20] H.I. Atabay, J.E.L. Corry, The prevalence of *Campylobacters* and *Arcobacters* in broiler chickens, J. Appl. Microbiol. 83 (1997) 619–626. https://doi.org/10.1046/j.1365-2672.1997.00277.x.

[21] H. Atabay, M. Waino, M. Madsen, Detection and diversity of various *Arcobacter* species in Danish poultry, Int. J. Food Microbiol. 109 (2006) 139–145. https://doi.org/10.1016/j.ijfoodmicro.2006.01.020.

[22] E. Çelik, S. Otlu, Isolation of *Arcobacter* spp. and identification of isolates by multiplex PCR from various domestic poultry and wild avian species, Ann. Microbiol. 70 (2020) 60. https://doi.org/10.1186/s13213-020-01603-7.

[23] K. Houf, S.L.W. On, T. Coenye, L. Debruyne, S. De Smet, P. Vandamme, *Arcobacter thereius* sp. nov., isolated from pigs and ducks, Int. J. Syst. Evol. Microbiol. 59 (2009) 2599–2604. https://doi.org/10.1099/ijs.0.006650-0.

[24] R.G. Russell, J.A. Kiehlbauch, C.J. Gebhart, L.J. DeTolla, Uncommon *Campylobacter* species in infant *Macaca nemestrina* monkeys housed in a nursery, J. Clin. Microbiol. 30 (1992) 3024–3027.

[25] K.F. Anderson, J.A. Kiehlbauch, D.C. Anderson, H.M. McClure, I.K. Wachsmuth, *Arcobacter* (*Campylobacter*) *butzleri*-associated diarrheal illness in a nonhuman primate population, Infect. Immun. 61 (1993) 2220–2223. https://doi.org/10.1016/S0378-1127(01)00672-7.

[26] M.D. Goni, I.J. Muhammad, M. Goje, A.A. Bitrus, S.M. Jajere, B.M. Adam, M.A. Abbas, Occurrence of emerging *Arcobacter* in dogs and cats and its public health implications: A Review, Adv. Anim. Vet. Sci. 5 (2017) 362–370. https://doi.org/10.17582/journal.aavs/2017/5.9.362.370.

[27] M.T. Fera, E. La Camera, M. Carbone, D. Malara, M.G. Pennisi, Pet cats as carriers of *Arcobacter* spp. in Southern Italy, J. Appl. Microbiol. 106 (2009) 1661–1666. https://doi.org/10.1111/j.1365-2672.2008.04133.x.

[28] D.M. Goni, S. Abdulaziz, G.K. Dhaliwal, Z. Zakaria, I.J. Muhammad, M.A. Mohamed, A.A. Bello, A.A. Bitrus, Occurrence of *Arcobacter* in dogs and cats in Selangor, Malaysia, and associated risk factors,

TURKISH J. Vet. Anim. Sci. 40 (2016) 769–775. https://doi.org/10.3906/vet-1602-7.

[29] K. Houf, S. De Smet, J. Baré, S. Daminet, Dogs as carriers of the emerging pathogen *Arcobacter*, Vet. Microbiol. 130 (2008) 208–213. https://doi.org/10.1016/j.vetmic.2008.01.006.

[30] H. Yildiz, S. Aydin, Pathological effects of *Arcobacter cryaerophilus* infection in rainbow trout (*Oncorhynchus mykiss* Walbaum), Acta Vet. Hung. 54 (2006) 191–199. https://doi.org/10.1556/AVet.54.2006.2.6.

[31] L. Collado, J. Guarro, M.J. Figueras, Prevalence of *Arcobacter* in meat and shellfish, J. Food Prot. 72 (2009) 1102–1106. https://doi.org/10.4315/0362-028X-72.5.1102.

[32] A. Levican, L. Collado, C. Aguilar, C. Yustes, A.L. Diéguez, J.L. Romalde, M.J. Figueras, *Arcobacter bivalviorum* sp. nov. and *Arcobacter venerupis* sp. nov., new species isolated from shellfish, Syst. Appl. Microbiol. 35 (2012) 133–138. https://doi.org/10.1016/j.syapm.2012.01.002.

[33] M.J. Figueras, L. Collado, A. Levican, J. Perez, M.J. Solsona, C. Yustes, *Arcobacter molluscorum* sp. nov., a new species isolated from shellfish, Syst. Appl. Microbiol. 34 (2011) 105–109. https://doi.org/10.1016/j.syapm.2010.10.001.

[34] L. Collado, I. Cleenwerck, S. Van Trappen, P. De Vos, M.J. Figueras, *Arcobacter mytili* sp. nov., an indoxyl acetate-hydrolysis-negative bacterium isolated from mussels, Int. J. Syst. Evol. Microbiol. 59 (2009) 1391–1396. https://doi.org/10.1099/ijs.0.003749-0.

[35] A.H. Shah, A.A. Saleha, Z. Zunita, M. Murugaiyah, *Arcobacter* – An emerging threat to animals and animal origin food products?, Trends Food Sci. Technol. 22 (2011) 225–236. https://doi.org/10.1016/j.tifs.2011.01.010.

[36] H.T.K. Ho, L.J.A. Lipman, W. Gaastra, The introduction of *Arcobacter* spp. in poultry slaughterhouses, Int. J. Food Microbiol. 125 (2008) 223–229. https://doi.org/10.1016/j.ijfoodmicro.2008.02.012.

[37] Y. Badilla-Ramírez, K.L. Fallas-Padilla, H. Fernández-Jaramillo, M.L. Arias-Echandi, Survival capacity of *Arcobacter butzleri* inoculated in poultry meat at two different refrigeration temperatures, Rev. Inst. Med. Trop. Sao Paulo. 58 (2016) 22–24. https://doi.org/10.1590/S1678-9946201658022.

[38] R. Scullion, C.S. Harrington, R.H. Madden, Prevalence of *Arcobacter* spp. in Raw Milk and Retail Raw Meats in Northern Ireland, J. Food Prot. 69 (2006) 1986–1990. https://doi.org/10.4315/0362-028X-69.8.1986.

[39] A.H. Shah, A.A. Saleha, M. Murugaiyah, Z. Zunita, A.A. Memon,

Prevalence and Distribution of *Arcobacter* spp. in Raw Milk and Retail Raw Beef, J. Food Prot. 75 (2012) 1474–1478. https://doi.org/10.4315/0362-028X.JFP-11-487.

[40] F. Giacometti, A. Lucchi, G. Manfreda, D. Florio, R.G. Zanoni, A. Serraino, Occurrence and genetic diversity of *Arcobacter butzleri* in an artisanal dairy plant in Italy, Appl. Environ. Microbiol. 79 (2013) 6665–6669. https://doi.org/10.1128/AEM.02404-13.

[41] F. Giacometti, A. Serraino, F. Pasquali, A. De Cesare, E. Bonerba, R. Rosmini, Behavior of *Arcobacter butzleri* and *Arcobacter cryaerophilus* in Ultrahigh-Temperature, Pasteurized, and Raw Cow's Milk Under Different Temperature Conditions, Foodborne Pathog. Dis. 11 (2014) 15–20. https://doi.org/10.1089/fpd.2013.1597.

[42] A. Serraino, F. Giacometti, P. Daminelli, M.N. Losio, G. Finazzi, G. Marchetti, A. V. Zambrini, R. Rosmini, Survival of *Arcobacter butzleri* during production and storage of artisan water Buffalo Mozzarella Cheese, Foodborne Pathog. Dis. 10 (2013) 820–824. https://doi.org/10.1089/fpd.2013.1485.

[43] F. Giacometti, A. Lucchi, A. Di Francesco, M. Delogu, E. Grilli, I. Guarniero, L. Stancampiano, G. Manfreda, G. Merialdi, A. Serraino, *Arcobacter butzleri*, *Arcobacter cryaerophilus*, and *Arcobacter skirrowii* circulation in a dairy farm and sources of milk contamination, Appl. Environ. Microbiol. 81 (2015) 5055–5063. https://doi.org/10.1128/AEM.01035-15.

[44] A. González, M.A. Ferrús, Study of *Arcobacter* spp. contamination in fresh lettuces detected by different cultural and molecular methods, Int. J. Food Microbiol. 145 (2011) 311–314. https://doi.org/10.1016/j.ijfoodmicro.2010.11.018.

[45] A. Mottola, G. Ciccarese, C. Sinisi, A.E. Savarino, P. Marchetti, V. Terio, G. Tantillo, R. Barrasso, A. Di Pinto, Occurrence and characterization of *Arcobacter* spp. from ready-to-eat vegetables produced in Southern Italy, Ital. J. Food Saf. 10 (2021). https://doi.org/10.4081/ijfs.2021.8585.

[46] N.H. Kim, S.M. Park, H.W. Kim, T.J. Cho, S.H. Kim, C. Choi, M.S. Rhee, Prevalence of pathogenic *Arcobacter* species in South Korea: Comparison of two protocols for isolating the bacteria from foods and examination of nine putative virulence genes, Food Microbiol. 78 (2019) 18–24. https://doi.org/10.1016/j.fm.2018.09.008.

[47] D. Uljanovas, G. Gölz, V. Brückner, A. Grineviciene, E. Tamuleviciene, T. Alter, M. Malakauskas, Prevalence, antimicrobial susceptibility and virulence gene profiles of *Arcobacter* species isolated from human stool samples, foods of animal origin, ready-to-eat salad mixes and environmental water, Gut Pathog. 13 (2021) 76.

https://doi.org/10.1186/s13099-021-00472-y.

[48] M.H. Lee, C. Choi, Survival of *Arcobacter butzleri* in apple and pear purees, J. Food Saf. 33 (2013) 333–339. https://doi.org/10.1111/jfs.12057.

[49] F. Fanelli, A. Di Pinto, A. Mottola, G. Mule, D. Chieffi, F. Baruzzi, G. Tantillo, V. Fusco, Genomic Characterization of *Arcobacter butzleri* Isolated From Shellfish: Novel Insight Into Antibiotic Resistance and Virulence Determinants, Front. Microbiol. 10 (2019) 1–17. https://doi.org/10.3389/fmicb.2019.00670.

[50] V. Lappi, J.R. Archer, E. Cebelinski, F. Leano, J.M. Besser, R.F. Klos, C. Medus, K.E. Smith, C. Fitzgerald, J.P. Davis, An Outbreak of Foodborne Illness Among Attendees of a Wedding Reception in Wisconsin Likely caused by *Arcobacter butzleri*, Foodborne Pathog. Dis. 10 (2013) 250–255. https://doi.org/10.1089/fpd.2012.1307.

[51] A. Levican, L. Collado, M.J. Figueras, The Use of Two Culturing Methods in Parallel Reveals a High Prevalence and Diversity of *Arcobacter* spp. in a Wastewater Treatment Plant, Biomed Res. Int. 2016 (2016) 1–9. https://doi.org/10.1155/2016/8132058.

[52] S.P. Donachie, J.P. Bowman, S.L.W. On, M. Alam, *Arcobacter halophilus* sp. nov., the first obligate halophile in the genus *Arcobacter*, Int. J. Syst. Evol. Microbiol. 55 (2005) 1271–1277. https://doi.org/10.1099/ijs.0.63581-0.

[53] H.M. Kim, C.Y. Hwang, B.C. Cho, *Arcobacter marinus* sp. nov, Int. J. Syst. Evol. Microbiol. 60 (2010) 531–536. https://doi.org/10.1099/ijs.0.007740-0.

[54] Z. Zhang, C. Yu, X. Wang, S. Yu, X.-H. Zhang, *Arcobacter pacificus* sp. nov., isolated from seawater of the South Pacific Gyre, Int. J. Syst. Evol. Microbiol. 66 (2016) 542–547. https://doi.org/10.1099/ijsem.0.000751.

[55] M.T. Fera, T.L. Maugeri, C. Gugliandolo, C. Beninati, M. Giannone, E. La Camera, M. Carbone, Detection of *Arcobacter* spp. in the coastal environment of the Mediterranean sea, Appl. Environ. Microbiol. 70 (2004) 1271–1276. https://doi.org/10.1128/AEM.70.3.1271-1276.2004.

[56] M.A. Assanta, D. Roy, M.-J. Lemay, D. Montpetit, Attachment of *Arcobacter butzleri*, a new waterborne pathogen, to water distribution pipe surfaces, J. Food Prot. 65 (2002) 1240–1247. https://doi.org/10.4315/0362-028X-65.8.1240.

[57] C.I. Collins, E.A. Murano, I. V. Wesley, Survival of *Arcobacter butzleri* and *Campylobacter* jejuni after irradiation treatment in vacuum-packaged ground pork, J. Food Prot. 59 (1996) 1164–1166. https://doi.org/10.4315/0362-028X-59.11.1164.

[58] C.A. Phillips, The effect of citric acid, lactic acid, sodium citrate and sodium lactate, alone and in combination with nisin, on the growth of *Arcobacter butzleri*, Lett. Appl. Microbiol. 29 (1999) 424–428. https://doi.org/10.1046/j.1472-765X.1999.00668.x.

[59] M.J. Figueras, A. Levican, I. Pujol, F. Ballester, M.J.R. Quilez, F. Gomez-Bertomeu, A severe case of persistent diarrhoea associated with *Arcobacter cryaerophilus* but attributed to *Campylobacter* sp. and a review of the clinical incidence of *Arcobacter* spp., New Microbes New Infect. 2 (2014) 31–37. https://doi.org/10.1002/2052-2975.35.

[60] P.-J. Kerkhof, A.-M. Van den Abeele, B. Strubbe, D. Vogelaers, P. Vandamme, K. Houf, Diagnostic approach for detection and identification of emerging enteric pathogens revisited: the (*Ali*)*arcobacter lanthieri* case, New Microbes New Infect. 39 (2021) 100829. https://doi.org/10.1016/j.nmni.2020.100829.

[61] A. Samie, C.L. Obi, L.J. Barrett, S.M. Powell, R.L. Guerrant, Prevalence of *Campylobacter* species, *Helicobacter pylori* and *Arcobacter* species in stool samples from the Venda region, Limpopo, South Africa: Studies using molecular diagnostic methods, J. Infect. 54 (2007) 558–566. https://doi.org/10.1016/j.jinf.2006.10.047.

[62] O. Vandenberg, A. Dediste, K. Houf, S. Ibekwem, H. Souayah, S. Cadranel, N. Douat, G. Zissis, J.-P. Butzler, P. Vandamme, *Arcobacter* species in humans, Emerg. Infect. Dis. 10 (2004) 1863–1867. https://doi.org/10.3201/eid1010.040241.

[63] V. Prouzet-Mauléon, L. Labadi, N. Bouges, A. Ménard, F. Mégraud, *Arcobacter butzleri* : underestimated enteropathogen, Emerg. Infect. Dis. 12 (2006) 307–309. https://doi.org/10.3201/eid1202.050570.

[64] Z.-D. Jiang, H.L. DuPont, E.L. Brown, R.K. Nandy, T. Ramamurthy, A. Sinha, S. Ghosh, S. Guin, K. Gurleen, S. Rodrigues, J.J. Chen, R. McKenzie, R. Steffen, Microbial etiology of travelers' diarrhea in Mexico, Guatemala, and India: Importance of Enterotoxigenic *Bacteroides fragilis* and *Arcobacter* Species, J. Clin. Microbiol. 48 (2010) 1417–1419. https://doi.org/10.1128/JCM.01709-09.

[65] P. Vandamme, P. Pugina, G. Benzi, R. Van Etterijck, L. Vlaes, K. Kersters, J.P. Butzler, H. Lior, S. Lauwers, Outbreak of recurrent abdominal cramps associated with *Arcobacter butzleri* in an Italian school, J. Clin. Microbiol. 30 (1992) 2335–2337. https://doi.org/10.1128/jcm.30.9.2335-2337.1992.

[66] K.K. Soelberg, T.K.L. Danielsen, R. Martin-Iguacel, U.S. Justesen, *Arcobacter butzleri* is an opportunistic pathogen: recurrent bacteraemia in an immunocompromised patient without diarrhoea, Access Microbiol. 2 (2020) 0–3. https://doi.org/10.1099/acmi.0.000145.

[67] W. Tee, R. Baird, M. Dyall-Smith, B. Dwyer, *Campylobacter cryaerophila* isolated from a human., J. Clin. Microbiol. 26 (1988) 2469–73. https://doi.org/10.1128/jcm.26.12.2469-2473.1988.

[68] J. Lerner, V. Brumberger, V. Preac-Mursic, Severe diarrhea associated with *Arcobacter butzleri*, Eur. J. Clin. Microbiol. Infect. Dis. 13 (1994) 660–662. https://doi.org/10.1007/BF01973994.

[69] H. Fernández, S. Krause, M. Paz Villanueva, *Arcobacter butzleri* an emerging enteropathogen: communication of two cases with chronic diarrhea, Brazilian J. Microbiol. 35 (2004) 216–218. https://doi.org/10.1590/S1517-83822004000200008.

[70] I. Wybo, J. Breynaert, S. Lauwers, F. Lindenburg, K. Houf, Isolation of *Arcobacter skirrowii* from a Patient with Chronic Diarrhea, J. Clin. Microbiol. 42 (2004) 1851–1852. https://doi.org/10.1128/JCM.42.4.1851-1852.2004.

[71] T. Kayman, S. Abay, H. Hizlisoy, H.İ. Atabay, K.S. Diker, F. Aydin, Emerging pathogen Arcobacter spp. in acute gastroenteritis: molecular identification, antibiotic susceptibilities and genotyping of the isolated *Arcobacters*, J. Med. Microbiol. 61 (2012) 1439–1444. https://doi.org/10.1099/jmm.0.044594-0.

[72] S.L.W. On, A. Stacey, J. Smyth, Isolation of *Arcobacter butzleri* from a neonate with bacteraemia, J. Infect. 31 (1995) 225–227. https://doi.org/10.1016/S0163-4453(95)80031-X.

[73] P.R. Hsueh, L.J. Teng, P.C. Yang, S.K. Wang, S.C. Chang, S.W. Ho, W.C. Hsieh, K.T. Luh, Bacteremia caused by *Arcobacter cryaerophilus* 1B, J. Clin. Microbiol. 35 (1997) 489–491. https://doi.org/10.1128/jcm.35.2.489-491.1997.

[74] E. Arguello, C.C. Otto, P. Mead, N.E. Babady, Bacteremia Caused by *Arcobacter butzleri* in an Immunocompromised Host, J. Clin. Microbiol. 53 (2015) 1448–1451. https://doi.org/10.1128/JCM.03450-14.

[75] S.K.P. Lau, Identification by 16S ribosomal RNA gene sequencing of *Arcobacter butzleri* bacteraemia in a patient with acute gangrenous appendicitis, Mol. Pathol. 55 (2002) 182–185. https://doi.org/10.1136/mp.55.3.182.

[76] T. Monzón, F. Coronel, A Patient with Type 1 Diabetes Continuing on Peritoneal Dialysis after more than 15 Years, Perit. Dial. Int. J. Int. Soc. Perit. Dial. 33 (2013) 220–222. https://doi.org/10.3747/pdi.2011.00331.

[77] 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.

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

[78] A.-M. Van den Abeele, D. Vogelaers, E. Vanlaere, K. Houf, Antimicrobial susceptibility testing of *Arcobacter butzleri* and *Arcobacter cryaerophilus* strains isolated from Belgian patients, J. Antimicrob. Chemother. 71 (2016) 1241–1244. https://doi.org/10.1093/jac/dkv483.

[79] P. Kietsiri, C. Muangnapoh, W. Lurchachaiwong, P. Lertsethtakarn, L. Bodhidatta, O. Suthienkul, N.C. Waters, S.T. Demons, B.A. Vesely, Characterization of *Arcobacter* spp. isolated from human diarrheal, non-diarrheal and food samples in Thailand, PLoS One. 16 (2021) e0246598. https://doi.org/10.1371/journal.pone.0246598.

[80] H. Shirzad Aski, M. Tabatabaei, R. Khoshbakht, M. Raeisi, Occurrence and antimicrobial resistance of emergent *Arcobacter* spp. isolated from cattle and sheep in Iran, Comp. Immunol. Microbiol. Infect. Dis. 44 (2016) 37–40. https://doi.org/10.1016/j.cimid.2015.12.002.

[81] S. Abay, T. Kayman, H. Hizlisoy, F. Aydin, *In vitro* Antibacterial Susceptibility of *Arcobacter butzleri* Isolated from Different Sources, J. Vet. Med. Sci. 74 (2012) 613–616. https://doi.org/10.1292/jvms.11-0487.

[82] V. Brückner, U. Fiebiger, R. Ignatius, J. Friesen, M. Eisenblätter, M. Höck, T. Alter, S. Bereswill, G. Gölz, M.M. Heimesaat, Prevalence and antimicrobial susceptibility of *Arcobacter* species in human stool samples derived from out- and inpatients: the prospective German *Arcobacter* prevalence study Arcopath, Gut Pathog. 12 (2020) 21. https://doi.org/10.1186/s13099-020-00360-x.

[83] D. Dekker, D. Eibach, K.G. Boahen, C.W. Akenten, Y. Pfeifer, A.E. Zautner, E. Mertens, R. Krumkamp, A. Jaeger, A. Flieger, E. Owusu-Dabo, J. May, Fluoroquinolone-resistant *Salmonella enterica*, *Campylobacter* spp., and *Arcobacter butzleri* from local and imported poultry meat in Kumasi, Ghana, Foodborne Pathog. Dis. 16 (2019) 352–358. https://doi.org/10.1089/fpd.2018.2562.

[84] S. Ferreira, J.A. Queiroz, M. Oleastro, F.C. Domingues, Insights in the pathogenesis and resistance of *Arcobacter*: a review, Crit. Rev. Microbiol. 42 (2015) 1–20. https://doi.org/10.3109/1040841X.2014.954523.

[85] 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.

[86] R.S.W. Tsang, J.M.C. Luk, D.L. Woodward, W.M. Johnson, Immunochemical characterization of a haemagglutinating antigen of *Arcobacter* spp., FEMS Microbiol. Lett. 136 (1996). https://doi.org/0378-1097/96.

[87] H.T.K. Ho, L.J.A. Lipman, H.G.C.J.M. Hendriks, P.C.J. Tooten, T. Ultee, W. Gaastra, Interaction of *Arcobacter* spp. with human and porcine intestinal epithelial cells, FEMS Immunol. Med. Microbiol. 50 (2007) 51–58. https://doi.org/10.1111/j.1574-695X.2007.00230.x.

[88] J. zur Bruegge, C. Hanisch, R. Einspanier, T. Alter, G. Gölz, S. Sharbati, *Arcobacter butzleri* induces a pro-inflammatory response in THP-1 derived macrophages and has limited ability for intracellular survival, Int. J. Med. Microbiol. 304 (2014) 1209–1217. https://doi.org/10.1016/j.ijmm.2014.08.017.

[89] G. Gölz, G. Karadas, M.E. Alutis, A. Fischer, A.A. Kühl, A. Breithaupt, U.B. Göbel, T. Alter, S. Bereswill, M.M. Heimesaat, *Arcobacter butzleri* induce colonic, extra-intestinal and systemic inflammatory responses in gnotobiotic IL-10 deficient mice in a strain-dependent manner, PLoS One. 10 (2015) e0139402. https://doi.org/10.1371/journal.pone.0139402.

[90] H.T.K. Ho, L.J.A. Lipman, M.M.S.M. Wãsten, A.J.A.M. van Asten, W. Gaastra, *Arcobacter* spp. possess two very short flagellins of which FlaA is essential for motility, FEMS Immunol. Med. Microbiol. 53 (2008) 85–95. https://doi.org/10.1111/j.1574-695X.2008.00405.x.

[91] L. Douidah, L. de Zutter, J. Bare, P. De Vos, P. Vandamme, O. Vandenberg, A.-M. Van den Abeele, K. Houf, Occurrence of Putative virulence genes in *Arcobacter* species isolated from humans and animals, J. Clin. Microbiol. 50 (2012) 735–741. https://doi.org/10.1128/JCM.05872-11.

[92] 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 Saf. 8 (2019). https://doi.org/10.4081/ijfs.2019.7840.

[93] F. Rovetto, A. Carlier, A.M. Van Den Abeele, K. Illeghems, F. Van Nieuwerburgh, L. Cocolin, K. Houf, Characterization of the emerging zoonotic pathogen *Arcobacter thereius* by whole genome sequencing and comparative genomics, PLoS One. 12 (2017) 1–27. https://doi.org/10.1371/journal. pone.0180493.

[94] L.C.M. Antunes, R.B.R. Ferreira, Biofilms and bacterial virulence, Rev. Med. Microbiol. 22 (2011) 12–16. https://doi.org/10.1097/MRM.0b013e3283410d22.

[95] I. Martinez-Malaxetxebarria, C. Girbau, A. Salazar-Sánchez, I. Baztarrika, I. Martínez-Ballesteros, L. Laorden, R. Alonso, A. Fernández-Astorga, Genetic characterization and biofilm formation of potentially pathogenic foodborne *Arcobacter* isolates, Int. J. Food Microbiol. 373 (2022) 109712. https://doi.org/10.1016/j.ijfoodmicro.2022.109712.

[96] J.Y. Merga, A.J.H. Leatherbarrow, C. Winstanley, M. Bennett, C.A. Hart, W.G. Miller, N.J. Williams, Comparison of *Arcobacter* isolation methods, and diversity of *Arcobacter* spp. in Cheshire, United Kingdom, Appl. Environ. Microbiol. 77 (2011) 1646–1650. https://doi.org/10.1128/AEM.01964-10.

[97] A.M. Lammerding, J.E. Harris, H. Lior, D.E. Woodward, L. Cole, C.A. Muckle, Isolation method for recovery of *Arcobacter butzleri* from fresh poultry and poultry products, in: *Campylobacters, Helicobacters*, Relat. Org., Springer US, Boston, MA, 1996: pp. 329–333. https://doi.org/10.1007/978-1-4757-9558-5_58.

[98] C.I. Collins, I. V. Wesley, E.A. Murano, Detection of *Arcobacter* spp. in ground pork by modified plating methods, J. Food Prot. 59 (1996) 448–452. https://doi.org/10.4315/0362-028X-59.5.448.

[99] K. Houf, L.A. Devriese, L. De Zutter, J. Van Hoof, P. Vandamme, Susceptibility of *Arcobacter butzleri*, *Arcobacter cryaerophilus*, and *Arcobacter skirrowii* to Antimicrobial Agents Used in Selective Media, J. Clin. Microbiol. 39 (2001) 1654–1656. https://doi.org/10.1128/JCM.39.4.1654-1656.2001.

[100] E. Boer, J.J.H.C. Tilburg, D.L. Woodward, H. Lior, W.M. Johnson, A selective medium for the isolation of *Arcobacter* from meats, Lett. Appl. Microbiol. 23 (1996) 64–66. https://doi.org/10.1111/j.1472-765X.1996.tb00030.x.

[101] L.G. Johnson, E.A. Murano, Development of a new medium for the isolation of *Arcobacter* spp., J. Food Prot. 62 (1999) 456–462. https://doi.org/10.4315/0362-028X-62.5.456.

[102] N. Salas-Massó, K.B. Andree, M.D. Furones, M.J. Figueras, Enhanced recovery of *Arcobacter* spp. using NaCl in culture media and re-assessment of the traits of *Arcobacter marinus* and *Arcobacter halophilus* isolated from marine water and shellfish, Sci. Total Environ. 566–567 (2016) 1355–1361. https://doi.org/10.1016/j.scitotenv.2016.05.197.

[103] F.U. Rahman, K.B. Andree, N. Salas-Massó, M. Fernandez-Tejedor, A. Sanjuan, M.J. Figueras, M.D. Furones, Improved culture enrichment broth for isolation of *Arcobacter*-like species from the marine environment, Sci. Rep. 10 (2020) 14547. https://doi.org/10.1038/s41598-020-71442-8.

[104] H.I. Atabay, J.E.L. Corry, Evaluation of a new arcobacter enrichment medium and comparison with two media developed for enrichment of *Campylobacter* spp., Int. J. Food Microbiol. 41 (1998) 53–58. https://doi.org/10.1016/S0168-1605(98)00034-8.

[105] H.I. Atabay, F. Aydin, K. Houf, M. Sahin, P. Vandamme, The prevalence of *Arcobacter* spp. on chicken carcasses sold in retail markets in Turkey, and identification of the isolates using SDS-PAGE, Int. J. Food Microbiol.

81 (2003) 21–28. https://doi.org/10.1016/S0168-1605(02)00163-0.

[106] J. Engberg, S.L.W. On, C.S. Harrington, P. Gerner-Smidt, Prevalence of *Campylobacter*, *Arcobacter*, *Helicobacter*, and *Sutterella* spp. in human fecal samples as estimated by a reevaluation of isolation methods for *Campylobacters*, J. Clin. Microbiol. 38 (2000) 286–291. https://doi.org/10.1128/jcm.38.1.286-291.2000.

[107] N. Singhal, M. Kumar, P.K. Kanaujia, J.S. Virdi, MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis, Front. Microbiol. 6 (2015) 1–16. https://doi.org/10.3389/fmicb.2015.00791.

[108] K. Houf, A. Tutenel, L. Zutter, J. Hoof, P. Vandamme, Development of a multiplex PCR assay for the simultaneous detection and identification of *Arcobacter butzleri*, *Arcobacter cryaerophilus* and *Arcobacter skirrowii*, FEMS Microbiol. Lett. 193 (2000) 89–94. https://doi.org/10.1111/j.1574-6968.2000.tb09407.x.

[109] D. PENTIMALLI, N. PEGELS, T. GARCÍA, R. MARTÍN, I. GONZÁLEZ, Specific PCR Detection of *Arcobacter butzleri*, *Arcobacter cryaerophilus*, *Arcobacter skirrowii*, and A*rcobacter cibarius* in Chicken meat, J. Food Prot. 72 (2009) 1491–1495. https://doi.org/10.4315/0362-028X-72.7.1491.

[110] A. González, J. Suski, M.A. Ferrús, Rapid and accurate detection of *Arcobacter* contamination in commercial chicken products and wastewater samples by real-time polymerase chain reaction, Foodborne Pathog. Dis. 7 (2010) 327–338. https://doi.org/10.1089/fpd.2009.0368.

[111] G. Brightwell, E. Mowat, R. Clemens, J. Boerema, D.J. Pulford, S.L. On, Development of a multiplex and real time PCR assay for the specific detection of *Arcobacter butzleri* and *Arcobacter cryaerophilus*, J. Microbiol. Methods. 68 (2007) 318–325. https://doi.org/10.1016/j.mimet.2006.09.008.

[112] X. Wang, D.J. Seo, M.H. Lee, C. Choi, A.B. Onderdonk, Comparison of Conventional PCR, Multiplex PCR, and Loop-Mediated Isothermal Amplification Assays for Rapid Detection of *Arcobacter* Species, J. Clin. Microbiol. 52 (2014) 557–563. https://doi.org/10.1128/JCM.02883-13.

[113] A. Levican, M. Figueras, Performance of five molecular methods for monitoring *Arcobacter* spp, BMC Microbiol. 13 (2013) 220. https://doi.org/10.1186/1471-2180-13-220.

[114] K. Houf, L. De Zutter, J. Van Hoof, P. Vandamme, Assessment of the genetic diversity among *Arcobacters* isolated from poultry products by using two PCR-based typing methods, Appl. Environ. Microbiol. 68 (2002) 2172–2178. https://doi.org/10.1128/AEM.68.5.2172-2178.2002.

[115] J.Y. Merga, N.J. Williams, W.G. Miller, A.J.H. Leatherbarrow, M. Bennett, N. Hall, K.E. Ashelford, C. Winstanley, Exploring the diversity of *Arcobacter butzleri* from cattle in the UK using MLST and whole Genome Sequencing, PLoS One. 8 (2013) e55240. https://doi.org/10.1371/journal.pone.0055240.

[116] L. Douidah, L. De Zutter, J. Baré, K. Houf, Towards a typing strategy for *Arcobacter* species isolated from humans and animals and assessment of the in vitro genomic stability, Foodborne Pathog. Dis. 11 (2014) 272–280. https://doi.org/10.1089/fpd.2013.1661.

[117] A. González, M.A. Ferrús, R. González, J. Hernández, Molecular fingerprinting of *Campylobacter* and *Arcobacter* isolated from chicken and water., Int. Microbiol. 10 (2007) 85–90. https://doi.org/10.2436/20.1501.01.12.

[118] W.G. Miller, I. V. Wesley, S.L.W. On, K. Houf, F. Mégraud, G. Wang, E. Yee, A. Srijan, C.J. Mason, First multi-locus sequence typing scheme for *Arcobacter* spp., BMC Microbiol. 9 (2009) 196. https://doi.org/10.1186/1471-2180-9-196.

[119] A. De Cesare, A. Parisi, F. Giacometti, A. Serraino, S. Piva, M. Caruso, E.P.L. De Santis, G. Manfreda, Multilocus sequence typing of *Arcobacter butzleri* isolates collected from dairy plants and their products, and comparison with their PFGE types, J. Appl. Microbiol. 120 (2016) 165–174. https://doi.org/10.1111/jam.12977.

[120] P.-J. Kerkhof, S.L.W. On, K. Houf, *Arcobacter vandammei* sp. nov., isolated from the rectal mucus of a healthy pig, Int. J. Syst. Evol. Microbiol. 71 (2021). https://doi.org/10.1099/ijsem.0.005113.

[121] J.D. Watson, F.H.C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid, Nature. 171 (1953) 737–738. https://doi.org/10.1038/171737a0.

[122] D.T. Zallen, Despite Franklin's work, Wilkins earned his Nobel, Nature. 425 (2003) 15–15. https://doi.org/10.1038/425015b.

[123] F. Sanger, G.G. Brownlee, B.G. Barrell, A two-dimensional fractionation procedure for radioactive nucleotides, J. Mol. Biol. 14 (1965) 303. https://doi.org/10.1016/S0022-2836(65)80253-4.

[124] W.M. Jou, G. Haegeman, M. Ysebaert, W. Fiers, Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein, Nature. 237 (1972) 82–88. https://doi.org/10.1038/237082a0.

[125] W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, M. Ysebaert, Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the

replicase gene, Nature. 260 (1976) 500–507. https://doi.org/10.1038/260500a0.

[126] J.M. Heather, B. Chain, The sequence of sequencers: The history of sequencing DNA, Genomics. 107 (2016) 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003.

[127] F. Sanger, A.R. Coulson, A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, J. Mol. Biol. 94 (1975) 441–448. https://doi.org/10.1016/0022-2836(75)90213-2.

[128] A.M. Maxam, W. Gilbert, A new method for sequencing DNA., Proc. Natl. Acad. Sci. 74 (1977) 560–564. https://doi.org/10.1073/pnas.74.2.560.

[129] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. 74 (1977) 5463–5467. https://doi.org/10.1073/pnas.74.12.5463.

[130] P. Nyrén, A. Lundin, Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis, Anal. Biochem. 151 (1985) 504–509. https://doi.org/10.1016/0003-2697(85)90211-8.

[131] M.A.J. Roberts, Recombinant DNA technology and DNA sequencing, Essays Biochem. 63 (2019) 457–468. https://doi.org/10.1042/EBC20180039.

[132] E.D. Hyman, A new method of sequencing DNA, Anal. Biochem. 174 (1988) 423–436. https://doi.org/10.1016/0003-2697(88)90041-3.

[133] M. Ronaghi, M. Uhlén, P. Nyrén, A Sequencing Method Based on Real-Time Pyrophosphate, Science (80-.). 281 (1998) 363–365. https://doi.org/10.1126/science.281.5375.363.

[134] G. Turcatti, A. Romieu, M. Fedurco, A.-P. Tairi, A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis †, Nucleic Acids Res. 36 (2008) e25–e25. https://doi.org/10.1093/nar/gkn021.

[135] M. Quail, M.E. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow, Y. Gu, A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers, BMC Genomics. 13 (2012) 341. https://doi.org/10.1186/1471-2164-13-341.

[136] W.J. Greenleaf, A. Sidow, The future of sequencing: convergence of intelligent design and market Darwinism, Genome Biol. 15 (2014) 303. https://doi.org/10.1186/gb4168.

[137] V. Bansal, C. Boucher, Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going?, IScience. 18 (2019) 37–41. https://doi.org/10.1016/j.isci.2019.06.035.

[138] D.J. Hedges, T. Guettouche, S. Yang, G. Bademci, A. Diaz, A. Andersen, W.F. Hulme, S. Linker, A. Mehta, Y.J.K. Edwards, G.W. Beecham, E.R. Martin, M.A. Pericak-Vance, S. Zuchner, J.M. Vance, J.R. Gilbert, Comparison of Three Targeted Enrichment Strategies on the SOLiD Sequencing Platform, PLoS One. 6 (2011) e18595. https://doi.org/10.1371/journal.pone.0018595.

[139] J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J.H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J.R. Nobile, B.P. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A. Williams, G.T. Roth, J. Bustillo, An integrated semiconductor device enabling non-optical genome sequencing, Nature. 475 (2011) 348–352. https://doi.org/10.1038/nature10242.

[140] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. DeWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Real-Time DNA Sequencing from Single Polymerase Molecules, Science (80-.). 323 (2009) 133–138. https://doi.org/10.1126/science.1162986.

[141] A.M. Giani, G.R. Gallo, L. Gianfranceschi, G. Formenti, Long walk to genomics: History and current approaches to genome sequencing and assembly, Comput. Struct. Biotechnol. J. 18 (2020) 9–19. https://doi.org/10.1016/j.csbj.2019.11.002.

[142] M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations, Science (80-.). 299 (2003) 682–686. https://doi.org/10.1126/science.1079700.

[143] A. Payne, N. Holmes, V. Rakyan, M. Loose, BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files, Bioinformatics. 35 (2019) 2193–2198. https://doi.org/10.1093/bioinformatics/bty841.

[144] N.J. Loman, M.J. Pallen, Twenty years of bacterial genome sequencing, Nat. Rev. Microbiol. 13 (2015) 787–794. https://doi.org/10.1038/nrmicro3565.

[145] S.S. Gautam, R. KC, K.W. Leong, M. Mac Aogáin, R.F. O'Toole, A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens, J. Biol. Methods. 6 (2019) e110. https://doi.org/10.14440/jbm.2019.276.

[146] E. Jaspers, J. Overmann, Ecological Significance of Microdiversity: Identical 16S rRNA Gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies, Appl. Environ. Microbiol. 70 (2004) 4831–4839. https://doi.org/10.1128/AEM.70.8.4831-4839.2004.

[147] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor, Bioinformatics. 34 (2018) i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

[148] R.M. Leggett, R.H. Ramirez-Gonzalez, B.J. Clavijo, D. Waite, R.P. Davey, Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics, Front. Genet. 4 (2013) 1–5. https://doi.org/10.3389/fgene.2013.00288.

[149] D.R. Zerbino, Using Velvet de novo assembler for short-reading sequence technologies, Curr. Protoc. Bioinforma. (2010) 1–13. https://doi.org/10.1002/0471250953.bi1105s31.Using.

[150] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. 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. https://doi.org/10.1089/cmb.2012.0021.

[151] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics. 29 (2013) 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

[152] T. Seemann, Prokka: Rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

[153] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: Rapid Annotations using Subsystems Technology, BMC Genomics. 9 (2008) 75. https://doi.org/10.1186/1471-2164-9-75.

[154] 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.

[155] R.L. Tatusov, M.Y. Galperin, D.A. Natale, E. V Koonin, The COG database : a tool for genome-scale analysis of protein functions and evolution, Nucleic Acids Res. 28 (2000) 33–36.

[156] M.Y. Galperin, Y.I. Wolf, K.S. Makarova, R. Vera Alvarez, D. Landsman, E. V. Koonin, COG database update: focus on microbial diversity, model organisms, and widespread pathogens, Nucleic Acids Res. 49 (2021) D274–D281. https://doi.org/10.1093/nar/gkaa1018.

[157] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (2019) D309–D314. https://doi.org/10.1093/nar/gky1085.

[158] C.A. Ruiz-Perez, R.E. Conrad, K.T. Konstantinidis, MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes, BMC Bioinformatics. 22 (2021) 11. https://doi.org/10.1186/s12859-020-03940-5.

[159] G. Tonkin-Hill, N. MacAlasdair, C. Ruis, A. Weimann, G. Horesh, J.A. Lees, R.A. Gladstone, S. Lo, C. Beaudoin, R.A. Floto, S.D.W. Frost, J. Corander, S.D. Bentley, J. Parkhill, Producing polished prokaryotic pangenomes with the Panaroo pipeline, Genome Biol. 21 (2020) 180. https://doi.org/10.1186/s13059-020-02090-4.

[160] A. Bazin, G. Gautreau, C. Médigue, D. Vallenet, A. Calteau, panRGP: a pangenome-based method to predict genomic islands and explore their diversity, BioRxiv. 1 (2020) 2020.03.26.007484. https://doi.org/10.1101/2020.03.26.007484.

[161] 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, PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph, PLOS Comput. Biol. 16 (2020) e1007732. https://doi.org/10.1371/journal.pcbi.1007732.

[162] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G.G. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: Rapid large-scale prokaryote pan genome analysis, Bioinformatics. 31 (2015) 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

[163] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63. https://doi.org/10.1038/nrg2484.

[164] K.P. Nickerson, S. Senger, Y. Zhang, R. Lima, S. Patel, L. Ingano, W.A. Flavahan, D.K.V. Kumar, C.M. Fraser, C.S. Faherty, M.B. Sztein, M. Fiorentino, A. Fasano, *Salmonella* Typhi colonization provokes extensive

transcriptional changes aimed at evading host mucosal immune defense during early infection of human intestinal tissue, EBioMedicine. 31 (2018) 92–109. https://doi.org/10.1016/j.ebiom.2018.04.005.

[165] A.J. Westermann, S.A. Gorski, J. Vogel, Dual RNA-seq of pathogen and host, Nat. Rev. Microbiol. 10 (2012) 618–630. https://doi.org/10.1038/nrmicro2852.

[166] S. Sarkar, M.T. Heise, Mouse models as resources for studying infectious diseases, Clin. Ther. 41 (2019) 1912–1922. https://doi.org/10.1016/j.clinthera.2019.08.010.

[167] Z. Dimitrov, I. Gotova, E. Chorbadjiyska, *In vitro* characterization of the adhesive factors of selected probiotics to Caco-2 epithelium cell line, Biotechnol. Biotechnol. Equip. 28 (2014) 1079–1083. https://doi.org/10.1080/13102818.2014.969948.

[168] T. Langerholc, P.A. Maragkoudakis, J. Wollgast, L. Gradisnik, A. Cencic, Novel and established intestinal cell line models e An indispensable tool in food science and nutrition, Trends Food Sci. Technol. 22 (2011) S11–S20. https://doi.org/10.1016/j.tifs.2011.03.010.

[169] P. Hoffmann, M. Burmester, M. Langeheine, R. Brehm, M.T. Empl, B. Seeger, G. Breves, Caco-2/HT29-MTX co-cultured cells as a model for studying physiological properties and toxin-induced effects on intestinal cells, PLoS One. 16 (2021) e0257824. https://doi.org/10.1371/journal.pone.0257824.

[170] 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. https://doi.org/10.1111/j.1472-765X.2009.02729.x.

[171] S. ER, A.T. KOPARAL, M. KIVANÇ, Cytotoxic effects of various lactic acid bacteria on Caco-2 cells, TURKISH J. Biol. 39 (2015) 23–30. https://doi.org/10.3906/biy-1402-62.

[172] B. Srinivasan, A.R. Kolli, M.B. Esch, H.E. Abaci, M.L. Shuler, J.J. Hickman, J.J. Srinivasan, B.; Kolli, A.R.; Esch, M.B.; Abaci, H.E.; Shuler, L.; Hickman, B. Srinivasan, A.R. Kolli, M.B. Esch, H.E. Abaci, M.L. Shuler, J.J. Hickman, TEER measurement techniques for in vitro barrier model systems, 2016. https://doi.org/10.1177/2211068214561025.TEER.

[173] L.W. Engel, N.A. Young, Human breast carcinoma cells in continuous culture: a review., Cancer Res. 38 (1978) 4327–39.

[174] S. Shioda, F. Kasai, K. Watanabe, K. Kawakami, A. Ohtani, M. Iemura, M. Ozawa, A. Arakawa, N. Hirayama, E. Kawaguchi, T. Tano, S. Miyata, M. Satoh, N. Shimizu, A. Kohara, Screening for 15 pathogenic viruses in human cell lines registered at the JCRB Cell Bank: characterization of in

vitro human cells by viral infection, R. Soc. Open Sci. 5 (2018) 172472. https://doi.org/10.1098/rsos.172472.

[175] E. Lang, S. Guyot, P. Alvarez-Martin, J.-M. Perrier-Cornet, P. Gervais, Caco-2 Invasion by *Cronobacter sakazakii* and *Salmonella enterica* Exposed to Drying and Heat Treatments in Dried State in Milk Powder, Front. Microbiol. 8 (2017) 1–8. https://doi.org/10.3389/fmicb.2017.01893.

[176] T.R. Chen, D. Drabkowski, R.J. Hay, M. Macy, W. Peterson, WiDr is a derivative of another colon adenocarcinoma cell line, HT-29, Cancer Genet. Cytogenet. 27 (1987) 125–134. https://doi.org/10.1016/0165-4608(87)90267-6.

[177] S.I. Wasserman, K.E. Barrett, P.A. Huott, G. Beuerlein, M.F. Kagnoff, K. Dharmsathaphorn, Immune-related intestinal Cl^- secretion. I. Effect of histamine on the T84 cell line, Am. J. Physiol. Physiol. 254 (1988) C53–C62. https://doi.org/10.1152/ajpcell.1988.254.1.C53.

[178] Y. Sambuy, I. De Angelis, G. Ranaldi, M.L. Scarino, A. Stammati, F. Zucco, The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics, Cell Biol. Toxicol. 21 (2005) 1–26. https://doi.org/10.1007/s10565-005-0085-6.

[179] K. Verhoeckx, P. Cotter, I. López-Expósito, C. Kleiveland, T. Lea, A. Mackie, T. Requena, D. Swiatecka, H. Wichers, The Impact of Food Bioactives on Health, Springer International Publishing, Cham, 2015. https://doi.org/10.1007/978-3-319-16104-4.

[180] M. Pinto, S. Robine Leon, M.D. Appay, Enterocyte-like differentiation and polarization of the human colon carcinoma cell line Caco-2 in culture, Biol. Cell. 47 (1983) 323–330.

[181] G. Ranaldi, R. Consalvo, Y. Sambuy, M.L. Scarino, Permeability characteristics of parental and clonal human intestinal Caco-2 cell lines differentiated in serum-supplemented and serum-free media, Toxicol. Vitr. 17 (2003) 761–767. https://doi.org/10.1016/S0887-2333(03)00095-X.

[182] P.H. Everst, H. Goossens, J.P. Butzler, D. Lloyd, S. Knutton, J.M. Ketley, P.H. Williams, Differentiated Caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *C. coli*, J. Med. Microbiol. 37 (1992) 319–325. https://doi.org/10.1099/00222615-37-5-319.

[183] A. Rosa, Interactions of *Escherichia coli* strains of non-EPEC serogroups that carry eae and lack the EAF and stx gene sequences with undifferentiated and differentiated intestinal human Caco-2 cells, FEMS Microbiol. Lett. 200 (2001) 117–122. https://doi.org/10.1016/S0378-1097(01)00212-9.

[184] M. Gagnon, A.Z. Berner, N. Chervet, C. Chassard, C. Lacroix, Comparison of the Caco-2 , HT-29 and the mucus-secreting HT29-MTX intestinal cell models to investigate *Salmonella* adhesion and invasion, J. Microbiol. Methods. 94 (2013) 274–279. https://doi.org/10.1016/j.mimet.2013.06.027.

[185] J. Elzinga, B. van der Lugt, C. Belzer, W.T. Steegenga, Characterization of increased mucus production of HT29-MTX-E12 cells grown under semi-wet interface with mechanical stimulation, PLoS One. 16 (2021) e0261191. https://doi.org/10.1371/journal.pone.0261191.

[186] J.-F. Sicard, G. Le Bihan, P. Vogeleer, M. Jacques, J. Harel, Interactions of intestinal bacteria with components of the intestinal mucus, Front. Cell. Infect. Microbiol. 7 (2017). https://doi.org/10.3389/fcimb.2017.00387.

[187] R.A. Cone, Barrier properties of mucus, Adv. Drug Deliv. Rev. 61 (2009) 75–85. https://doi.org/10.1016/j.addr.2008.09.008.

[188] A. Alemka, M. Clyne, F. Shanahan, T. Tompkins, N. Corcionivoschi, B. Bourke, Probiotic colonization of the adherent mucus layer of HT29-MTX-E12 cells attenuates *Campylobacter jejuni* virulence properties, Infect. Immun. 78 (2010) 2812–2822. https://doi.org/10.1128/IAI.01249-09.

[189] B.O. Schroeder, Fight them or feed them: How the intestinal mucus layer manages the gut microbiota, Gastroenterol. Rep. 7 (2019) 3–12. https://doi.org/10.1093/gastro/goy052.

[190] A.P. Moran, A. Gupta, L. Joshi, Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract, Gut. 60 (2011) 1412–1425. https://doi.org/10.1136/gut.2010.212704.

[191] P. Paone, P.D. Cani, Mucus barrier, mucins and gut microbiota: the expected slimy partners?, Gut. 69 (2020) 2232–2243. https://doi.org/10.1136/gutjnl-2020-322260.

[192] M.A.M. Vieira, T.A.T. Gomes, A.J.P. Ferreira, T. Knöbl, A.L. Servin, V. Liévin-Le Moal, Two atypical enteropathogenic *Escherichia coli* strains induce the production of secreted and membrane-bound mucins to benefit their own growth at the apical surface of human mucin-secreting intestinal HT29-MTX Cells, Infect. Immun. 78 (2010) 927–938. https://doi.org/10.1128/IAI.01115-09.

[193] J.A. Naughton, K. Mariño, B. Dolan, C. Reid, R. Gough, M.E. Gallagher, M. Kilcoyne, J.Q. Gerlach, L. Joshi, P. Rudd, S. Carrington, B. Bourke, M. Clyne, Divergent Mechanisms of Interaction of *Helicobacter pylori* and *Campylobacter jejuni* with mucus and mucins, Infect. Immun. 81 (2013) 2838–2850. https://doi.org/10.1128/IAI.00415-13.

[194] I. González-Rodríguez, B. Sánchez, L. Ruiz, F. Turroni, M. Ventura, P.

Ruas-Madiedo, M. Gueimonde, A. Margolles, Role of extracellular transaldolase from *Bifidobacterium bifidum* in mucin adhesion and aggregation, Appl. Environ. Microbiol. 78 (2012) 3992–3998. https://doi.org/10.1128/AEM.08024-11.

[195] M. Kankainen, L. Paulin, S. Tynkkynen, I. von Ossowski, J. Reunanen, P. Partanen, R. Satokari, S. Vesterlund, A.P.A. Hendrickx, S. Lebeer, S.C.J. De Keersmaecker, J. Vanderleyden, T. Hamalainen, S. Laukkanen, N. Salovuori, J. Ritari, E. Alatalo, R. Korpela, T. Mattila-Sandholm, A. Lassig, K. Hatakka, K.T. Kinnunen, H. Karjalainen, M. Saxelin, K. Laakso, A. Surakka, A. Palva, T. Salusjarvi, P. Auvinen, W.M. de Vos, Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein, Proc. Natl. Acad. Sci. 106 (2009) 17193–17198. https://doi.org/10.1073/pnas.0908876106.

[196] M. Candela, F. Perna, P. Carnevali, B. Vitali, R. Ciati, P. Gionchetti, F. Rizzello, M. Campieri, P. Brigidi, Interaction of probiotic Lactobacillus and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production, Int. J. Food Microbiol. 125 (2008) 286–292. https://doi.org/10.1016/j.ijfoodmicro.2008.04.012.

[197] G. Theodoropoulos, K.L. Carraway, Molecular signaling in the regulation of mucins, J. Cell. Biochem. 102 (2007) 1103–1116. https://doi.org/10.1002/jcb.21539.

[198] A. Levican, A. Alkeskas, C. Gunter, S.J. Forsythe, M.J. Figueras, Adherence to and invasion of human intestinal cells by *Arcobacter* species and their virulence genotypes, Appl. Environ. Microbiol. 79 (2013) 4951–4957. https://doi.org/10.1128/AEM.01073-13.

[199] A. Villarruel-López, M. Márquez-González, L.E. Garay-Martínez, H. Zepeda, A. Castillo, L.M. De La Garza, E.A. Murano, R. Torres-Vitela, Isolation of *Arcobacter* spp. from retail meats and cytotoxic effects of isolates against Vero cells, J. Food Prot. 66 (2003) 1374–1378. https://doi.org/10.4315/0362-028X-66.8.1374.

[200] A. Villarruel-López, M. Márquez-González, L.E. Garay-Martínez, H. Zepeda, A. Castillo, L. Mota De La Garza, E.A. Murano, R. Torres-Vitela, Isolation of *Arcobacter* spp. from retail meats and cytotoxic effects of isolates against Vero cells, J. Food Prot. 66 (2003) 1374–1378. https://doi.org/10.4315/0362-028X-66.8.1374.

[201] 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.

[202] 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.

[203] V. Lievin-Le Moal, A.L. Servin, Pathogenesis of Human Enterovirulent Bacteria: Lessons from Cultured, Fully Differentiated Human Colon Cancer Cell Lines, Microbiol. Mol. Biol. Rev. 77 (2013) 380–439. https://doi.org/10.1128/mmbr.00064-12.

1.5 Aims and structure of the Ph.D Thesis

The present Ph.D thesis takes into consideration the study of *Arcobacteraceae* species and in particular of *A. butzleri* at different levels. The goal was to obtain genomic and transcriptomic information of different *A. butzleri* strains. The choice to focus on *A. butzleri* was based on the higher isolation rate from different animals, foods and clinical cases compared to other *Arcobacteraceae* species, as reported in literature.

A comparative genomics analysis of 32 *A. butzleri* strains is reported in Chapter 2 jointly with the evaluation of colonization and invasion in contact with *in vitro* mucus producer and not mucus producer cell models. This experimental approach was followed to explore possible links between *A. butzleri* colonization and invasion and genome content. Furthermore, the presence of putative virulence genes and their possible correlation with the strain isolation source was investigated.

The genomes analysis was followed by transcriptomic analysis to evaluate the function of genes annotated on *A. butzleri* genomes. More specifically, in chapter 3 the global gene expression profile by RNAseq of three *A. butzleri* strains is presented. The strains were selected based on data regarding their adhesion and invasion ability and gene expression was evaluated in bacterial cells in contact with a human mucus-producing gut model after 30 and 90 minutes.

In chapter 4 the genome analysis of twenty *Arcobacteraceae* species recently reclassified to genus level is presented. The specific aim was to obtain information regarding the genomic features of the whole bacterial family.

In Chapter 5 general conclusions are presented to show the advancements regarding *Arcobacter* spp. knowledge from the presented data.

2. Functional pangenome analysis reveals high virulence plasticity of *Arcobacter butzleri* and affinity to human mucus

Published as:

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

Genomics. Buzzanca, D. et al. (2021). Elsevier Inc., 113(4), pp. 2065–2076.

doi: 10.1016/j.ygeno.2021.05.001.

2.1. Introduction

Arcobacter butzleri (recently proposed as *Aliarcobacter butzleri*) is a Gram negative bacterium belonging to the *Campylobacteraceae* family often isolated from human stool, animal feces, drinking water, and food [1,2]. It is the most widespread species within the genus *Arcobacter* 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 (cross-contamination 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].

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 and invasion and infer the virulence potential of strains [9]. Although this approach is a simplification of gut systems, it remains fundamental in the phenotypic investigation of host-pathogen interaction [10]. In this context, the intestinal mucus appears to be relevant and may influence the ability of *A. 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 intestinal tract; in the small intestine and in the colon are present one and two mucus layers, 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].

Survey studies have been performed to isolate *Arcobacter* 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.

2.2. Results and discussion

2.2.1 Simulated intestinal colonization is enhanced by human mucus

Thirty-two *A. butzleri* strains previously collected from human and animal feces, pig intestine, animal skin and meat (Table 2.1) were tested on human gut models to define their capability to colonize (cell adhesion) and invade intestinal cells. More specifically, the mixed culture of Caco-2 /HT29-MTX cells and Caco-2 were used as mucus producing (MP) and not-mucus producing (NMP) models, respectively. All *A. butzleri* strains were able to colonize both models after 90 minutes of co-incubation, while strains 7 (isolated from bovine), 26, 28, 31 (isolated from Human feces) presented a non detectable invasion on the MP models; moreover, invasion by strain 28 was not detectable also on NMP models (Figure 2.1). However, only a minor part of the bacterial cells colonizing the models (0.64 ± 0.34 % on average) invaded the human cells, corresponding to an average decrease of 3.7 ± 1.5 Log CFU/cm^2 from the initial inoculum (Supplementary Table 2.1). The multiplicity of infection (MOI) was not the same for all strains.

Although this parameter has been shown previously to play a role in the transepithelial resistance of cell lines after 48 hours of contact with *A. butzleri* [16], it appears not to have an effect during adhesion studies, conducted under short term (1-3 hours) of bacterial contact [17,18]. Overall, our data are in agreement with previous reports and confirm the ability of *A. butzleri* to colonize different cell lines with an invasion efficiency similar to the phylogenetically close species of *Campylobacter jejuni* [19–22].

Comparing colonization data (expressed as ΔLog CFU/cm^2) from MP and NMP models, *A. butzleri* showed an overall greater ($P < 0.001$) colonization capability in presence of human mucus **(Fig. 2.1A)**. 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 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 (Figure 2.1B). Strains from these two sources highlighted an equal proportion of highly colonizing (positive or close to zero ΔLog) and low colonizing phenotypes (negative ΔLog), regardless of the model used. Positive values observed for some strains suggest bacterial growth during the host-pathogen interaction timeframe, again more evident in presence of mucus. Finally, considering differences between strains, strain 2 colonized statistically more ($P < 0.05$) than strain 31 in the NMP model.

The effect of mucus in enhancing colonization has already been observed in *Arcobacter butzleri* [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

chemoattractant function of the mucus towards *Arcobacter* spp. since it was recovered not only from the inner content but also from the mucus layer of pig intestines [5]. The statistically significant higher mucus-model colonization observed for part of the pig isolates also suggests a rather strain-dependent mucus affinity that may result in exploitation of its protective action against intestinal peristalsis under *in vivo* conditions.

Table 2.1. *A. butzleri* strains used in this study. In the table are shown the number of the strains (nr.), C-country of origin (Country), the source of sampling (Source), the specific sampling matrix (Isolation source) and additional information such as official strain codes and information related to the patients from whom the strain was isolated.

Strain code in this study	Source	Additional information	Country
1	Human faeces	Stool sample, (Strain LMG 14714)	Greece
2	Human faeces	Stool sample, (Strain LMG 11119)	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

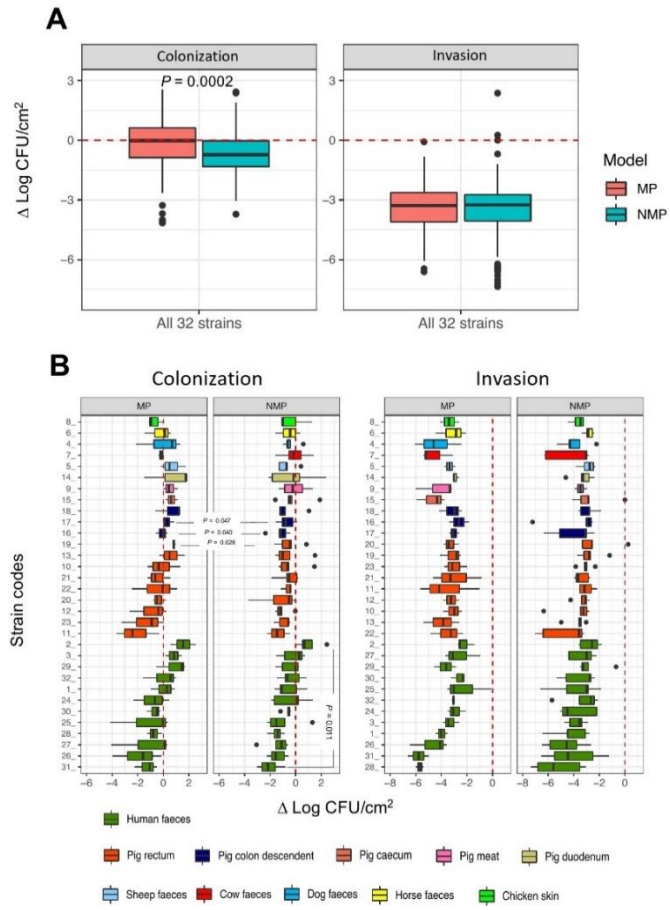


Figure 2.1. Colonization and invasion capabilities on mucus producer (MP) and not-mucus producer (NMP) models are expressed as Δ Log CFU/cm^2 (medians ± interquartile range; n=3; dots=outliers) and shown for all 32 strains together **(A)** and individually for each strain **(B)**. The red dotted line marks the ΔLog equal to 0: a condition in which all bacterial cells added colonized/invaded the model. 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 box-plots color are displayed in the caption. Significant differences between models and among the strains are reported in the graph (*P*-value) employing Wilcoxon's test.

2.2.2 Genomes characteristics and 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 2.2). 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 putative encoded proteomes (Figure 2.2A, Supplementary figure 2.1). We observed a remarkable abundance (19.4 % on average) of proteins with unknown function or characterized only for general functions, a fact that highlights the limited characterization of *A. butzleri* proteome [8]. Following, signal transduction mechanisms (average of 10.14%) is the most abundant COG class, 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 abundant (9.16%) than those related to carbohydrate if compared to the COG distribution observed in other

bacteria [26]. This is consistent with the limited or null consumption of carbohydrates shown by *A. butzleri* and other *Arcobacter* species, which instead utilize organic acids and amino acids as main carbon sources [14,25]. Moreover, the classes of signal transduction mechanism and cell motility represent more than 10 % of the predicted proteins. Only 14 genomes (Supplementary Figure 2.1B) 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 play a pivotal role in the host/pathogen interaction since the bacterial movement in the gel-like matrix such as the shallow mucus layer can allow a faster pathogen invasion [28].

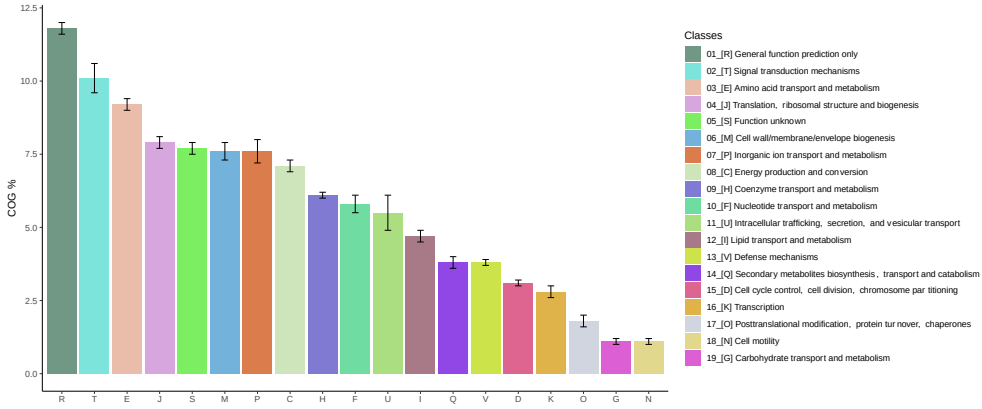


Figure 2.2. Bar-plots displaying the average (± standard deviation) distribution of COG classes in all 32 annotated genomes (% of putative proteins assigned to a class compared to the total putative proteins). Coding keys of classes colors are shown in the caption.

2.2.3 Genome-wide analysis shows an open pangenome

The core- and accessory-genomes sizes were estimated (Table 2.2) by clustering the predicted 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, comprising most of the hypothetical proteins (up to 90 %) and composed of more than 55 % of singletons (gene family exclusively present in one genome). Similar partitioning of *A. butzleri* pangenome has recently been observed on a set of 49 genomes [15]. This leads us to speculate a wide 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 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 will be needed soon to confirm these first observations.

The number of genomes here investigated is adequate to infer the structural organization of the pangenome with Markov Random Field networks (Figure 2.3A), which display the localization of each gene family (nodes) by following a pattern of continuity (edges connect loci that are frequently 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 strains shows organized clusters of persistent gene families (present in more than 95 % of the strains) either surrounded or, less frequently, interrupted by islands of dispensable genes (shell and cloud genome). It is noteworthy the presence of a large pangenome plasticity island that represents hotspots of alternative structural organizations along the genomes analyzed, thus possible sites of HGT (Figure 2.3B). In addition to a predictable presence of hypothetical and not functionally characterized proteins, this island encompasses several accessory gene families generally involved in the COG class of cell wall/membrane and envelope biogenesis, besides proteins more specifically associated with pili/flagella glycosylation, LPS glycosylation/assembling and exopolysaccharide (capsule) secretion (Figure 2.3C). Additionally, we detect up to 323 regions of genome plasticity (RGP) that can be referred as genomic islands shared by at least two of the genomes [34,35]. As observed for the whole pangenome, these RGP overall encompass gene families involved in the cell wall/membrane and envelope biogenesis, which however represents the second class after genes involved in replication, recombination, and repair of DNA (Figure 2.3C). The presence of genes involved in the biosynthesis of 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 dispensable genetic structures that can be acquired or lost, to face host-to-host transition and colonize new ecological niches [36,37]. For instance, loss of the flagellin glycosylation genes may determine phenotypic changes that decrease recognition of strains by the host–immune system [38], while the polysaccharide chain of LPS (inserted between hydrophobic lipid and hydrophilic O-antigen) possess hypervariable structures that reflect the specific pathogen serological signature [39].

Table 2.2. Pangenome partitions estimated by two computational methods.

	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

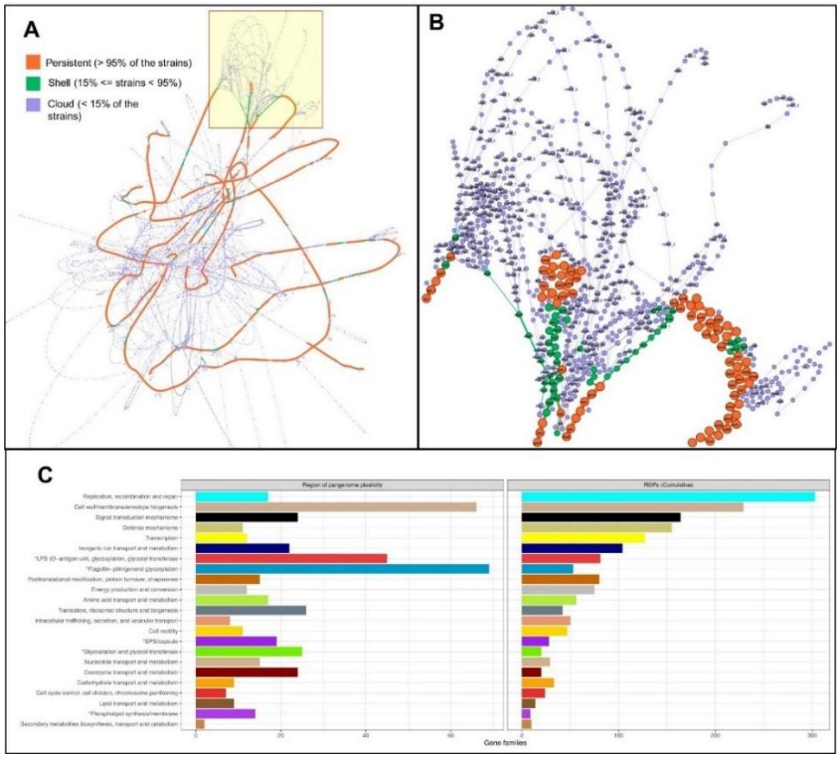


Figure 2.3. Partitioned pangenome network **(A)** displaying the genomic diversity of the 32 strains. Nodes represent the gene families and are colored according to the partition (caption), while their size is proportional to the number of genomes in which are present. Edges connect gene families colocalized in the pangenome and their thickness is proportional to the number of genomes sharing that link. Edges are colored as described for nodes, except for edges between partitions (mixed colors). The frame highlights a broad plasticity region of the pangenome (zoomed-in **B**) harboring shell/cloud gene families alternatively present in the 32 genomes (pangenome plasticity region; Supplementary Table 2.2). Input files (nodes.csv and edges.csv) set up for network visualization in Gephi (https://gephi.org/) are provided on Zenodo (http://doi.org/10.5281/zenodo.4301795). Bar-plots (**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 2.2).

Other genomic regions were entirely constituted by singletons and in some genomes (strains 18, 17, 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. Interestingly, few or no singletons are found in pig duodenum/caecum isolates and 8 of the 9 genomes from pig rectum (Supplementary Figure 2.2). On the contrary, by excluding the singletons we observed that 425 and 140 accessory gene families are significantly (Scoary statistics; *P* [FDR] < 0.05) overrepresented in the genomes of pig rectum and pig duodenum/caecum isolates, respectively. Besides, no overrepresentation of gene families is observed in each of two major ecological sources of isolation, i.e. human stool and all pig intestine. Accordingly, the phylogenetic trees (Figure 2.4) do not show a clear segregation between these two groups, but regardless of the type of input sequences (i.e. 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), group III (strains 12, 19, 20, 21 from pig) and group IV (strains 11, 13, 23 from pig rectum) was observed.

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 *Arcobacter butzleri* may colonize specific segments of pig intestine, as already observed at the species level for *Arcobacter* spp. [5]. Moreover, the low genomic plasticity of these three groups and the fact that pig intestine (particularly the rectum section) can be a favorable niche

for this pathogen with limited 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 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 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 humans, as recently reported at large-scale for *Campylobacter* spp. [38,41].

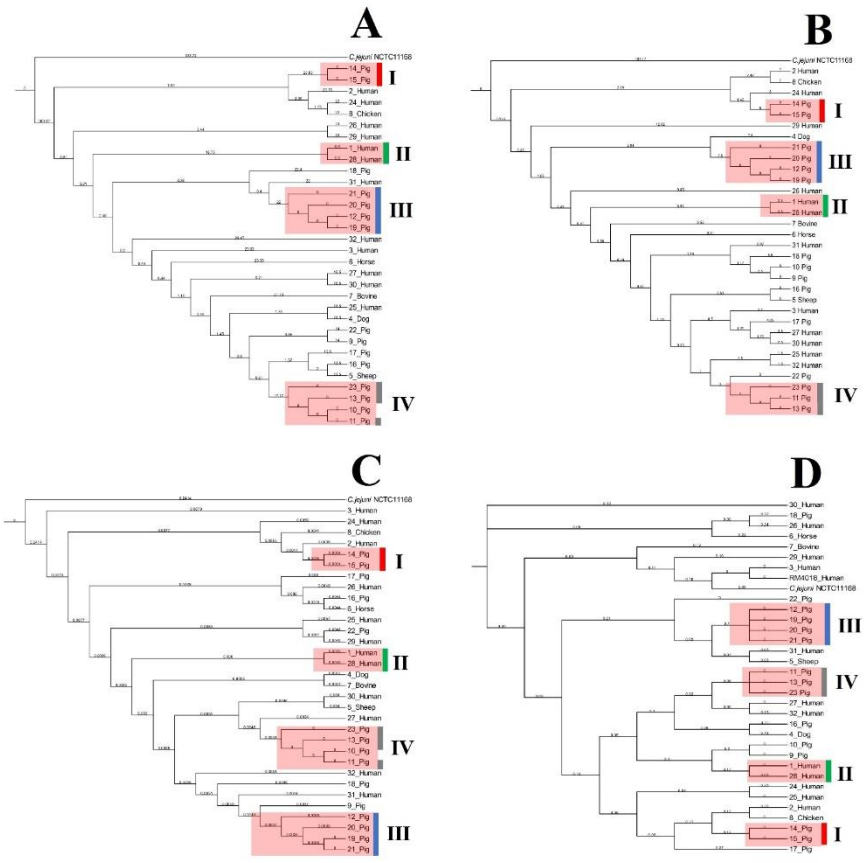


Figure 2.4. Phylogenetic trees of whole genomes (A), core genomes (B), MLST sequences (C) and SNPs (D) of the 32 *A. butzleri* strains. The original source of isolation is indicated and groups of strains that show recurrent clustering patterns are highlighted with colors and named by roman numbers: **I** (strains 14, 15); **II** (strains 1, 28); **III** (strains 12, 19, 20, 21) and **IV (**strains 11, 13, 23**).**

2.2.4 Repertoire of virulence genetic traits

To detect possible genomic signatures linked to *A. butzleri* virulent phenotypes we manually curated the annotated genes by focusing on those sequences putatively associated with host-pathogen interaction in this and

other pathogenic bacteria (Supplementary Table 2.3). This produced a list of 100 genes (of which 39 are accessory genes) putatively involved in functions related to human mucosa adhesion/invasion, interaction with host mucosa/mucus, flagellum and motility, as well as proteins more widely correlated to virulence of *A. butzleri* and other pathogens, such as hemolysis, secretion and regulatory systems (Figure 2.5). It is noteworthy that the 32 presence/absence profiles (Pearson's correlation-based dendrogram) cluster as previously observed in the whole-genome phylogenetic dendrogram and almost all these genes are included in accessory gene families of the pangenome. Accordingly, we might speculate that the biodiversity within *A. butzleri* populations is partly shaped from the exchangeable virulome as a genomic tracking of the host-to-host transitions undergone by each strain. Nevertheless, strains origin and other genes not directly involved in virulence mechanisms may play an important role in the phylogenetic segregation of the strains.

2.2.5 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 *Arcobacter* 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.

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 % 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 based studies and other genomic

surveys [44,45], these latter four genes are less prevalent across the 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 colonization phenotypes. This is not surprising since they encode for functions useful in following 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 to underestimating virulence potential and diffusion of *A. butzleri* due to false negatives amplifications [14,15].

2.2.6 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 2.3) partially confirms the previous observations, with a grouping pattern unrelated to the initially shared origin, but instead resembling the phylogenetic segregation described above (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 through the human mucus (Figure 2.1A). 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 (Figure 2.1B) [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].

2.2.7 Secretion systems involved in pathogenicity

Several genes of our proposed virulome are part of secretion systems and can play a role in the host-pathogen 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.

2.2.8 Genomic signatures recognized by host immune response

A consistent fraction of putative virulence genes are involved in the flagellar assembling/motility (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, the flagellar proteins are important virulence factors related to human pathogens motility in the proximal mucus layer and recognition by host immune response [13]. Thus, it is intriguing that gene families encoding flagellins are included in the core genome of *A. butzleri*, while we found genes responsible for their glycosylation in the accessory and plastic genomic regions. This suggests the heterogeneous glycans compositions of flagellum may lead to a strain-specific antigenic fraction of this bacterial component [58].

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 2.4, Supplementary Table 2.4). 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 biosynthesis protein (*msbB*), LPS transferases (*rfaC*), sugars/glycosides transferases/epimerase/reductase (*rfaF, sunS, pglJ, lacA, epsJ, rfbB, rfbC, rmlD, kfoC*). The lack of a single component of the O-antigen genes cluster (ABC transporters, glycotransferase, glycosyltransferase) can dramatically affect the infectiveness of Gram-negative pathogens [60,61]. Therefore, the role of such variability regarding the genes flanking the O-antigen ligase genes in 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 genomic comparison, highlights this region of the plastic virulome as one of the most useful to define strain-specific virulence signatures.

2.2.9 Genes involved in multiple virulence mechanisms and regulation

Other meaningful elements of the *A. butzleri* virulome (Figure 2.5) 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 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 [69]. Considering that *mprA* exerts a transcription regulation activity comparable to *phoP*, the 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 *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 bacteria, such as the virulence sensor *bvgS* [70] or the couple of genes *dsbA/dsbB*, here annotated as DSBA-like thioredoxin domain protein and thiol-disulfide oxidoreductase *resA*, respectively. 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.

Figure 2.5. 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.

2.3. 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 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 genomic features for the host adaptation of this pathogen. Here, also, to confirm these first findings, we link the variable virulome to strains phenotypes, by identifying in the LPS assembling pathway one potential strain-specific signature. Despite the intrinsic limit of pangenomic based comparison 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 recognition markers, such as master regulators.

Therefore, these outcomes will provide concrete guidelines for more comprehensive omics investigation of the *A. butzleri* lifestyle in human mucosa. Future studies are needed to confirm the characteristics of *A.*

butzleri detected with a genome study approach. For this purpose, analyzes focused on strains with specific knock-out genes will allow to confirm their function in *A. butzleri*. Moreover, the HGT already demonstrated in bacterial co-cultures of *C. jejuni*, requires additional study in *A. butzleri* [73]. *In vitro* study will elucidate the ability of *A. butzleri* to incorporate DNA sequences in their genome.

2.4. Materials and methods

2.4.1 Bacterial strains

The *A. butzleri* strains (**Table 2.1**) 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) [74].

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.

2.4.2 Cell lines and human gut models

Human colon carcinoma cell lines Caco-2 (86010202, ECACC, European Collection of Authenticated Cell Cultures, Public Health England), HT29 (ATCC® HTB38, ECACC) and HT29 MTX (12040401, ECACC) were cultured in Dulbecco's Modified Eagle's Medium (DMEM 6429; Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% of fetal bovine serum (FBS; F7524 Sigma-Aldrich) and EmbryoMax Penicillin-

Streptomycin Solution, 100X (TMS-AB2-C, Sigma-Aldrich). The cell lines were grown in 25 and 75 cm^2 culture flasks (Corning, New York, New York USA) at 37 °C in a humidified atmosphere containing 5% CO_2 and 95% air and sub passaged every 3-4 days (Eppendorf, Galaxy 170 S, Hamburg, Germany) [75].

Two *in vitro* monolayer human epithelial structures were prepared: a mucus-secreting (MP) co-culture of differentiated Caco-2 and HT29-MTX cells in a 9/1 ratio; and two non-mucus-secreting 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 [76]. Briefly, the cells were seeded at a density of 35,000 cells cm^{-2} and grown in complete culture medium under the same conditions described above, for 14–15 days with regular changes of the media, until functional polarization was reached and models could be considered differentiated and ready for the experiments [77]. Before (3-4 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 antibiotics to allow the pathogen growth.

2.4.3 Assessment of colonization and invasion capability

The working suspensions of the strains were inoculated on MP and NMP cell models. Depending on 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 Log_{10} CFU mL^{-1} (**Supplementary Table 2.1)**. Due to this experimental limitation, the multiplicity of infection (MOI) was not the same for all strains tested. Colonization-invasion assays were performed on two different model wells for each biological replicate. After 90 minutes of co-incubation at 37

°C in a normal 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 biological replicates.

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 cm^2 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 μg 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 [20,78,79].

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

2.4.4 Genome sequencing, annotation and bioinformatic analysis

Genomic DNA (gDNA) extraction of *A. butzleri* strains was performed by the beads-beating, phenol-chloroform DNA extraction method followed by a RNAse A (5 µg μl^{-1}, MRNA092 Epicenter, Madison, Wisconsin, U.S.A) treatment to digest RNA in the DNA samples, with an incubation of 30 min at 37 °C. The DNA quantification was performed with the employment of Nanodrop (ND 1000, Thermo SCIENTIFIC). The gDNA quality check, to confirm the absence of degradation and 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 used as DNA intercalating.

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 quantification 1 µg of gDNA was used for the library preparation using NEBNext® library prep Kit, 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. The PCR product purification was performed with the use of the AMPure XP system, afterwards, the libraries were analyzed by Agilent 2100 Bioanalyzer (size distribution) and quantified using real-time PCR.

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

Reads were *de novo* assembled with SPAdes (version 3.11.0) [80] and the quality of the contigs was checked with QUAST software to obtain statistics related to the genomes assembly process and data quality, such

as coverage, total genome bp length and the number of contigs [81] (**Supplementary Table 2.2**).

Genomes were annotated using the Prokka (version 1.11) suite [82] and putative encoded proteins have been manually checked through on UniProt (https://www.uniprot.org/), Pfam (https://pfam.xfam.org/), and CDD database (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to understand their functional role [83]. The CRISPR-CAS sequences have been detected with the software CRISPRCasFinder 1.1.2 (https://crisprcas.i2bc.paris-saclay.fr/), while phage sequences were retrieved with Phaster (https://phaster.ca/) [84,85]. Additional analysis on the metabolic pathway was performed on the putative predicted proteome with the software RPS-BLAST 2.2.15 on WebMga (http://weizhong-lab.ucsd.edu/webMGA/), to obtain the related COGs (Clusters of Orthologous Groups) codes and classes [86–88].

Proteins inferred by Prokka were then processed with the parallel use of Roary (version 3.13.0) and 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 the pangenome was inferred through the matrix generated by PPanGGOlin and visualized using the 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]. Associations between binary matrices (presence/absence) of accessory gene families or RGP (singletons excluded) and the main sources of isolation (human stool, pig intestine, and its main sections) were assessed with Scoary scripts [89] and considered significant for *P*-value [FDR 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 2.4**). 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 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [90].

2.4.5 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) [91].

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 [92], by using the sequences used for *Arcobacter* spp., specifically *aspA, atpA, glnA, gltA, glyA, pgm, tkt* [93]. After the obtainment of the MLST numeric codes **(Supplementary Table 2.5)**, 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) [94].

Phylogenetic trees were visualized with iTOL (version 5.5.1) to obtain the image format choice [95], while the software Morpheus (https://software.broadinstitute.org/morpheus) was used in the heatmap production [96].

2.4.6 Statistical analysis

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

Normality and homogeneity of the data from colonization and invasion assays were checked using Shapiro-Wilk's W and Levene's tests, respectively. Kruskal–Wallis (K-W) and ANOVA were used to assess the overall variation and differences between the multiple groups, for nonparametric and 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 (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.

2.5. Availability of data and material

Raw sequence reads were deposited at the Sequence Read Archive of the National Center for Biotechnology Information (Bioproject accession number: PRJNA660594). The genomes assembled sequences, the sequences of the predicted transcripts and amino acidic sequences (.faa, .fna, .gff, .gbf, .sqn, .tbl, .ffn), and the files used to construct the pangenome network (edges.csv, nodes.csv) are available on Zenodo (https://zenodo.org/) at http://doi.org/10.5281/zenodo.4301795. The study showed in this chapter has been adapted from the version published on the journal Genomics in 2021 (10.1016/j.ygeno.2021.05.001).

2.6 References

[1] D. Chieffi, F. Fanelli, V. Fusco, *Arcobacter butzleri*: Up-to-date taxonomy, ecology, and pathogenicity of an emerging pathogen, Compr. Rev. Food Sci. Food Saf. (2020) 1541-4337.12577. https://doi.org/10.1111/1541-4337.12577.

[2] A. Oren, G.M. Garrity, List of new names and new combinations previously effectively, but not validly, published, Int. J. Syst. Evol. Microbiol. 70 (2020) 1–5. https://doi.org/10.1099/ijsem.0.003881.

[3] 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.

[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.

[5] S. De Smet, L. De Zutter, K. Houf, Spatial Distribution of the Emerging Foodborne Pathogen Arcobacter in the Gastrointestinal Tract of Pigs, Foodborne Pathog. Dis. 9 (2012) 1097–1103. https://doi.org/10.1089/fpd.2012.1184.

[6] N. Shange, P. Gouws, L.C. Hoffman, *Campylobacter* and *Arcobacter* species in food-producing animals: prevalence at primary production and during slaughter, World J. Microbiol. Biotechnol. 35 (2019) 146. https://doi.org/10.1007/s11274-019-2722-x.

[7] C.L. Hilton, B.M. Mackey, A.J. Hargreaves, S.J. Forsythe, The recovery of *Arcobacter butzleri* NCTC 12481 from various temperature treatments, J. Appl. Microbiol. 91 (2001) 929–932. https://doi.org/10.1046/j.1365-2672.2001.01457.x.

[8] T.P. Ramees, K. Dhama, K. Karthik, R.S. Rathore, A. Kumar, M. Saminathan, R. Tiwari, Y.S. Malik, R.K. Singh, *Arcobacter* : an emerging food-borne zoonotic pathogen, its public health concerns and advances in diagnosis and control – a comprehensive review, Vet. Q. 37 (2017) 136–161. https://doi.org/10.1080/01652176.2017.1323355.

[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, Front. Physiol. 9 (2018). https://doi.org/10.3389/fphys.2018.01584.

[10] G.P. Donaldson, S.M. Lee, S.K. Mazmanian, Gut biogeography of the bacterial microbiota, Nat. Rev. Microbiol. 14 (2016) 20–32. https://doi.org/10.1038/nrmicro3552.

[11] A.P. Moran, A. Gupta, L. Joshi, Sweet-talk: role of host glycosylation in

bacterial pathogenesis of the gastrointestinal tract, Gut. 60 (2011) 1412–1425. https://doi.org/10.1136/gut.2010.212704.

[12] G.C. Hansson, Role of mucus layers in gut infection and inflammation, Curr. Opin. Microbiol. 15 (2012) 57–62. https://doi.org/10.1016/j.mib.2011.11.002.

[13] J.-F. Sicard, G. Le Bihan, P. Vogeleer, M. Jacques, J. Harel, Interactions of Intestinal Bacteria with Components of the Intestinal Mucus, Front. Cell. Infect. Microbiol. 7 (2017). https://doi.org/10.3389/fcimb.2017.00387.

[14] F. Fanelli, A. Di Pinto, A. Mottola, G. Mule, D. Chieffi, F. Baruzzi, G. Tantillo, V. Fusco, Genomic Characterization of *Arcobacter butzleri* Isolated From Shellfish: Novel Insight Into Antibiotic Resistance and Virulence Determinants, Front. Microbiol. 10 (2019) 1–17. https://doi.org/10.3389/fmicb.2019.00670.

[15] 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. 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.

[17] M.A. Bianchi, D. Del Rio, N. Pellegrini, G. Sansebastiano, E. Neviani, F. Brighenti, A fluorescence-based method for the detection of adhesive properties of lactic acid bacteria to Caco-2 cells, Lett. Appl. Microbiol. 39 (2004) 301–305. https://doi.org/10.1111/j.1472-765X.2004.01589.x.

[18] 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. 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.

[21] A. Klančnik, I. Gobin, B. Jeršek, S. Smole Možina, D. Vučković, M. Tušek Žnidarič, M. Abram, Adhesion of *Campylobacter jejuni* Is Increased in Association with Foodborne Bacteria, Microorganisms. 8 (2020) 201. https://doi.org/10.3390/microorganisms8020201.

[22] P.H. Everst, H. Goossens, J.P. Butzler, D. Lloyd, S. Knutton, J.M. Ketley, P.H. Williams, Differentiated Caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *C. coli*, J. Med. Microbiol. 37 (1992) 319–325. https://doi.org/10.1099/00222615-37-5-319.

[23] A. Alemka, N. Corcionivoschi, B. Bourke, Defense and Adaptation: The Complex Inter-Relationship between *Campylobacter jejuni* and Mucus, Front. Cell. Infect. Microbiol. 2 (2012) 1–6. https://doi.org/10.3389/fcimb.2012.00015.

[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. https://doi.org/10.1371/journal.pone.0085487.

[25] F. Rovetto, A. Carlier, A.M. Van Den Abeele, K. Illeghems, F. Van Nieuwerburgh, L. Cocolin, K. Houf, Characterization of the emerging zoonotic pathogen *Arcobacter thereius* by whole genome sequencing and comparative genomics, PLoS One. 12 (2017) 1–27. https://doi.org/10.1371/journal. pone.0180493.

[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.

[27] M.T. Cabeen, C. Jacobs-wagner, The Bacterial Cytoskeleton, Annu. Rev. Of Genetics. (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. https://doi.org/10.1016/j.celrep.2019.04.106.

[29] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G.G. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: Rapid large-scale prokaryote pan genome analysis, Bioinformatics. 31 (2015) 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

[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, PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph, PLOS Comput. Biol. 16 (2020) e1007732. https://doi.org/10.1371/journal.pcbi.1007732.

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

[32] T. Gouliouris, K.E. Raven, C. Ludden, B. Blane, J. Corander, C.S. Horner, J. Hernandez-Garcia, P. Wood, N.F. Hadjirin, M. Radakovic, M.A.

Holmes, M. de Goffau, N.M. Brown, J. Parkhill, S.J. Peacock, Genomic surveillance of *Enterococcus faecium* reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom, MBio. 9 (2018) 1–15. https://doi.org/10.1128/mBio.01780-18.

[33] M.R. Davies, L. McIntyre, A. Mutreja, J.A. Lacey, J.A. Lees, R.J. Towers, S. Duchêne, P.R. Smeesters, H.R. Frost, D.J. Price, M.T.G. Holden, S. David, P.M. Giffard, K.A. Worthing, A.C. Seale, J.A. Berkley, S.R. Harris, T. Rivera-Hernandez, O. Berking, A.J. Cork, R.S.L.A. Torres, T. Lithgow, R.A. Strugnell, R. Bergmann, P. Nitsche-Schmitz, G.S. Chhatwal, S.D. Bentley, J.D. Fraser, N.J. Moreland, J.R. Carapetis, A.C. Steer, J. Parkhill, A. Saul, D.A. Williamson, B.J. Currie, S.Y.C. Tong, G. Dougan, M.J. Walker, Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics, Nat. Genet. 51 (2019) 1035–1043. https://doi.org/10.1038/s41588-019-0417-8.

[34] M. Juhas, J.R. Van Der Meer, M. Gaillard, R.M. Harding, D.W. Hood, D.W. Crook, Genomic islands: Tools of bacterial horizontal gene transfer and evolution, FEMS Microbiol. Rev. 33 (2009) 376–393. https://doi.org/10.1111/j.1574-6976.2008.00136.x.

[35] A. Bazin, G. Gautreau, C. Médigue, D. Vallenet, A. Calteau, panRGP: a pangenome-based method to predict genomic islands and explore their diversity, BioRxiv. 1 (2020) 2020.03.26.007484. https://doi.org/10.1101/2020.03.26.007484.

[36] U. Sood, P. Hira, R. Kumar, A. Bajaj, D.L.N. Rao, R. Lal, M. Shakarad, Comparative genomic analyses reveal core-genome-wide genes under positive selection and major regulatory hubs in outlier strains of *Pseudomonas aeruginosa*, Front. Microbiol. 10 (2019). https://doi.org/10.3389/fmicb.2019.00053.

[37] U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, Genomic islands in pathogenic and environmental microorganisms, Nat. Rev. Microbiol. 2 (2004) 414–424. https://doi.org/10.1038/nrmicro884.

[38] E. Mourkasa, A.J. Taylorb, G. Mérica, S.C. Baylissa, B. Pascoea, L. Mageirosa, J.K. Callanda, M.D. Hitchingse, A. Ridleyf, A. Vidalf, K.J. Forbesg, N.J.C.S. Strachanh, C.T. 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 *Campylobacter jejuni*, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 11018–11028. 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.

[41] D.J. Woodcock, P. Krusche, N.J.C. Strachan, K.J. Forbes, F.M. Cohan, G. Méric, S.K. 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) 9650. 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 Abeele, K. Houf, Occurrence of Putative Virulence Genes in *Arcobacter* Species Isolated from Humans and Animals, J. Clin. Microbiol. 50 (2012) 735–741. https://doi.org/10.1128/JCM.05872-11.

[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 Saf. 8 (2019). https://doi.org/10.4081/ijfs.2019.7840.

[45] C. Girbau, C. Guerra, I. Martínez-Malaxetxebarria, R. Alonso, A. Fernández-Astorga, Prevalence of ten putative virulence genes in the emerging foodborne pathogen *Arcobacter* isolated from food products, Food Microbiol. 52 (2015) 146–149. 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. 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. Chen, Q. Zhang, Point mutations in the major outer membrane protein drive hypervirulence of a rapidly expanding clone of *Campylobacter jejuni*, Proc. Natl. Acad. Sci. 113 (2016) 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.-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) 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. 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 (2010) 120.

https://doi.org/10.1186/1471-2164-11-120.

[50] A.P. Corfield, S.A. Wagner, J.R. Clamp, M.S. Kriaris, L.C. Hoskins, Mucin degradation in the human colon: production of sialidase, sialate O-acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria., Infect. Immun. 60 (1992) 3971–3978. https://doi.org/10.1128/iai.60.10.3971-3978.1992.

[51] M. Sandkvist, Type II Secretion and Pathogenesis, Infect. Immun. 69 (2001) 3523–3535. https://doi.org/10.1128/IAI.69.6.3523.

[52] T. Yoshida, Y. Ayabe, M. Yasunaga, Y. Usami, H. Habe, H. Nojiri, T. Omori, Genes involved in the synthesis of the exopolysaccharide methanolan by the obligate methylotroph *Methylobacillus* sp. strain 12S, Microbiology. 149 (2003) 431–444. https://doi.org/10.1099/mic.0.25913-0.

[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 of a long, galactose-rich exopolysaccharide in *Lactobacillus rhamnosus* GG and functional analysis of the priming glycosyltransferase, Appl. Environ. Microbiol. 75 (2009) 3554–3563. https://doi.org/10.1128/AEM.02919-08.

[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) 82–88. https://doi.org/10.1016/j.ijfoodmicro.2013.01.003.

[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. https://doi.org/10.3389/fmicb.2019.01078.

[56] R.A. Batchelor, Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species, Microbiology. 150 (2004) 3507–3517. https://doi.org/10.1099/mic.0.27112-0.

[57] C.E. Alvarez-Martinez, P.J. Christie, Biological Diversity of Prokaryotic Type IV Secretion Systems, Microbiol. Mol. Biol. Rev. 73 (2009) 775–808. https://doi.org/10.1128/mmbr.00023-09.

[58] S. Merino, J. Tomás, Gram-Negative Flagella Glycosylation, Int. J. Mol. Sci. 15 (2014) 2840–2857. https://doi.org/10.3390/ijms15022840.

[59] 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, Biochim. Biophys. Acta - Biomembr. 1828 (2013) 1143–1152. 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. https://doi.org/10.1111/1574-6976.12070.

[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. https://doi.org/10.1093/femsre/fuw007.

[62] 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 Virulence of *Shigella dysenteriae*, Infect. Immun. 68 (2000) 6329–6336. 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 Uropathogenic *Escherichia coli*: Aerobactin and Heme Transport and TonB Are Required for Virulence in the Mouse, Infect. Immun. 69 (2001) 6179–6185. 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 TonB-Dependent 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/10.1186/1756-0500-5-327.

[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. 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. Groisman, Dissecting the PhoP regulatory network of Escherichia coli and *Salmonella enterica*, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2862–2867. 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) 202–212. https://doi.org/10.1128/JB.187.1.202.

[70] Y. Eguchi, T. Okada, S. Minagawa, T. Oshima, H. Mori, K. Yamamoto, A. Ishihama, R. Utsumi, Signal Transduction Cascade between EvgA / EvgS and PhoP / PhoQ Two-Component Systems of *Escherichia coli*, J.

Bacteriol. 186 (2004) 3006–3014. https://doi.org/10.1128/JB.186.10.3006.

[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 Compartmentalized Host and Pathogen Strategies during *Salmonella* Infection, MBio. 11 (2020). https://doi.org/10.1128/mBio.03348-19.

[73] D.L. Wilson, J.A. Bell, V.B. Young, S.R. Wilder, L.S. Mansfield, J.E. Linz, Variation of the natural transformation frequency of *Campylobacter jejuni* in liquid shake culture, Microbiology. 149 (2003) 3603–3615. https://doi.org/10.1099/mic.0.26531-0.

[74] K. Houf, R. Stephan, Isolation and characterization of the emerging foodborn pathogen *Arcobacter* from human stool, J. Microbiol. Methods. 68 (2007) 408–413. https://doi.org/10.1016/j.mimet.2006.09.020.

[75] 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). https://doi.org/10.1371/journal.pone.0094457.

[76] 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.

[77] 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.

[78] S. Backert, D. Hofreuter, Molecular methods to investigate adhesion, transmigration, invasion and intracellular survival of the foodborne pathogen *Campylobacter jejuni*, J. Microbiol. Methods. 95 (2013) 8–23. https://doi.org/10.1016/j.mimet.2013.06.031.

[79] A. Levican, A. Alkeskas, C. Gunter, S.J. Forsythe, M.J. Figueras, Adherence to and Invasion of Human Intestinal Cells by *Arcobacter* Species and Their Virulence Genotypes, Appl. Environ. Microbiol. 79 (2013) 4951–4957. https://doi.org/10.1128/AEM.01073-13.

[80] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. 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.

https://doi.org/10.1089/cmb.2012.0021.

[81] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics. 29 (2013) 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

[82] T. Seemann, Prokka: Rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

[83] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. 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. https://doi.org/10.1093/nar/gku1221.

[84] 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. https://doi.org/10.1093/nar/gkm360.

[85] 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–W21. https://doi.org/10.1093/nar/gkw387.

[86] A. Marchler-Bauer, CDD: a Conserved Domain Database for protein classification, Nucleic Acids Res. 33 (2004) D192–D196. https://doi.org/10.1093/nar/gki069.

[87] 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.

[88] 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. https://doi.org/10.1186/1471-2164-12-444.

[89] O. Brynildsrud, J. Bohlin, L. Scheffer, V. Eldholm, Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary, Genome Biol. 17 (2016) 238. https://doi.org/10.1186/s13059-016-1108-8.

[90] S. McGinnis, T.L. Madden, BLAST: At the core of a powerful and diverse set of sequence analysis tools, Nucleic Acids Res. 32 (2004) 20–25. https://doi.org/10.1093/nar/gkh435.

[91] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. 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.

[92] M. V Larsen, S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R.L. Marvig, L. Jelsbak, T. 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. https://doi.org/10.1128/JCM.06094-11.

[93] W.G. Miller, I. V. Wesley, S.L.W. On, K. Houf, F. Mégraud, G. Wang, E. Yee, A. Srijan, C.J. Mason, First multi-locus sequence typing scheme for Arcobacter spp., BMC Microbiol. 9 (2009) 196. https://doi.org/10.1186/1471-2180-9-196.

[94] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments, PLoS One. 5 (2010) e9490. https://doi.org/10.1371/journal.pone.0009490.

[95] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, Nucleic Acids Res. 47 (2019) W256–W259. https://doi.org/10.1093/nar/gkz239.

[96] M.C. Ryan, M. Stucky, C. Wakefield, J.M. Melott, R. Akbani, J.N. Weinstein, B.M. Broom, Interactive Clustered Heat Map Builder: An easy web-based tool for creating sophisticated clustered heat maps, F1000Research. 8 (2020) 1750. https://doi.org/10.12688/f1000research.20590.2.

3. Transcriptome evaluation of *Arcobacter butzleri* in contact with a mucus producer human gut model

3.1. Introduction

Arcobacter butzleri (recently proposed as *Aliarcobacter butzleri*) is a Gram negative species belonging to *Arcobacteraceae* family, isolated from human stool samples and from different kind of food such as meat, milk and vegetables. For this reason *A. butzleri* is currently considered a foodborne pathogen [1,2]. In humans, this bacterium has been associated with different gastrointestinal symptoms such as diarrhea and abdominal pains [2,3]. *A. butzleri* was isolated from different animals such as clams, chicken, pig, bovine, dog and cat [3]. The animals from which *A. butzleri* is isolated are usually asymptomatic making difficult the presence estimation of this bacterium on farms [3]. The increase of *A. butzleri* antibiotic resistance and the reported human clinical cases linked to *A. butzleri* infections highlight the necessity to expand the currently limited knowledge regarding this microorganism [2,3].

Knowledge about *A. butzleri* has increased particularly in recent years due to whole-genome sequencing studies allowing genome exploration and functional annotation of different sequences associated with virulence and antibiotic resistance [4–6]. Ten currently considered putative virulence genes (*cadF, ciaB, cj1349, hecA, hecB, mviN, pldA, irgA, tlyA* and *iroE*) are the most studied and detected in many works reinforcing the need to obtain new data on these genes and to detect new ones in order to obtain sequences directly attributable to virulent bacterial phenotype to better understand the virulence mechanism of *A. butzleri* and obtain sequences on which design new molecular detection methods [2,7,8]. This need has been partially met in recent years through comparative genomics studies of *A. butzleri* in which genes with different functions, directly and indirectly related to virulence, have been identified. Sequences encoding

for functions related to expression regulation, flagella functionality, chemotaxis, urease activity and iron transport have been highlighted as putative virulence genes. Part of these putative genes are characterized by sequence hypervariability among strains, this is particularly evident for the genes encoding the PorA protein and genes encoding lipopolysaccharides synthesis and O-antigen production [5,6].

Genomic analyses, when coupled to physiological observation, can lead to the identification of putative virulence related mechanisms. For this purpose, the use of *in vitro* cell models is nowadays an important step in determining the effective bacterial capacity to adhere and enter host cells. In the case of intestinal pathogens, *in vitro* human gut models are often used to test these microorganisms. In particular, cell lines from human colon adenocarcinoma like Caco-2 (differentiation into enterocyte-like cells) and HT29-MTX-E12 (mucus production) are used in pure or combined culture to mimic the bacterial infection process [9]. Further, it is possible to retrieve the bacterial cells during colonization/invasion to perform transcriptomic analyses [10]. Such *in vitro* studies with *A. butzleri* have shown its ability to adhere and invade host cells, with a particular affinity towards mucus produced by HT29-MTX-E12 cell line [6,9,11].

Gene annotation in *A. butzleri* allowed the identification of different putative virulence genes however, the gene expression of *A. butzleri* under infection conditions remains to be investigated [4–6].

The present study aims to analyse the *A. butzleri* transcriptome in contact with intestinal *in vitro* mucus producer human cells layer mimicking the infection process after 30 minutes (30') and 90 minutes (90') of host-bacteria contact.

3.2. Results and Discussion

3.2.1 Colonization and invasion ability of *A. butzleri* and transcriptome analysis

The strains LMG 11119, LMG 10828^T and 31 were isolated from human clinical samples in different geographic areas (Italy, U.S.A and Belgium, respectively). These strains have been selected for their differences linked to genomic (Figure 3.1A) and virulence features as presented in a previous work [6]. Previously, a higher adhesion of *A. butzleri* to mucus-producing mixed cell models Caco-2/HT29-MTX-E12 (9/1 ratio) was observed, when compared to colonization data tested on Caco-2 (non-mucus producing cell lines) models and mixed Caco-2/HT29 (9/1 ratio) [6]. The higher colonization of *A. butzleri* in presence of mucus is in agreement with observations in other Gram negative pathogens including *Campylobacter jejuni* and *Helicobacter pylori* that can interact with intestinal mucus mucins (glycoproteins) [12–15].

An *in vitro* mucus producing cell model (Caco-2/HT29-MTX-E12) was used to evaluate the ability of *A. butzleri* to colonize and invade the cell layer after 30 'and 90' of bacteria-host cells contact. All strains showed the ability to colonize (bacteria present on/in the cells) the cell layer at both times, but strain LMG 11119 presented the highest colonization after 30' (*p-value < 0.05*) and 90' (*p-value < 0.01*) (Supplementary table 3.1). Regarding invasion, strain LMG 10828^T did not show a detectable load related to internalized bacteria after 30' of bacteria-cell line contact. At 90' all three strains were detected intracellularly; this is a confirmation of the ability to invade human cell lines. Although strain LMG 11119 showed a tendency towards greater invasion it was not statistically significant (*p-*

value > 0.05) (Figure 3.1B) (Supplementary Table 3.1).The data highlight a different behavior of *A. butzleri* strains tested, which could also be relevant for their virulence potential. Overall, the *in vitro* ability of *A. butzleri* strains to colonize mucus producing human gut cell models was confirmed [6].

The different colonization and internalization behavior of the host cells by the three *A. butzleri* strains could be related with differences in gene expression. Therefore, an analysis of the entire transcriptome of the strains was carried out. This assessment of the transcriptome was performed for the strains in contact with the human cell line at timepoints 30'and 90'.

In addition, the gene expression of the strains was evaluated after growth in Arcobacter agar media and in Dulbecco's Modified Eagle Medium (DMEM) after 2 hours of adaptation by incubation at 37 °C (5% CO_2). In this way, changes in gene expression not directly linked to contact with the human cell line could be detected. Further, the DMEM adaptation was used as a control condition to evaluate the presence of differentially expressed genes (DEGs) when the strains were placed in contact with the human cell line.

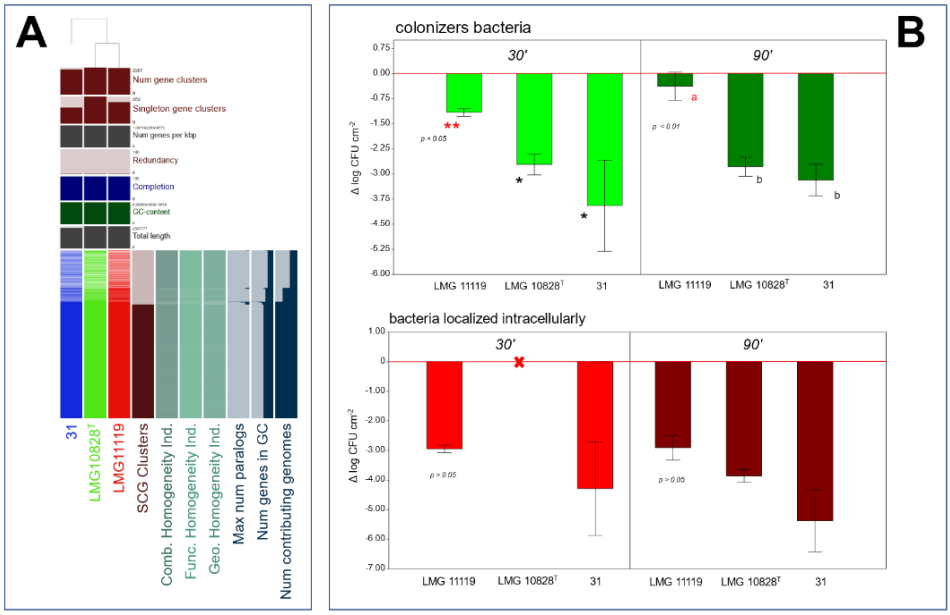


Figure 3.1. Genomic comparison between *A. butzleri* strains tested (A) and *in vitro* colonization – invasion assay (B). (**A**) The genomes of LMG 11119, LMG 10828^T (reference strain) and 31 have been compared with Anvi'o tool. The dendrogram shows gene cluster presence/absence, as well as general information about genomes. The bar chart (**B**) shows the colonization (colonizers bacteria, adhered and internalized bacteria to host cells) and bacteria localized intracellularly after 30' (light colors) and 90' (dark colors). The values are indicated as Δlog CFU cm^{-2} correspondents to adhered and internalized bacteria (see paragraph 4.3 for calculation specifications). The error bars represent the standard errors while red lines indicate a Δlog of 0 (bacterial load detected equal to initial bacterial inoculum load inoculated into cell models). The red X indicates a bacterial load not detected indicating the absence of bacteria adhering to or entering cells. Furthermore, the figure shows statistical analysis *p*-values with statistical differences between strains indicated near bars.

3.2.2 *A. butzleri* DEGs upon shift from Arcobacter Agar to DMEM

The *A. butzleri* transcriptome was analyzed after 2 h of incubation in DMEM at 37 °C (adaptation phase performed before *in vitro* cell model

assay), showing an increase of the general gene expression in the shift from Arcobacter agar. Differences in the number of DEGs (LogFC > 3 and < 3, *p*-value < 0.05, *FDR < 0.05*) were observed between strains. More specifically, strain LMG 11119 showed a general overexpression with 44.69% DEGs (all overexpressed) of the total annotated genes, while the DEGs number for strain LMG 10828^T and strain 31 were 7.68% (of which 97.75% overexpressed) and 10.77%(of which 98.71% overexpressed), respectively (Figure 3.2A). The differentially expressed genes belonged to different gene classes (Clusters of Orthologous Genes COG codes), most of which were included in the pathways related to translation, ribosomal structure and biogenesis, cell wall – membrane envelope biogenesis, coenzyme transport and metabolism, signal transduction mechanisms, amino acid transport, and metabolism and inorganic ion transport and metabolism (Figure 3.2B). Some DEGs were related to genes whose virulence function has already been hypothesized in functional genomic analyses [5,6] (Figure 3.2C; Supplementary Table 3.2).

A significant upregulation (LogFC > 3, *p*-value < 0.05, *FDR < 0.05*) of genes putatively encoding for LPS O-antigen biosynthesis and therefore potentially linked to host cell adhesion, was observed in strain LMG 11119. These genes (*lacA*, *cotSA*, *pgaB*, *sunS*, *rfaG*, *mdoB*, and O-antigen ligase gene) were upregulated jointly with *macA* and two gene copies of *macB* [6]. The proteins encoded by the latter two genes have been linked to the transmembrane transport of LPS and similar glycolipids, and to virulence and antibiotic resistance in Gram negative bacteria [16]. Moreover, other LPS related genes (*rmlA*, *rmlC*, *rmolD*, *gmhA1*, *rfaF* and *gmhB*) resulted differentially expressed with a lower overexpression (logFC between 1.5 and 3).

Genes currently considered virulence-related in *A. butzleri* linked to chemotaxis (*cheD*, *cheY*, *cheR*, *mcp4*), and to two-component regulatory systems (*phoP*, *phoQ*, *mprA* and *ttrS*) were overexpressed in strain LMG 11119. In addition, the genes *exbB* and *exbD*, adjacent in the genome, were overexpressed in LMG 11119; these genes are linked to TonB protein activity (virulence, iron transport) [17]. Finally, genes *mviN* (*murJ*), *tlyA* and *cadF* (*oprF*) showed to be overexpressed, and the respective proteins are widely considered *Arcobacter* virulence factors [5–7]. Overall, the gene upregulation observed for strain LMG 11119 is in accordance with previous observations regarding the virulence behavior of other Gram negative bacteria such as *Escherichia coli* and *Salmonella enterica*, showing an increase of the virulence gene expression when in contact with DMEM or Eagle's minimal essential media (MEM) [18,19]. Moreover, the passage from *Arcobacter* media to DMEM characterized by different carbon sources and conditions, in particular sugars and amino acids, pH, salts, jointly with the passage from a solid to a liquid medium may explain the bacterium's need to modulate gene expression in response to new environmental conditions [19–21]. We further hypothesize a role of these overexpressed genes by strain LMG 11119 on the subsequent first phase of contact of the bacterial cells with the human-derived cell layer. The fast response to changing environmental conditions of this strain that could also activate putative virulence genes, suggests an advantage also during infection and passage through the gut.

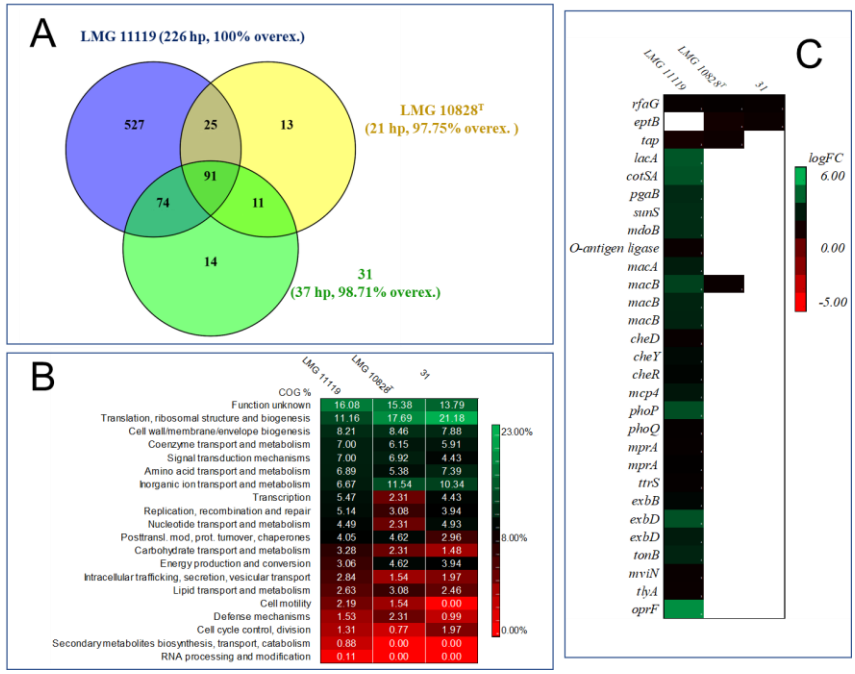


Figure 3.2. DEGs Venn diagram of *A. butzleri* strains DEGs comparison (A), DEGs pathway classes (B) and virulence-related DEGs heatmap (C) of *A. butzleri* incubated in DMEM after 2 h. The Venn diagram (**A**) shows the presence of DEGs shared between different strains and present in a single strain (hypothetical proteins excluded). The percentage of over-expressed genes (percentages calculated has been calculated on total DEGs including hypothetical proteins) are indicated in brackets next to the strain name. The heatmap **B** shows percentages relating to COG gene classes of total DEGs detected after 2 h of *A. butzleri* incubation in DMEM. The *A. butzleri* differentially expressed genes in DMEM after 2 h of incubation that based on the described genomic functions could play a role in *A. butzleri* virulence are shown in the heatmap **C**.

3.2.3 Expression of genes currently considered virulence associated

In all three *A. butzleri* strains genes differentially expressed after 30' and 90' of contact with the cell line were detected. Strain LMG 11119, characterized by higher colonization of human cell lines, showed a higher ratio of overexpressed DEGs (48/62 at 30', 128/161 at 90' from the

inoculum in cell models). On the other hand, this ratio was lower in strains LMG 10828^{T} (39/102 at 30', 78/146 at 90') and 31 (34/55 at 30', 65/166 at 90'), indicating a higher number of genes that are downregulated. Most of the genes were found to be differentially expressed in only one strain, consequently, the DEGs common for the three strains were not numerous (Supplementary table 3.3-4; Figure 3.3A). The majority of the differentially expressed genes (over and underexpressed) were linked to energy production and conversion (Figure 3.3B). Also, genes currently considered *A. butzleri* putative virulence genes were among the DEGs (Figure 3.3C).

More specifically, considering the ten putative virulence genes (*cadF, ciaB, cj1349, hecA, hecB, mviN, pldA, irgA, tlyA* and *iroE*) currently used in the detection of virulence traits in different *Arcobacteraceae* species, overexpression for some of them was observed [7,22]. Gene *oprF* (corresponding to *cadF*, adhesion to host) and *irgA* (iron – regulation functions) were overexpressed at 30' together with *iroE* (iron acquisition and infection maintenance) after 90' in strain LMG 10828^{T}. The *oprF* gene was overexpressed at 30' and 90' also in strain 31 but not in strain LMG 11119.

Regulatory genes related to bacterial virulence were differentially expressed. Gene *walR*, part of a two-component regulatory system linked to *Staphylococcus* virulence, and a gene encoding a histidine kinase (LOV-HK) associated to *Brucella* virulence, are adjacent in LMG 11119 genome and were overexpressed after 90' [23,24].

A transaldolase related gene was overexpressed at 90' in strain LMG 11119. This enzyme has been linked to the ability to colonize host mucus.

Overexpression of this gene in the strain characterized by higher colonization on cell models secreting mucus supports this thesis [12].

3.2.4 TonB-complex overexpression is linked to a higher colonization

The TonB system is composed of TonB, ExbB and ExbD proteins, linked to the inner bacterial membrane and is involved in iron transport and assimilation. For the TonB complex functionality, in addition to ExbB and ExbD a siderophore transport protein linked to the outer membrane (TonB), is necessary [25]. The genes *exbB, exbD* and *tonB* are adjacent in *A. butzleri* genome and are preceded by a promoter upstream of *exbB* gene (Figure 3.3C-D).

In strain LMG 11119 the overexpression of *tonB* and *exbB* genes was observed after 30' while *exbD* was not differentially expressed. At 90'all three genes were overexpressed in LMG 11119. Strain LMG 10828^{T} showed only the overexpression of *tonB* gene at 90', while *tonB* at 90' and *exbD* at 30' and 90' were underexpressed in strain 31. Additionally, other genes related to the TonB system were differentially expressed. The gene *fiu* (Catecholate siderophore receptor Fiu) was overexpressed at 30' and 90' in the strains LMG 11119 and LMG 10828^{T}. Gene *fiu* was linked to the active transport across the outer membrane of complexed iron thanks to the use of catecholate siderophores and is considered linked to TonB function [26]. Moreover, other genes encoding siderophores-related proteins linked to iron assimilation were overexpressed in the strain LMG 11119. These overexpressed genes were *pupB* (receptor for the siderophores ferric pseudobactin BN8), *fcuA* (ferrichrome receptor), *feoA* at 30' and 90', while *feoB* (putative Fe^{2+} uptake protein) was overexpressed at 90' [27]. The gene *fcuA* was under-expressed at both timings in the strain LMG 10828^{T}. Gene *cirA* (Colicin I receptor

precursor), linked to iron assimilation, was overexpressed at 30' and 90' in strain LMG 10828^T and at 90' in LMG 11119 [26,28].

These findings suggest an important role of genes belonging to the TonB protein complex. The physiological difference between strains was accompanied by the overexpression at the same time of TonB and iron assimilation related genes. This gene expression pattern is in accordance with previous studies of TonB in other Gram negative bacteria such as *E. coli* and *Pseudomonas plecoglossicida* where TonB complex mutants showed an attenuated virulence [29,30]. Furthermore, iron transport and uptake are considered important in several pathogenic bacteria including *Yersinia pestis* which produces siderophores as an iron uptake system, while the deletion of a Ferric uptake regulator (Fur) has been linked to the reduction of virulence of *Riemerella anatipestifer*, an avian pathogen [31]. Further, in mice, *Salmonella* responds to iron-depleting conditions in the host by iron-acquisition genes upregulation [32]. It is however correct to remember that a small amount of iron is present in DMEM in the form of ferric nitrate, in a concentration of 0.25 μM, this substance may have contributed, together with bacterial contact with host cells to the activation of iron-related genes [33]. The data regarding genes linked to iron uptake and the results obtained in other pathogenic bacteria in literature suggest an important role of iron transport and assimilation during *A. butzleri* virulence [29,30,34].

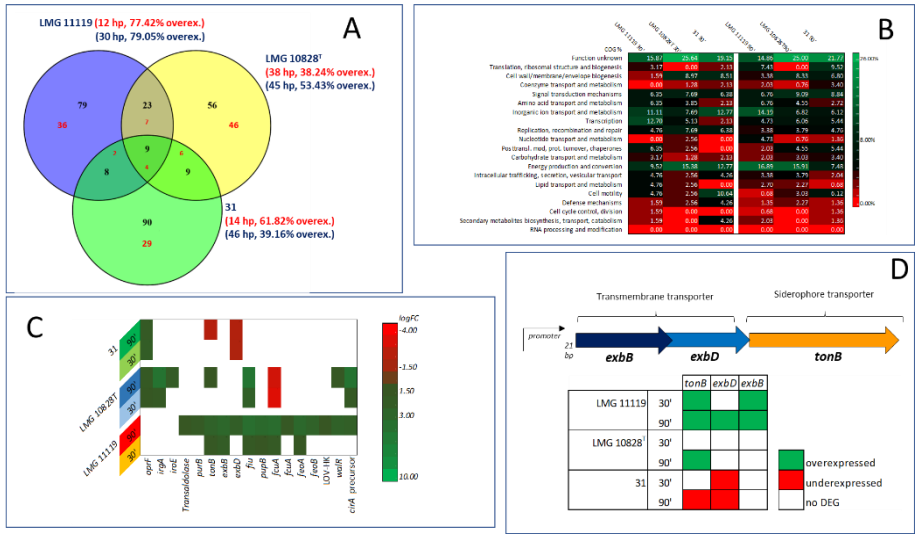


Figure 3.3. DEGs Venn diagram of *A. butzleri* strains DEGs comparison (A), DEGs pathway classes (B), (C) virulence-related DEGs heatmap of *A. butzleri* in contact with host cells and (D) *tonB* operon organization. The Venn diagram (**A**) shows the presence of DEGs shared between different strains and present in a single strain detected after 30' (in red) and 90' of *A. butzleri* contact with *in vitro* host cells (hypothetical proteins excluded), and the percentage of over-expressed genes (percentages calculated has been calculated on total DEGs, including hypothetical proteins) are indicated in brackets next to the strain name (red = 30'; blue = 90').
The heatmap **B** shows percentages relating to COG gene classes of total DEGs detected in virulence conditions at 30' and 90'. The heatmap **C** shows differentially expressed genes of *A. butzleri* after 30' and 90' of contact with host cells and considered *A. butzleri* putative virulence genes from functional gene annotation analysis. The organization of *tonB* operon is shown in figure **D**, the times shown in the table (30 'and 90') indicate the overexpression (green) or underexpression (red) of the indicated gene. The promoter detection with BPROM software indicates a putative promoter structure 21 bp upstream of *exbB* gene.

3.2.5 Differential gene expression of organic acids related genes

The ability of *A. butzleri* to grow by consuming lactate is known, while it is not considered able to grow in the presence of acetate alone, with the exception of strain ED-1 which can grow in presence of lactate, acetate and moderately in the presence of succinate. Therefore, certain heterogeneity at the level of strains regarding the metabolism of *A. butzleri* is outlined [35,36]. Human cells are also characterized by the presence in their metabolism of pathways linked to acetate and lactate [37]. An increase in lactate concentration has been linked to human cells stress response [38], however, human cells can also consume this organic acid [39]. Acetate is the main substrate that supports acetyl coenzyme A (Ac-CoA) metabolism in human cells, while its consumption under stress conditions has also been observed [37,40,41].

In the genome of *A. butzleri* genes linked to the metabolism of acetate and lactate are present and part of them were differentially expressed in colonization-invasion tests of the present work [35] (Figure 3.4A). The genes belonging to *yjcH-actP* operon that are present in different copies in the *A. butzleri* genome, were differentially expressed in the strains tested (Figure 3.4B). The gene *yjcH* (Inner membrane protein YjcH) involved in acetate dissimilation and transport and *actP* (Cation/acetate symporter ActP) linked to acetate transport in *E. coli*, were overexpressed at 30' and 90' in the three *A. butzleri* strains [42]. Gene *actP* showed a low overexpression at 30' in LMG 11119 (LogFC 1.49) which, however, showed the overexpression of two *actP* copies at 90 minutes. The gene *acs* (Acetyl-CoA synthetase) was differentially expressed only at 30' in strain LMG 11119 with a low overexpression (LogFC 1.37). ACS protein catalyzes the reaction which transforms acetate, coenzyme A and ATP into

AMP, pyrophosphate and Acetyl-CoA; this gene is present in the same operon with *actP* and *yjcH* in *E. coli* while in *A. butzleri* is present in a genomic region not adjacent to *actP* and *yjcH*. Furthermore, *yjcH* and *actP* were overexpressed in one gene copy at 30' and 90' in strains 31 and LMG 10828^{T}, while two gene copies of these genes were overexpressed at 90' in the strain LMG 11119 .

Based on the results obtained regarding the genes associated with acetate metabolism, it was decided to quantify acetate in the medium under the various conditions tested. The quantification results are shown in Figure 3.4C. The chemical analyses performed on DMEM obtained during *in vitro* test confirm what was observed by the expression of organic acid related genes (Figure 3.4A-B). Acetate in DMEM was not detectable after 90' for strain LMG 11119. At 30' a lower concentration was observed when compared to DMEM adaptation (without the presence of human cells) the same time (p-value < *0.02*). For the other two strains differences in the acetate concentration between 30' and 90' were not significant (p-value > *0.05*). The decrease of acetic acid concentration after 90' of infection only in strain LMG 11119 suggests a role of this organic acid in this strain that showed a larger number of organic acid-related DEGs. This result is similar to the observation in *E. coli*, where the operon *acs-yjcH-actP* has been linked to acetate assimilation system and in consequences to *E. coli* replication inside macrophages. Deletions of this operon in *E. coli* have been linked to virulence and colonization capability decrease in avian lungs [42]. Moreover, acetate consumption was linked with *E. coli* adherence promotion to Caco-2 cells [43]. The consumption of acetate as carbon source leads to speculate on a possible energetic advantage for *A. butzleri*.

The comparison between DMEM after 30' of bacteria - host cell contact and DMEM with host cell alone showed an increase of acetate after 30' of infection in LMG 11119 (18.54 times increase, *p*-value < *0.05*). The increase in acetate concentration after 30' of infection suggests a host stress response and an increase in acetate concentration during colonization and invasion of the cell line [41]. Acetate concentration was higher after 90' of LMG 10828^T when compared to LMG1119 (*p*-value < *0.01*). Strains 31, LMG 10828^T showed a lower acetic acid concentration after 30' of contact with host cell compared with bacteria adapted in DMEM (decrease of 81.00% and 76.91%; *p*-value < *0.03*). However, a comparison between *A. butzleri* alone in DMEM and normal DMEM showed a higher acetic acid concentration after 30' of bacteria adaptation, suggesting production of acetic acid by the three strains (Figure 3.4C). These aspects show the *A. butzleri* ability to produce acetic acid as well as to consume it as already illustrated in literature [36]. It is important to observe that in the *in vitro* assay of this study acetate consumption can be linked to acetate produced by host cells or normally present in DMEM fetal serum [44]. However, as stated, human cells are also capable to consume acetate [37] (Figure 3.4C).

An increase in lactic acid concentration was observed at 90' for LMG 11119 and LMG 10828^T compared with these *A. butzleri* strains adapted in DMEM (*p*-value < *0.03*) (Figure 3.4C). Lactic acid concentration after 90' was higher for strains LMG 11119 and LMG 10828^T when compared to the human cell line in DMEM as well (average increase of 310 times; *p*-value < *0.01*). Similarly, strain 31 showed an increase of lactic acid at 30' (an increase of 22.79 times; *p*-value < *0.01*) (Figure 3.4C). The increase in lactic acid concentration without related *A. butzleri* DEGs,

suggests the production of this molecule by host cell line. This aspect has been already observed in Caco-2 cell line as a response to *E. coli* infection [45]. Considering the bacteria in DMEM without host cells, our data showed a lactic acid concentration decrease after 90' of incubation, confirming the ability of *A. butzleri* to consume lactate from fetal serum present in DMEM (Figure 3.4C) (*p*-value < *0.01*) [36,46,47].

The analysis of glucose in DMEM incubated with only bacteria did not show wide concentration variation compared with non inoculated DMEM with the exception of a decrease after 90' for strain LMG 11119 (13.53% of initial glucose, *p*-value < *0.01*), while DMEM incubated with the strain 31 without host cells showed a minimum glucose decrease passing from control to DMEM after 90 'of incubation (*p*-value < *0.01*) (Supplementary figure 3.1). This aspect follows the low number of carbohydrate transport and metabolism DEGs detected in the three strains and different conditions (average 2.46%) (Figure 3.2B and 3.3B; Supplementary Table 3.5). The lack of glucose fermentation has already been observed in *A. butzleri* [36,47]. The differences in glucose consumption among *A. butzleri* strains here shown, is an aspect previously observed in *Campylobacter* species where a small part of strains can consume glucose [48].

The consumption related to pyruvic acid was more pronounced compared to glucose. A decrease in pyruvic acid concentration was observed passing from DMEM without bacteria to 90' bacterial adaptation. A decrease of pyruvic acid concentration of 99.16% and 21.43% passing from the inoculated DMEM at time 0 to 30' and 90' was observed (*p*-value < *0.001*) respectively in strains LMG 11119 and 31. The strain LMG 10828^{T} showed a significant decrease (50.01%) only between normal DMEM and DMEM after 90 'of incubation (*p-value* < *0.01*) without statistically

significant differences after 30’ of incubation (Supplementary figure 3.1). These results are in accordance with the already known ability of *A. butzleri* to consume pyruvic acid [47].

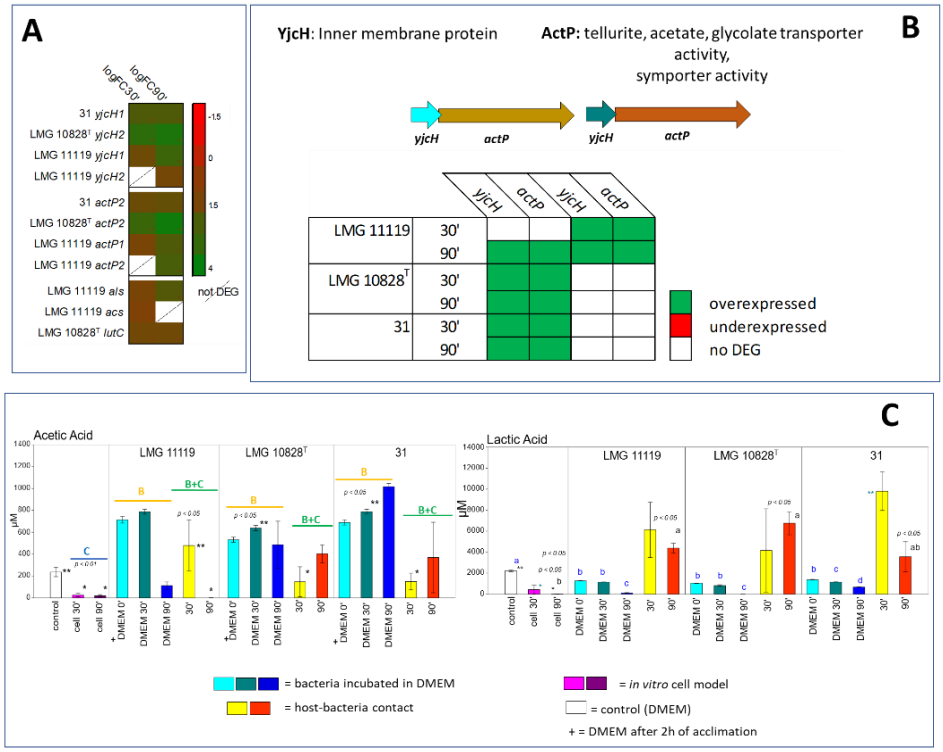


Figure 3.4. Acetic and lactic acid concentration and related DEGs. The heatmap (**A**) shows lactate and acetate-related DEGs of *A. butzleri* in contact with host cells (*p*-value < *0.05*, FDR corrected). The association of *actP* and *yjcH* gene copies is shown in figure **B**, the times shown (30 'and 90') indicate the relative overexpression (green) of the indicated gene. The other component of this operon *acs* gene present in *E. coli*, was found to be overexpressed in the LMG 11119 strain but is not present adjacent to *actP* and *yjcH*. The bar chart (**C**) shows concentrations in DMEM of acetic and lactic acid in µM. In the figure are indicated the strain codes and sampling times 0' (after 2h of adaptation), 30' and 90'. The different conditions are indicated by different colors (color codes below figure) while the lines above the bars of the graph show the analysis of DMEM from host cells (blue; C), bacteria (yellow; B) and bacteria in contact with the host cell (green; B+C). The error bars represent the standard errors, *p*-value with statistical differences between conditions are indicated near bars with different colors of the letters indicate different comparisons.

3.3. Conclusions

The present work led to an evaluation of the entire transcriptome of three *A. butzleri* strains that exhibited some physiological differences during *in vitro* human mucus producer gut model tests. The colonization ability of the strains, without high invasiveness, has been confirmed together with the involvement of some genes currently considered to be virulence-related by genomic analyzes. The strain LMG 11119 characterized by higher colonization showed a fast response to environmental stimuli in DMEM suggesting an important role of environment-related gene expression response that could also play a role in adaptation under host colonization conditions.

Relevant the presence of overexpressed genes linked to iron transport in the strain characterized by higher colonization that allows to hypothesize the importance of iron metabolism during *A. butzleri* infections. Other relevant DEGs were organic acid-related genes; as already observed in other Gram negative bacteria, these differentially expressed genes suggest the consumption of host derived organic acids. The expression profile of genes related to organic acids metabolism correlated well with the consumption pattern of the respective acids during colonization of the cell lines. Moreover, the analysis of organic acids also allowed the observation of host cells response to the stress condition caused by contact with *A. butzleri* linked to the release of lactate by human cells.

The data show the importance of functional annotation genomes analysis confirming some genes already considered virulence-related. *A. butzleri* gene expression data will provide more information on metabolic pathways active during virulence clarifying the mechanisms in place during pathogen-host contact. Future knockout studies of different genes

will deepen the knowledge about their functions. Even though there is a significant number of sequences with an unknown function, for certain genes an implication in *A. butzleri* virulence process can be suggested. Although this work was performed on simplified *in vitro* models in comparison to *in vivo* conditions, our data can be used for subsequent targeted *in vitro* and *in vivo* studies as well as suggesting new gene sequences for the detection and study of *A. butzleri*.

3.4. Material and Methods

3.4.1 *A. butzleri* cultivation and experiment bacteria suspension preparation

The *A. butzleri* strains, isolated from human in different geographic areas, LMG 10828^{T} (U.S.A), LMG 11119 (Italy) and 31 (Belgium) were obtained from the Belgian Coordinated Collection of Microorganisms (BCCM; Laboratory for Microbiology, Ghent University, Belgium) and cultivated in microaerophilic conditions at 30 °C on Arcobacter agar (CM0965, Oxoid, Basingstoke, UK) with C.A.T supplement (SR0174, Oxoid).

The bacteria pellets have been collected for each experiment according to the following method. A colony was inoculated in 5 ml of Arcobacter broth and incubated at 30 °C. After 48 h, 500 µl of culture were inoculated on Arcobacter agar plates supplemented with C.A.T. After 48 h of incubation in microaerobic conditions, the strains were collected with 1 ml of Ringer's solution (1.15525, Millipore, Burlington, Massachusetts, U.S.A). After two Ringer's solution washing (13000 rpm, 10') the approximate bacterial load of each working suspension was determined by OD at 630 nm evaluation with ELx880 microtiter plate reader (Savatec, Turin, Italy) using an internal standard curve.

3.4.2 *In vitro* human gut models production and cultivation

Two human colon carcinoma cell lines Caco-2 (86010202, ECACC, European Collection of Authenticated Cell Cultures, Public Health England) and HT29-MTX-E12 (12040401, ECACC) were cultured in Dulbecco's Modified Eagle's Medium (DMEM 6429; Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% of fetal bovine serum (FBS; F7524 Sigma-Aldrich) and EmbryoMax Penicillin-Streptomycin Solution (TMS-AB2-C, Sigma-Aldrich). The cell lines were grown in culture flasks (Corning, New York, New York, USA) at 37 °C in a humidified atmosphere (5% CO_2) sub-passaged every 3–4 days (Eppendorf, Galaxy 170 S, Hamburg, Germany). The mucus producer *in vitro* epithelial structure monolayer was prepared with Caco-2 and HT29-MTX-E12 cells in a 9:1 ratio [49]. Briefly, the cells were seeded at a density of 35.000 cells cm^{-2} on 21.2 cm^2 cell culture dishes to obtain sufficient material for RNA analysis and on 1.93 cm^2 models for *A. butzleri* colonization-invasion evaluation and grown in complete culture media under the same conditions described above for 14–15 days. At the reach of functional polarization, the models were considered differentiated and used for the experiments [50]. Three days before the *in vitro* test the models were washed twice with PBS 1X (Phosphate buffered saline) and the culture media used during cell maintenance was replaced with the same media without antibiotics to allow bacterial growth.

3.4.3 *In vitro* colonization - invasion assay

The *in vitro* colonization – invasion assay has been performed as previously described in a similar test about *A. butzleri* on mucus producer gut models with some modifications [6].

The three *A. butzleri* strains collected from Arcobacter Agar plates were diluted in DMEM media and incubated for 2 h of acclimation period (37°C, 5% CO_2) obtaining DMEM time 0 control for the transcriptome analysis. The *A. butzleri* strains in DMEM (T0), were inoculated on *in vitro* mucus producer human cells replacing ¾ of the model media, obtaining an average load of log 6.54 CFU cm^{-2} (st.err., 0.29) (Supplementary Table 3.1). The colonization-invasion evaluation was performed in parallel after 30' and 90' of bacterial infection in three biological replicates on two different model wells per sampling time. After 30' and 90' of bacteria contact with host cells at 37 °C (5% CO_2), the bacteria not adhering to the cells were removed by two PBS washing steps. The colonization capability (bacteria ability to adhere and enter in host cells, T1) has been evaluated by incubating one cell model well with 0.5 ml of 0.25% Triton X- 100 (v v^{-1}, in PBS). After 30' of incubation at 37°C, the resulting suspension was analyzed to evaluate the colonization bacterial load (CFU, colony-forming unit method) plating the dilutions in Ringer's solution on Arcobacter Agar media (C.A.T supplemented). The bacterial load count has been performed after 48 h of incubation at 30 °C in microaerobic conditions. The *A. butzleri* invasion capability (number of bacteria cells entered in the host cells, T2) has been evaluated in parallel adding 0.5 ml of DMEM (without FBS) containing 300 μg ml^{-1} of gentamicin sulfate (G1914, Sigma-Aldrich) to kill all the bacteria outside the host cells. After 2 h of gentamicin treatment at 37 °C, two PBS washing steps have been performed to eliminate antibiotic traces and the internalized *A. butzleri* cells load was evaluated following the CFU evaluation methods exposed above. The raw count data was expressed as log CFU cm^{-2}. The colonization Δ load has been calculated as T1 logCFU cm^{-2} – T0 logCFU

cm^{-2}, while internalized bacteria Δ load has been calculated as T2 logCFU cm^{-2} – T0 logCFU cm^{-2}.

During the *in vitro* tests, biological samples were taken for transcriptomic analysis in three biological replicates. These samples were taken under the same conditions as the tests previously exposed on cell models processed in parallel to those for the evaluation of colonization and invasion of *A. butzleri* strains. Briefly, after 30' and 90' of *A. butzleri* co-incubation with human cell models in 21.2 cm^2 cell dishes, the cell layer was washed twice with PBS 1X, keeping DMEM eliminated from models for chemical analysis. After this washing step the cell layer containing adhered and internalized bacteria has been collected with 1 ml of PBS 1X and centrifugate at 13000 rpm for 5' at 4 °C, once PBS has been eliminated 0.5 ml of RNAlater (AM7021, Invitrogen; Waltham, Massachusetts, USA) has been added to the cell-bacteria pellet and stored at -20 °C for subsequent analyzes. The RNAlater conservation procedure has been applicated also to 1 ml of *A. butzleri* DMEM suspension after adaptation as described above and to 1 ml of *A. butzleri* suspension collected from Arcobacter Agar before adaptation through centrifugation (13000 rpm for 5' at 4 °C) and used as controls.

3.4.4 Genome sequencing and bioinformatical analysis

The genomic DNA of the *A. butzleri* strains was extracted following a beads-beating, phenol-chloroform – isoamyl alcohol DNA extraction method. After nucleic acids extraction an RNAse A treatment through incubation at 37 °C for 30' was performed to digest RNA (5 μg μl^{-1}, MRNA092 Epicenter, Madison, Wisconsin, U.S.A).

A first DNA quantification was performed using Nanodrop (ND1000, Thermo Fisher Scientific), followed by an electrophoretic run (100 V, 30′)

on agarose gel 0.8% (w v^{-1}, 0710, VWR) in TAE 1X (Tris – Acetic acid – EDTA, K915, VWR) with gelRed as DNA intercalating (41003, Biotium, Fremont, California, USA), this electrophoretic run was performed as DNA quality check. Illumina Novaseq 6000 whole-genome sequencing (paired-end 150 bp, coverage 100X, performed by Novogene company, Cambridge, UK) was performed after Qubit 2.0 quantification. 1 μg of gDNA was used for the library preparation using (NEB, Ipswich, MA, USA), randomly fragmented by shearing (350 pb) after this fragmentation the samples were polished. The fragments were A-tailed and ligated with NEBNext adapter, and PCR enriched (P5 and indexed P7 oligos), while the purification of PCR product was performed with AMPure XP system. The libraries were analyzed by Agilent 2100 Bioanalyzer (size distribution evaluation) and quantified with real-time PCR.

3.4.5 RNA sequencing

The RNA sequencing (RNAseq) was performed by GENEWIZ, LLC. (South Plainfield, NJ, USA) on RNA extracted from collected *in vitro* test samples as described in the dedicated section. The total RNA was extracted through the use of Qiagen RNeasy Plus Universal mini kit following the manufacturer's instructions (73404, Qiagen, Hilden, Germany). The samples were quantified with Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and checked for RNA integrity with 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). The rRNA depletion was performed with FastSelect rRNA H/M/R Kit (334385, Qiagen, Hilden, Germany), followed by the sequencing libraries preparation with NEBNext Ultra RNA Library Prep Kit for Illumina following the manufacturer's instructions (E7770, NEB, Ipswich, MA, USA). The enriched RNA was fragmented for 15' at 94 °C, first and

second-strand cDNA were synthesized after this step of fragmentation and once obtained were end-repaired and adenylated to 3' ends. After adenylation the universal adapter was ligated to cDNA fragments followed by index addition and PCR library enrichment. The libraries obtained were validated with the employ of Agilent Tapestation 4200 (Agilent Technologies, Palo Alto, CA, USA), the libraries quantification was performed with Qubit 2.0 Fluorometer and quantitative PCR (Applied Biosystems, Carlsbad, CA, USA). The sequencing libraries were multiplexed and clustered on the flow-cell and intended for sequencing with Illumina NovaSeq 6000, ~20M read pairs sample, ~40M read pairs/sample in case of human cells-host co-incubation (2x150 Pair-End). The image analysis and base calling were performed with the use of the NovaSeq Control Software v1.6, raw sequences data (.bcl) from Novaseq instrument was converted in fastq files demultiplexed with Illumina bcl2fastq program v2.20 (one mismatch allowed for index sequence identification).

3.4.6 Genome and Transcriptome bioinformatic and statistical analysis

The whole-genome sequencing reads produced were quality filtered with Solexa QA software, PRINSEQ++ v1.2 has been used to remove sequences shorter than 60 bp together with dereplicated sequences [51,52]. The reads were *de novo* assembled at our laboratory with SPAdes (v3.11.0) the contigs obtained were evaluated for their quality with QUAST (v5.0.0) and annotated using Prokka tool (v1.11) [53–55] (Supplementary table 3.6). The promoters detection has been performed with BPROM tool [56]. The graphical visualization of the genomes has been performed with Anvi'o v7.1 (mcl 10) [57]. Gene enrichment analysis was performed with

emapper v2.1.6 tool while SWISS-MODEL has been used to detect particular sequences in DEGs of interest [58,59]. The operons visualization was performed with Artemis v18.1.0 [60].

The analysis of *A. butzleri* transcriptome has been evaluated as follows. Raw reads were quality checked with QUAST (v5.0.0) adapters were removed with cutadapt v3.1 [54,61]. Raw reads were quality filtered by SolexaQA++ v3.1.7.1 and PRINSEQ v0.20.4, respectively (Phred score < 20, < 51bp) [51,52]. Reads that were human contaminants have been discarded by using by using Bowtie 2 v2.3.5 in end-to-end sensitive mode [62]. Clean reads were then aligned against the respective annotated strain used by Bowtie2 in end-to-end, sensitive mode. The number of reads mapped to each gene were extracted with samtools v1.2. The raw count transcripts table was normalized with RPKM method taking into consideration the experimental conditions and the characteristics of the obtained transcriptomic libraries [63]. The normalized transcripts count was elaborated with EdgeR Bioconductor R package v3.14 to evaluate the presence of possible DEGs [64,65]. Genes characterized by logFC (*p*-value < *0.05*, FDR < 0.05) equal or greater than 1.5 and equal or lower than -1.5 were considered DEGs, genes of interest with values different than those indicated have been cited in the text by specifying the logFC value.

3.4.7 High Performance Liquid Chromatography

The DMEM obtained from *A. butzleri* and cell models incubation and co-incubation was centrifugate at 13000 rpm, 10’ at 4°C, the supernatant was sterile filtered (0.2 µm) and chemically analyzed by HPLC (High Performance Liquid Chromatography).

The HPLC system (Thermo Electron Corporation, Waltham, MA, USA) was equipped with a SCM 1000 degasser, a P2000 binary gradient pump, a multiple autoinjector (AS3000), a photodiode array (PDA) detector (Thermo Electron Corporation, UV6000LP) and a refractive index detector (RI-150). The mobile phase was 0.013 N H_2SO_4 and the sample injection volume was 20 μL. The isocratic elution method was applied with a flow rate of 0.6 ml min^{-1} for 30'. The analysis has been performed using a reverse-phase Aminex HPX-87H column (300mm* 7.8mm) equipped with a Microguard cartridge (Bio-Rad Laboratories, Hercules, CA, USA) working at 65°C. The data have been elaborated with ChromQuest chromatography data system (ThermoQuest software 5.0, Inc., San Jose, CA, USA), while identification was achieved by comparison with the retention time of authentic standards from Sigma-Aldrich (Milan, Italy).

3.4.8 General Statistical analysis

The statistical analysis has been performed with the software RStudio v2021.09.0 (R 3.6.1,https://www.r-project.org/). Shapiro-Wilk's W and Modified Levene's tests (Brown-Forsythe) have been used to check respectively normality and homogeneity of the data. Kruskal–Wallis (K-W) and Anova test were used to evaluate overall differences and variations between multiple groups, while Wilcoxon Rank sum test (WRS) and two-sample t-test have been used to evaluate differences between 2 groups. These tests were used for nonparametric (K-W, WRS) and parametric data (Anova, t-test). Dunn's and Tukey's tests were used as *post hoc* analyses for nonparametric and parametric data respectively. The bar chart graphs (average values, standard errors) have been produced using Past3 v3.17 [66].

3.5. Availability of data and material

Whole Genome Raw sequence reads were deposited at the Sequence Read Archive of the National Center for Biotechnology Information to accession numbers SRX9057116 (LMG 10828^{T}), SRX9057105 (LMG 11119) and SRX9057128 (31). The RNAseq raw data are available to the bioproject number PRJNA703833 (3 = LMG 10828^{T}, 2 = LMG 11119 and 31).

The assembled sequences of genomes, functional annotation files, and logFC table are available on Zenodo to 10.5281/zenodo.5882246 (https://zenodo.org/).

3.6 References

[1] A. Pérez-Cataluña, N. Salas-Massó, A.L. Diéguez, S. Balboa, A. Lema, J.L. Romalde, M.J. Figueras, Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos, Front. Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.02077.

[2] D. Chieffi, F. Fanelli, V. Fusco, *Arcobacter butzleri* : Up-to-date taxonomy, ecology, and pathogenicity of an emerging pathogen, Compr. Rev. Food Sci. Food Saf. (2020) 1541-4337.12577. https://doi.org/10.1111/1541-4337.12577.

[3] T.P. Ramees, K. Dhama, K. Karthik, R.S. Rathore, A. Kumar, M. Saminathan, R. Tiwari, Y.S. Malik, R.K. Singh, *Arcobacter* : an emerging food-borne zoonotic pathogen, its public health concerns and advances in diagnosis and control – a comprehensive review, Vet. Q. 37 (2017) 136–161. https://doi.org/10.1080/01652176.2017.1323355.

[4] F. Fanelli, A. Di Pinto, A. Mottola, G. Mule, D. Chieffi, F. Baruzzi, G. Tantillo, V. Fusco, Genomic Characterization of *Arcobacter butzleri* Isolated From Shellfish: Novel Insight Into Antibiotic Resistance and Virulence Determinants, Front. Microbiol. 10 (2019) 1–17. https://doi.org/10.3389/fmicb.2019.00670.

[5] 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. https://doi.org/10.1016/j.meegid.2020.104213.

[6] D. Buzzanca, C. Botta, I. Ferrocino, V. Alessandria, K. Houf, K. Rantsiou, Functional pangenome analysis reveals high virulence plasticity of *Aliarcobacter butzleri* and affinity to human mucus, Genomics. 113 (2021) 2065–2076. https://doi.org/10.1016/j.ygeno.2021.05.001.

[7] L. Douidah, L. de Zutter, J. Bare, P. De Vos, P. Vandamme, O. Vandenberg, A.-M. Van den Abeele, K. Houf, Occurrence of Putative Virulence Genes in *Arcobacter* Species Isolated from Humans and Animals, J. Clin. Microbiol. 50 (2012) 735–741. https://doi.org/10.1128/JCM.05872-11.

[8] C. Girbau, C. Guerra, I. Martínez-Malaxetxebarria, R. Alonso, A. Fernández-Astorga, Prevalence of ten putative virulence genes in the emerging foodborne pathogen *Arcobacter* isolated from food products, Food Microbiol. 52 (2015) 146–149. https://doi.org/10.1016/j.fm.2015.07.015.

[9] V. Lievin-Le Moal, A.L. Servin, Pathogenesis of Human Enterovirulent Bacteria: Lessons from Cultured, Fully Differentiated Human Colon Cancer Cell Lines, Microbiol. Mol. Biol. Rev. 77 (2013) 380–439.

https://doi.org/10.1128/mmbr.00064-12.

[10] T. Langerholc, P.A. Maragkoudakis, J. Wollgast, L. Gradisnik, A. Cencic, Novel and established intestinal cell line models e An indispensable tool in food science and nutrition, Trends Food Sci. Technol. 22 (2011) S11–S20. https://doi.org/10.1016/j.tifs.2011.03.010.

[11] 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.

[12] J.-F. Sicard, G. Le Bihan, P. Vogeleer, M. Jacques, J. Harel, Interactions of Intestinal Bacteria with Components of the Intestinal Mucus, Front. Cell. Infect. Microbiol. 7 (2017). https://doi.org/10.3389/fcimb.2017.00387.

[13] J.A. Naughton, K. Mariño, B. Dolan, C. Reid, R. Gough, M.E. Gallagher, M. Kilcoyne, J.Q. Gerlach, L. Joshi, P. Rudd, S. Carrington, B. Bourke, M. Clyne, Divergent Mechanisms of Interaction of *Helicobacter pylori* and *Campylobacter jejuni* with Mucus and Mucins, Infect. Immun. 81 (2013) 2838–2850. https://doi.org/10.1128/IAI.00415-13.

[14] B. Dolan, J. Naughton, N. Tegtmeyer, F.E.B. May, M. Clyne, The Interaction of *Helicobacter pylori* with the Adherent Mucus Gel Layer Secreted by Polarized HT29-MTX-E12 Cells, PLoS One. 7 (2012) e47300. https://doi.org/10.1371/journal.pone.0047300.

[15] A. Alemka, M. Clyne, F. Shanahan, T. Tompkins, N. Corcionivoschi, B. Bourke, Probiotic Colonization of the Adherent Mucus Layer of HT29MTXE12 Cells Attenuates *Campylobacter jejuni* Virulence Properties, Infect. Immun. 78 (2010) 2812–2822. https://doi.org/10.1128/IAI.01249-09.

[16] N.P. Greene, E. Kaplan, A. Crow, V. Koronakis, Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective, Front. Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.00950.

[17] 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/10.1186/1756-0500-5-327.

[18] L.Q. Leverton, J.B. Kaper, Temporal Expression of Enteropathogenic *Escherichia coli* Virulence Genes in an In Vitro Model of Infection, Infect. Immun. 73 (2005) 1034–1043. https://doi.org/10.1128/IAI.73.2.1034-1043.2005.

[19] J.M.A. Blair, G.E. Richmond, A.M. Bailey, A. Ivens, L.J. V. Piddock, Choice of Bacterial Growth Medium Alters the Transcriptome and Phenotype of *Salmonella enterica* Serovar Typhimurium, PLoS One. 8

(2013) e63912. https://doi.org/10.1371/journal.pone.0063912.

[20] M.M. Kendall, C.C. Gruber, C.T. Parker, V. Sperandio, Ethanolamine Controls Expression of Genes Encoding Components Involved in Interkingdom Signaling and Virulence in Enterohemorrhagic *Escherichia coli* O157:H7, MBio. 3 (2012). https://doi.org/10.1128/mBio.00050-12.

[21] E.-B. Goh, G. Yim, W. Tsui, J. McClure, M.G. Surette, J. Davies, Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics, Proc. Natl. Acad. Sci. 99 (2002) 17025–17030. https://doi.org/10.1073/pnas.252607699.

[22] P. Kietsiri, C. Muangnapoh, W. Lurchachaiwong, P. Lertsethtakarn, L. Bodhidatta, O. Suthienkul, N.C. Waters, S.T. Demons, B.A. Vesely, Characterization of *Arcobacter* spp. Isolated from human diarrheal, non-diarrheal and food samples in Thailand, PLoS One. 16 (2021) e0246598. https://doi.org/10.1371/journal.pone.0246598.

[23] A. Delauné, S. Dubrac, C. Blanchet, O. Poupel, U. Mäder, A. Hiron, A. Leduc, C. Fitting, P. Nicolas, J.-M. Cavaillon, M. Adib-Conquy, T. Msadek, The WalKR System Controls Major Staphylococcal Virulence Genes and Is Involved in Triggering the Host Inflammatory Response, Infect. Immun. 80 (2012) 3438–3453. https://doi.org/10.1128/IAI.00195-12.

[24] J. Rinaldi, M. Arrar, G. Sycz, M.L. Cerutti, P.M. Berguer, G. Paris, D.A. Estrín, M.A. Martí, S. Klinke, F.A. Goldbaum, Structural Insights into the HWE Histidine Kinase Family: The *Brucella* Blue Light-Activated Histidine Kinase Domain, J. Mol. Biol. 428 (2016) 1165–1179. https://doi.org/10.1016/j.jmb.2016.01.026.

[25] G. Iraola, R. Pérez, H. Naya, F. Paolicchi, E. Pastor, S. Valenzuela, L. Calleros, A. Velilla, M. Hernández, C. Morsella, Genomic Evidence for the Emergence and Evolution of Pathogenicity and Niche Preferences in the Genus *Campylobacter*, Genome Biol. Evol. 6 (2014) 2392–2405. https://doi.org/10.1093/gbe/evu195.

[26] N. Noinaj, M. Guillier, T.J. Barnard,, S.K. Buchanan, TonB-Dependent Transporters: Regulation, Structure, and Function, Annu. Rev. Microbiol. 64 (2010) 43–60. https://doi.org/10.1146/annurev.micro.112408.134247.

[27] C.K.Y. Lau, K.D. Krewulak, H.J. Vogel, Bacterial ferrous iron transport: the Feo system, FEMS Microbiol. Rev. 40 (2016) 273–298. https://doi.org/10.1093/femsre/fuv049.

[28] Q. Wang, J.G. Frye, M. McClelland, R.M. Harshey, Gene expression patterns during swarming in *Salmonella typhimurium*: Genes specific to surface growth and putative new motility and pathogenicity genes, Mol. Microbiol. 52 (2004) 169–187. https://doi.org/10.1111/j.1365-2958.2003.03977.x.

[29] L. Hu, L. Zhao, Z. Zhuang, X. Wang, Q. Fu, H. Huang, L. Lin, L. Huang,

Y. Qin, J. Zhang, Q. Yan, The Effect of tonB Gene on the Virulence of *Pseudomonas plecoglossicida* and the Immune Response of *Epinephelus coioides*, Front. Microbiol. 12 (2021) 1–13. https://doi.org/10.3389/fmicb.2021.720967.

[30] A.G. Torres, P. Redford, R.A. Welch, S.M. Payne, TonB-Dependent Systems of Uropathogenic *Escherichia coli*: Aerobactin and Heme Transport and TonB Are Required for Virulence in the Mouse, Infect. Immun. 69 (2001) 6179–6185. https://doi.org/10.1128/IAI.69.10.6179-6185.2001.

[31] M. Huang, M. Liu, J. Liu, D. Zhu, Q. Tang, R. Jia, S. Chen, X. Zhao, Q. Yang, Y. Wu, S. Zhang, J. Huang, X. Ou, S. Mao, Q. Gao, D. Sun, M. Wang, A. Cheng, Functional characterization of Fur in iron metabolism, oxidative stress resistance and virulence of *Riemerella anatipestifer*, Vet. Res. 52 (2021) 48. https://doi.org/10.1186/s13567-021-00919-9.

[32] S.M. Zughaier, P. Cornelis, Editorial: Role of Iron in Bacterial Pathogenesis, Front. Cell. Infect. Microbiol. 8 (2018) 3880–3891. https://doi.org/10.3389/fcimb.2018.00344.

[33] T. Ackermann, S. Tardito, Cell Culture Medium Formulation and Its Implications in Cancer Metabolism, Trends in Cancer. 5 (2019) 329–332. https://doi.org/10.1016/j.trecan.2019.05.004.

[34] Z. Tan, S.M. Chekabab, H. Yu, X. Yin, M.S. Diarra, C. Yang, J. Gong, Growth and Virulence of *Salmonella* Typhimurium Mutants Deficient in Iron Uptake, ACS Omega. 4 (2019) 13218–13230. https://doi.org/10.1021/acsomega.9b01367.

[35] W.G. Miller, C.T. Parker, M. Rubenfield, G.L. Mendz, M.M.S.M. Wösten, D.W. Ussery, J.F. Stolz, T.T. Binnewies, P.F. Hallin, G. Wang, J.A. Malek, A. Rogosin, L.H. Stanker, R.E. Mandrell, The complete genome sequence and analysis of the epsilonproteobacterium *Arcobacter butzleri*, PLoS One. 2 (2007) e1358. https://doi.org/10.1371/journal.pone.0001358.

[36] L. Szydlowski, T.C.T. Lan, N. Shibata, I. Goryanin, Metabolic engineering of a novel strain of electrogenic bacterium *Arcobacter butzleri* to create a platform for single analyte detection using a microbial fuel cell, Enzyme Microb. Technol. 139 (2020) 109564. https://doi.org/10.1016/j.enzmictec.2020.109564.

[37] X. Liu, D.E. Cooper, A.A. Cluntun, M.O. Warmoes, S. Zhao, M.A. Reid, J. Liu, P.J. Lund, M. Lopes, B.A. Garcia, K.E. Wellen, D.G. Kirsch, J.W. Locasale, Acetate Production from Glucose and Coupling to Mitochondrial Metabolism in Mammals, Cell. 175 (2018) 502-513.e13. https://doi.org/10.1016/j.cell.2018.08.040.

[38] J.M. Ratter, H.M.M. Rooijackers, G.J. Hooiveld, A.G.M. Hijmans, B.E. de Galan, C.J. Tack, R. Stienstra, *In vitro* and *in vivo* Effects of Lactate

on Metabolism and Cytokine Production of Human Primary PBMCs and Monocytes, Front. Immunol. 9 (2018) 2564. https://doi.org/10.3389/fimmu.2018.02564.

[39] G.A. Brooks, Lactate as a fulcrum of metabolism, Redox Biol. 35 (2020) 101454. https://doi.org/10.1016/j.redox.2020.101454.

[40] Z.T. Schug, B. Peck, D.T. Jones, Q. Zhang, S. Grosskurth, I.S. Alam, L.M. Goodwin, E. Smethurst, S. Mason, K. Blyth, L. McGarry, D. James, E. Shanks, G. Kalna, R.E. Saunders, M. Jiang, M. Howell, F. Lassailly, M.Z. Thin, B. Spencer-Dene, G. Stamp, N.J.F. van den Broek, G. Mackay, V. Bulusu, J.J. Kamphorst, S. Tardito, D. Strachan, A.L. Harris, E.O. Aboagye, S.E. Critchlow, M.J.O. Wakelam, A. Schulze, E. Gottlieb, Acetyl-CoA Synthetase 2 Promotes Acetate Utilization and Maintains Cancer Cell Growth under Metabolic Stress, Cancer Cell. 27 (2015) 57–71. https://doi.org/10.1016/j.ccell.2014.12.002.

[41] S. Bose, V. Ramesh, J.W. Locasale, Acetate Metabolism in Physiology, Cancer, and Beyond, Trends Cell Biol. 29 (2019) 695–703. https://doi.org/10.1016/j.tcb.2019.05.005.

[42] X. Zhuge, Y. Sun, M. Jiang, J. Wang, F. Tang, F. Xue, J. Ren, W. Zhu, J. Dai, Acetate metabolic requirement of avian pathogenic *Escherichia coli* promotes its intracellular proliferation within macrophage, Vet. Res. 50 (2019) 31. https://doi.org/10.1186/s13567-019-0650-2.

[43] F. Yang, L. Yang, Z. Chang, L. Chang, B. Yang, Regulation of virulence and motility by acetate in enteropathogenic *Escherichia coli*, Int. J. Med. Microbiol. 308 (2018) 840–847. https://doi.org/10.1016/j.ijmm.2018.07.010.

[44] A.M. Hosios, M.G. Vander Heiden, Acetate metabolism in cancer cells, Cancer Metab. 2 (2014) 27. https://doi.org/10.1186/s40170-014-0027-y.

[45] X. He, D.O. Mishchuk, J. Shah, B.C. Weimer, C.M. Slupsky, Cross-talk between *E. coli* strains and a human colorectal adenocarcinoma-derived cell line, Sci. Rep. 3 (2013) 3416. https://doi.org/10.1038/srep03416.

[46] B. Hiebl, S. Peters, O. Gemeinhardt, S.M. Niehues, F. Jung, Impact of serum in cell culture media on in vitro lactate dehydrogenase (LDH) release determination, J. Cell. Biotechnol. 3 (2017) 9–13. https://doi.org/10.3233/JCB-179002.

[47] 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.

[48] J.M. Garber, H. Nothaft, B. Pluvinage, M. Stahl, X. Bian, S. Porfirio, A. Enriquez, J. Butcher, H. Huang, J. Glushka, E. Line, J.A. Gerlt, P. Azadi, A. Stintzi, A.B. Boraston, C.M. Szymanski, The gastrointestinal pathogen *Campylobacter jejuni* metabolizes sugars with potential help from

commensal Bacteroides vulgatus, Commun. Biol. 3 (2020) 2. https://doi.org/10.1038/s42003-019-0727-5.

[49] P. Hoffmann, M. Burmester, M. Langeheine, R. Brehm, M.T. Empl, B. Seeger, G. Breves, Caco-2/HT29-MTX co-cultured cells as a model for studying physiological properties and toxin-induced effects on intestinal cells, PLoS One. 16 (2021) e0257824. https://doi.org/10.1371/journal.pone.0257824.

[50] M. Pinto, S. Robine Leon, M.D. Appay, Enterocyte-like differentiation and polarization of the human colon carcinoma cell line Caco-2 in culture, Biol. Cell. 47 (1983) 323–330.

[51] M.P. Cox, D.A. Peterson, P.J. Biggs, SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data, BMC Bioinformatics. 11 (2010) 485. https://doi.org/10.1186/1471-2105-11-485.

[52] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, Bioinformatics. 27 (2011) 863–864. https://doi.org/10.1093/bioinformatics/btr026.

[53] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. 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. https://doi.org/10.1089/cmb.2012.0021.

[54] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics. 29 (2013) 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

[55] T. Seemann, Prokka: Rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

[56] V. Solovyev, A. Salamov, I. Seledtsov, D. Vorobyev, A. Bachinsky, AUTOMATIC ANNOTATION OF BACTERIAL COMMUNITY SEQUENCES AND APPLICATION TO INFECTIONS DIAGNOSTIC, in: Proc. Int. Conf. Bioinforma. Model. Methods Algorithms, SciTePress - Science and and Technology Publications, 2011: pp. 346–353. https://doi.org/10.5220/0003333703460353.

[57] A.M. Eren, Ö.C. Esen, C. Quince, J.H. Vineis, H.G. Morrison, M.L. Sogin, T.O. Delmont, Anvi'o: an advanced analysis and visualization platform for 'omics data, PeerJ. 3 (2015) e1319. https://doi.org/10.7717/peerj.1319.

[58] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and

phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (2019) D309–D314. https://doi.org/10.1093/nar/gky1085.

[59] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (2018) W296–W303. https://doi.org/10.1093/nar/gky427.

[60] T. Carver, S.R. Harris, M. Berriman, J. Parkhill, J.A. McQuillan, Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data, Bioinformatics. 28 (2012) 464–469. https://doi.org/10.1093/bioinformatics/btr703.

[61] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.Journal. 17 (2011) 10. https://doi.org/10.14806/ej.17.1.200.

[62] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods. 9 (2012) 357–359. https://doi.org/10.1038/nmeth.1923.

[63] S. Zhao, Z. Ye, R. Stanton, Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols, RNA. 26 (2020) 903–909. https://doi.org/10.1261/rna.074922.120.

[64] Y. Chen, A.T.L. Lun, G.K. Smyth, From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline, F1000Research. 5 (2016) 1438. https://doi.org/10.12688/f1000research.8987.2.

[65] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics. 26 (2010) 139–140. https://doi.org/10.1093/bioinformatics/btp616.

[66] and P.D.R. Øyvind Hammer, David A.T. Harper, PAST: paleontological statistics software package for education and data analysis, Palaeontol. Electron. 4 (2001) 1352–1357.

4. Arcobacteraceae pangenome analysis demonstrates genomes heterogeneity and reduction in genome size of species isolated from animals and humans

4.1. Introduction

The recently proposed family *Arcobacteraceae* includes Gram-negative bacterial species isolated from different environmental niches and hosts [1,2]. Species belonging to this family have recently been separated from the family *Campylobacteraceae* due to their genomic differences observed in the past years. The different sources of isolation and genomic heterogeneity of species within *Arcobacteraceae* has even led to the proposal of six bacterial genera [1]. Genomic features evaluation of species belonging to *Arcobacteraceae* has been characterized by an increased amount of information in the last years but are predominantly focused on the pathogenicity and general phylogeny, often leading to conflicting conclusions between authors [1,3]. In particular the division of the original genus *Arcobacter* into six genera: *Aliarcobacter, Halarcobacter, Malaciobacter, Pseudarcobacter, Poseidonibacter and Arcobacter,* is still heavily discussed [1,3,4]. Recently, some authors have even proposed to re-group these bacterial genera into one genus "*Arcobacter*" as the division does not reflect significant biological or clinical features. On and colleagues came to this conclusion after the analysis of average aminoacidic identity, BLAST-based average nucleotide identity, percentage of conserved proteins, alignment fractions, G-C percentages, *in silico* DNA–DNA hybridization values and genome-wide average nucleotide identity [3,4]. For this reason, the term "*Arcobacter*" will be used in this study to indicate the genus of the species belonging to the *Arcobacteraceae* family.

Part of these species has been exclusively isolated from environmental matrices such as water, while species such as *Arcobacter bivalviorum*, *Arcobacter mytili*, *Arcobacter molluscorum* and *Arcobacter venerupis*

have been collected from seafood such as mussels and shellfish [2,5–7]. Some species, in particular those belonging to the proposed genus *Aliarcobacter,* are present in farm animals in which they were initially associated with illness causing diarrhea, abortions and mastitis, although currently considered as none pathogenic for animals [2,8]. However, the species *Arcobacter butzleri, Arcobacter cryaerophilus, Arcobacter thereius*, and *Arcobacter skirrowii* have been isolated from diarrheic human stool samples and cases of septicemia, with food from animal origin and drinking water as suggested sources of infection [9,10]. Furthermore, *A. butzleri*, *Arcobacter cibarius*, *A. cryaerophilus*, *Arcobacter skirrowii*, *Arcobacter thereius*, and *Arcobacter trophiarum* are rather commonly present in livestock, causing no reduced production parameters, and can also be present at levels up to 10 colony forming units per gram on meat and meat products [2,11]. *Arcobacter cryaerophilus* and *A. butzleri* have also been isolated from other matrices, including raw milk and drinking water, and are the species most commonly isolated and studied so far [2,12,13].

At present comparative genomics of *Arcobacteraceae* has been performed on only a few species so far, arguing the necessity of a broad evaluation of this bacterial family [14–16]. Therefore, the aim of the present study is a pangenome analysis of the members of the *Arcobacteraceae* family. As at present only limited or even no additional isolates, other than the type strains, are available for the majority of the *Arcobacteraceae* species, the study is performed on the type strains. Furthermore, this approach allows more reliable assessment of the *Arcobacteraceae* pangenome, as inclusion of multiple isolate genomes available from only part of the species can lead to biased data on the different pangenome partitions (e.g., increase of

number of the soft-core genes). The focus is on the analysis of the species belonging to the proposed group *Aliarcobacter* due to their human and veterinary clinical and foodborne pathogenic relevance.

4.2. Results and discussion

4.2.1. *Arcobacteraceae* genomes analyses suggest the presence of a single genus

General information regarding the 20 *Arcobacteraceae* type strains, as well as on the extra *A. cryaerophilus* strain to cover the heterogeneity of this species [17], are shown in table 1 and figure 1. These type strains represent the groups 1 to 5 as proposed by Pérez-Cataluña *et al.* [1]. *A. vandammei* was placed in group 1 considering its characteristics and its phylogenetic position (Figure 2) [18]. Though several species have been isolated from multiple sources and various hosts, the initial isolation sources recorded for each type strain were considered. Species that were isolated from human clinical cases were considered in the "clinical" group, even if the reference strain was initially isolated from a different host or environmental matrix (Figure 1).

The *Arcobacteraceae* genomes come from the assembly of Illumina raw sequences reads (Supplementary Table 1). This strategy was followed in order to reduce differences related to the level of genomes completeness and error rates deriving from the sequencing technologies [19,20]. Moreover, the use of raw reads allowed the assemblies from the same assembler to be obtained [21]. The genomes "normalization" (raw sequence reads from Illumina, use of the same assembler) has therefore made possible the comparison of bacteria, reducing technical differences. It has to be considered that the use of draft genomes does not allow to obtain all the information that can derive from complete genomes.

However, the objectives of this study were linked to the detection of differences between groups considered phylogenetically different and in this contest the use of draft genome is common and generally accepted [1,22–24]. As a matter of fact, the draft genomes have already been used to detect differences between strains belonging to *Arcobacteraceae* species [14–16,25].

The genome comparison shows a generally smaller genome size of the species of group 1 (2.12 Mb, st. dev ± 0.2 Mb, *p-value < 0.05*) compared with members of the other 4 groups, without a significant correlation with the GC percentages. Moreover, the genes number of group 1 is in accordance with its small genomes size (average 2169 genes, st. dev ± 156, *p-value < 0.05*) (Supplementary Table 1). Group 1 is composed of strains isolated from animals and some of them have already been associated with human infections. In particular, *A. butzleri*, *A. cryaerophilus*, *A. lanthieri*, *A. skirrowii*, and *A. thereius* are currently considered emerging food-borne pathogens [9,26]. The analysis indicates a loss of genes, and a smaller genome, a phenomenon that was proposed as linked to genomes of pathogenic bacteria compared to their nonpathogenic or less pathogenic relatives [27,28]. A subgroup of the group 1 (subgroup S; *A. cryaerophilus*, *A. porcinus*, *A. skirrowii*, *A. thereius* and *A. trophiarum*) in turn, shows a reduced size of the genome and a smaller number of genes compared to the other seven species of the group and compared with the species belonging to groups 2-5 (*p < 0.05*) (Supplementary Table 1).

The gene class pathway analysis (KAAS, KEGG codes) of the orthogroups (OGs) in the 21 genomes revealed sequences belonging to 115 gene classes, of which the most abundant are linked to metabolic pathways (319 OGs), biosynthesis of amino acids (86 OGs), biosynthesis of cofactors (79

OGs), microbial metabolism in diverse environments (71 OGs), ribosome (51 OGs), carbon metabolism (41 OGs). In agreement with these results, gene enrichment analysis performed to obtain Clusters of Orthologous Genes (COG) category codes revealed a presence of orthogroups present in all species, related to translation, ribosomal structure, and biogenesis, amino acid transport and metabolism, coenzyme transport and metabolism, energy production and conversion (Supplementary figure 1). This suggests the presence of a core genome composed of genes with functions linked to fundamental metabolic functions. This aspect is predictable considering the importance of these genes in different bacteria for their primary metabolism. Moreover, the importance of amino acid transport and metabolism as opposed to carbohydrate transport and metabolism (33 orthogroups, COGs) previously reported, has been confirmed by the present data, where a large number of sequences related to amino acid metabolism were observed [9,16,29]. However, the analysis showed a lack of information in the COGs database about specific orthogroups that result to be unknown. Further studies are required to elucidate the function of these genes (Figure 1).

Dendrograms constructed on amino acidic and 16S rRNA nucleotide sequences displayed the absence of a clear separation between the members of the 5 proposed groups (Figure 2). The absence of a clear separation is observable in the different genomic distances in the dendrograms between groups and internal to the groups. The genomic distance in some cases was higher within the same group than between different groups (Figure 2). Furthermore, the orthogroups and gene cluster analysis showed a separation of the species *A. butzleri* and *Arcobacter lacus* from their catalogued group (Figure 1-2). These aspects indicate an

absence of phylogenetic stability of the newly proposed genera, as also suggested by other authors [3,4].

Table 4.1. *Arcobacteraceae* species information. In this table the strains codes of the 20 species belonging to the *Arcobacteraceae* family and two outgroups sequences that were used are also shown. The codes about Sequence Read Archive (SRA) are shown. Accession numbers indicated with "*" have been retrieved from ENA (https://www.ebi.ac.uk/ena/browser/home), the other strains sequenced in this work are available at the NCBI bioproject PRJNA808439. The column "group" reports the group of strains as indicated by Pérez-Cataluña and colleagues [1].

Species	Strain	Group	Run accession
Arcobacter butzleri	LMG 10828^T	1	SRR18076128
Arcobacter cibarius	LMG 21996^T	1	SRR3664169*
Arcobacter cryaerophilus	*LMG* 24291^T	1	SRR7985382*
Arcobacter cryaerophilus	LMG 10829	1	SRR7985571*
Arcobacter porcinus	LMG 24487^T	1	SRR18076131
Arcobacter skirrowii	LMG 6621^T	1	SRR18076130
Arcobacter thereius	LMG 24486^T	1	SRR18076129
Arcobacter trophiarum	LMG 25534^T	1	SRR18076127
Arcobacter vandammei	LMG 31429^T	1	SRR18076126
Arcobacter vitoriensis	LMG 30050^T	1	SRR18076123
Arcobacter faecis	LMG 28519^T	1	SRR18076124
Arcobacter lacus	LMG 29062^T	1	SRR5221256*
Arcobacter lanthieri	LMG 28516^T	1	SRR18076125
Arcobacter ellisii	LMG 26155^T	2	SRR7588928*
Arcobacter venerupis	LMG 26156^T	2	SRR5914676*
Arcobacter suis	LMG 26152^T	2	SRR7591528*
Arcobacter halophilus	CCUG 53805^T	3	SRR7587110*
Arcobacter molluscorum	LMG 25693^T	3	SRR7591199*
Arcobacter mytili	LMG 24559^T	3	SRR7588217*
Arcobacter bivalviorum	LMG 26154^T	4	SRR7586655*
Arcobacter nitrofigilis	LMG 7704^T	5	NC_014166.1*
Campylobacter jejuni	NCTC 11168^T	outgroup	NC_002163.1*
Helicobacter pylori	MT 5135^T	outgroup	NZ_CP071982.1*

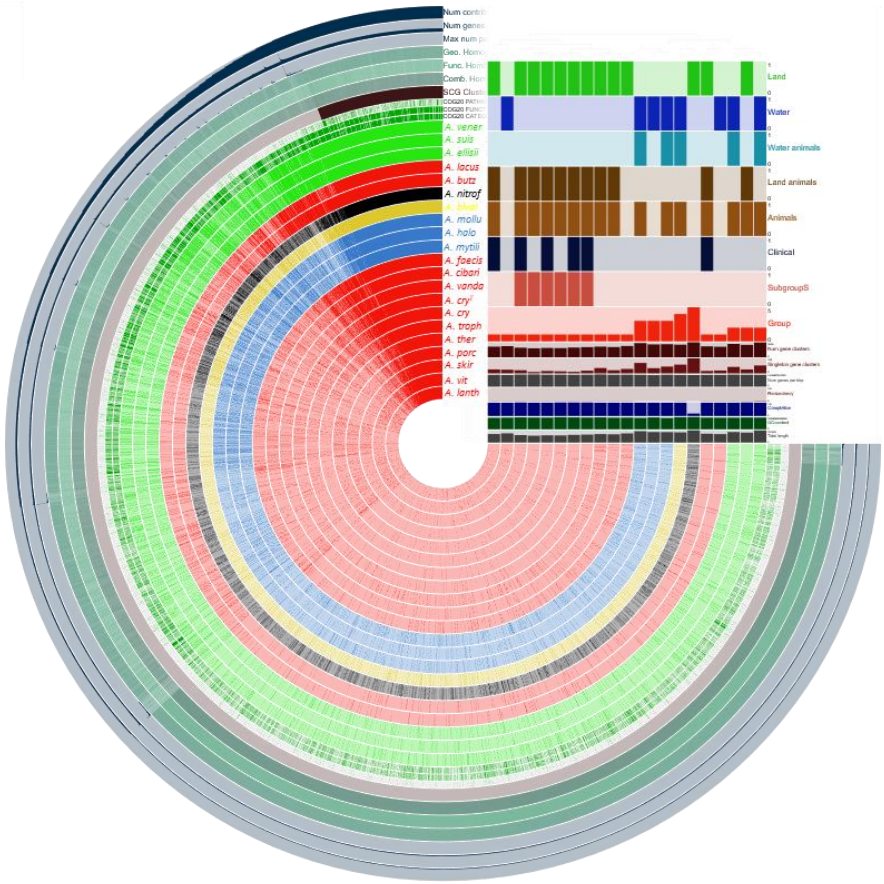


Figure 4.1. Anvi'o Analysis. The figure shows genomes characteristics of the *Arcobacteraceae* species object of study sorted for gene cluster presence and absence. In this figure are observable the different sources of isolation and the different species groups as well as general information regarding the genomes (size, GC content).

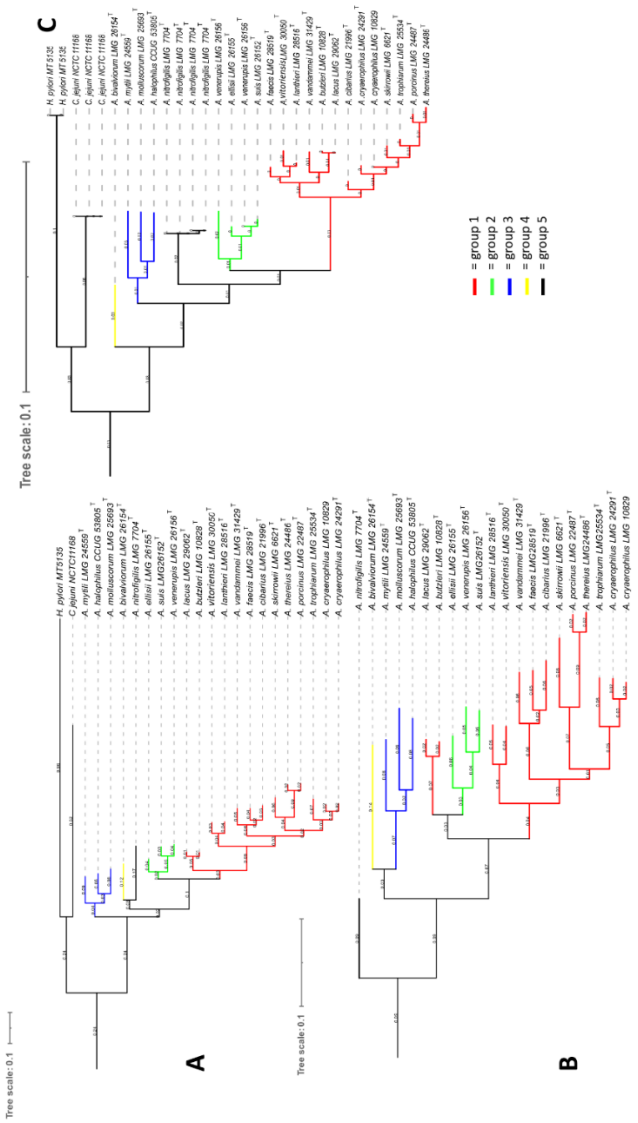


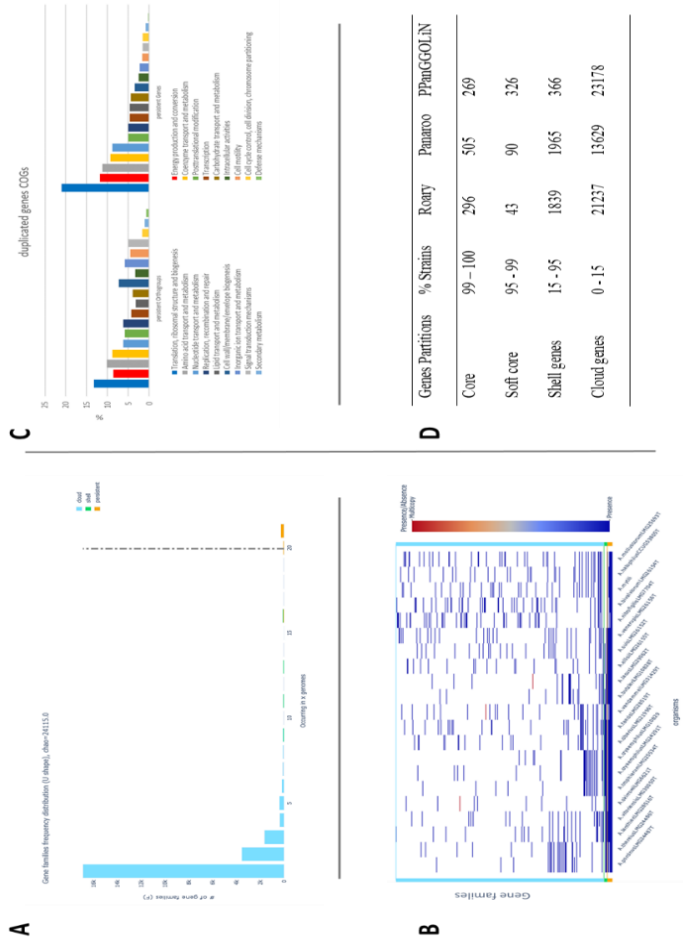
Figure 4.2. Dendrograms of *Arcobacteraceae* species studied. The trees in the figure have been computed from different input sequences: amino acid annotated sequences, best bcgTree (**A**), tree inferred from all orthogroups (**B**) and 16S rRNA sequences (**C**). The groups of species from Pérez-Cataluña and colleagues are indicated by branches colors as indicated in the figure [1].

4.2.2. Pangenomes partition shows a wide presence of persistent genes

Pangenomes analysis has been performed on the gene sequences of 20 *Arcobacteraceae* species obtained during the functional annotation process. The results show a presence of 505 core genes with Panaroo tool. However, pangenome analysis performed with Roary and PPanGGOLiN show only the presence of 296 and 269 core genes, respectively, demonstrating again that largely different outcomes that can be obtained when different analysis tools are applied, as also previously reported [30,31]. As conflicting data about the functional pangenome were obtained, OGs (orthoFinder) and gene family comparison have been performed respectively with orthoFinder and PPanGGOLiN. The results showed 1324 persistent OGs (32.19% of assigned OGs), 501 persistent gene families (average of the nr. of persistent families present in the 21 genomes (persistent = present in a range between 90 and 99% of total species) and a relevant number of gene families in cloud partition (average 1669 gene families) [32] (Figure 4.3). The high number of cloud gene families representing genes shared only between few genomes demonstrates a flexible genome linked to the adaptation of the species at different environmental matrices and hosts. The analysis of the number of gene families among the strains led to observe a lower presence of persistent gene families in *A. nitrofigilis* (376) compared with the other species. This result is in agreement with the lower "completion" of its genome from the analysis with Anvi'o software and confirmed by PPanGGOLiN completeness analysis (66%). The analysis of gene families duplication showed the presence of a wide number of genes in the persistent partition, leading to the hypothesis of a gene duplication strategy linked to a phenomenon of redundancy to keep certain metabolic functions

active even in the event of mutations. This is not completely surprising, considering the gene enrichment results on persistent genes and GOs which detected a wide presence of sequences related to fundamental functions, most of which linked to translation, ribosomal structure and biogenesis, energy production and conversion and amino acid transport and metabolism (Figure 4.3). The last class listed further supports again the importance of amino acid metabolism for the metabolism of this bacterial family [9,16,29].

Taking into account the inclusion of different *Arcobacteraceae* species in the present analysis, these results are not really different from other pangenome studies, such as the pangenome of *Campylobacter jejuni* and *Campylobacter coli* or the *Enterococcus* spp. pangenome [33–36]. The present results support the claim that the *Arcobacteraceae* species belong to one single genus within the *Arcobacteraceae* family, as also argued by On and colleagues [3].



Genes Partitions	% Strains	Roary	Panaroo	PPanGGOLiN
Core	99 – 100	296	505	269
Soft core	95 - 99	43	90	326
Shell genes	15 - 95	1839	1965	366
Cloud genes	0 - 15	21237	13629	23178

Figure 4.3. Pangenome analysis of the *Arcobacteraceae* species. The figure shows different data about pangenome analysis performed on 21 isolates belonging *Arcobacteraceae* family. The histogram **(A)** shows gene families frequency, shell and persistent genes are indicated respectively by green and orange bar on the right. Figure **B** shows the presence of multiple genes copies (blue = present, red = absent), it's possible to observe that most of the multiple copies are in the lower part of the graph in correspondence with the genomic portion relating to the core and shell genome. The histogram **C** shows the percentage of the different persistent OGs and gene pathway families, below this figure are showed the different gene partitions obtained by different tools **(D)**.

4.2.3. Animal related species show a different gene classes composition

The significantly reduced genome size of the species in group 1 indicates the loss of genes in those animal associated species and a link to their pathogenic nature. This aspect has also been reported before for other pathogenic bacterial species [27,28]. The reduction of genome size is linked to an evolutionary adaptation to the host by pathogenic species. Particularly relevant is the possibility to lose genes encoding proteins detected by the host immune system as well as the loss of specific pathways related to environmental survival and spreading [28,37]. To understand the relation between the absence of genes and certain characteristics, it is essential perform an analysis on the gene classes on the group with a reduced genome to be compared to the gene classes of other groups (Supplementary Table 4.2). Data of group 1 showed the presence of a smaller number of sequences related to amino acid transport and metabolism, energy production and conversion, carbohydrate transport and metabolism, intracellular trafficking, secretion and vesicular transport, post-translational modification, signal transduction mechanisms together with lipid transport and metabolism, defense mechanisms, cell motility, protein turnover, chaperones, cell wall-membrane-envelope biogenesis and translation, ribosomal structure and biogenesis ($p < 0.05$). Part of these pathway classes has also been lost in other pathogenic bacteria [28]. The lipid transport metabolism genes loss, together with cell wall and membrane envelope biogenesis genes lost, can be linked to an "escape" strategy from the host immune system changing part of LPS O-antigen structures. The loss of LPS-related genes has been observed in pathogenic *Yersinia pestis* [28,38]. Genes linked to chemotaxis, energy

production, carbohydrate, amino acid metabolism and flagella production can be lost during bacterial evolution to adapt to changing environmental niches from the normal environment to host and host related environmental conditions [28].

However, the ratio of the orthogroups gene classes (% of specific classes on COGs total, EggNOG annotated) including lipid transport and metabolism, cell motility, cell wall-membrane-envelope biogenesis, translation, cell division, chromosome partitioning and ribosomal structure, biogenesis, and cell cycle control results higher in group 1 compared to the species of the groups 2 to 5 ($p < 0.05$). This suggests not a simple gene loss but an evolutionary adaptation to environmental and host conditions. The higher percentage of genes linked to lipid transport and metabolism and cell motility suggests a possible role during host colonization as previously shown for *A. butzleri* virulence [14,16].

Group 1 includes five species (*A. cryaerophilus*, *A. porcinus*, *A. skirrowii*, *A. thereius* and *A. trophiarum*) with a smaller genome, and exposed as subgroup S. The species of subgroup S, compared to the other seven species of group 1, show a lower number of genes linked to energy production and conversion, carbohydrate transport and metabolism, defense mechanism, cell wall-membrane-envelope biogenesis and signal transduction mechanisms ($p < 0.05$). This suggests a further genome evolution adaptation. Moreover, the loss of these genes suggests the possibility of an underestimation of the importance of these species in veterinary and human clinical cases due to a reduced capability to grow on laboratory media conditions [39]. Regarding this lack of genes related to slow growth *in vitro* in the subgroup S, the comparison with the most isolated species, *A. butzleri*, shows a lack of group-specific carbohydrates-

related genes. Five OGs have been detected only in *A. butzleri*, these OGs are linked to ADP-glyceromanno-heptose-6-epimerase-activity, polysaccharide-deacetylase, haloacid-dehalogenase-like-hydrolase, membrane protein EamA and an ABC-transporter. Moreover, only a carbohydrate-related OG (Membrane transport protein) MFS is present in subgroup S and absent in *A. butzleri*. The absence of these metabolic genes could explain the difficult isolation of species belonging to subgroup S and consequently to their underestimation in clinical cases, suggesting the necessity of molecular based methods for their detection.

4.2.4 Specific pathways linked to different *Arcobacteraceae* groups

Species within the family *Arcobacteraceae* have specific pathways correlated to different groups, and, by consequence, to different environmental conditions linked to isolation sources (Spearman's correlation of pathway completeness percentages, S. corr., $p < 0.05$) (Figure 4.4). Species within group 1 show a positive correlation (S. corr. 0.64) to carbapenem resistance demonstrating the presence of antibiotics resistance-related genes present in the species often isolated from animals. Furthermore, the lysine biosynthesis pathway is positively correlated (S. corr. 0.57), whereas cobalamin biosynthesis and assimilatory nitrate reduction are negatively correlated (S. corr. -0.77, -0.71). This negative correlation was also observed in the subgroup S (S. corr. -0.46, -0.62). Furthermore, cobalamin biosynthesis and assimilatory nitrate reduction pathway show a positive correlation in group 2 (S. corr. 0.62, 0.61 respectively), composed of species isolated from shellfish (*A. ellisi*, *A. venerupis*) and pork meat (*A. suis*), but not related to pathogenicity in humans and present in mammals. The different correlation among groups of cobalamin-related genes, suggests the opportunity of animal-related

species to consume cobalamin from the host, as also reported for other pathogens [40]. The species included in group 3, isolated from the environment and animals, show a positive correlation with threonine and ectoine biosynthesis (S. corr. 0.58, 0.63) (Supplementary Table 4.3). Ectoine is an osmosis-stress protective molecule, and the presence of this pathway is not surprising considering their link to an environment characterized by a high salt concentration [41]. Interesting, the ectoine pathway is negative correlated with the *Arcobacteraceae* species group 1 (S. corr. - 0.49). Group 1 (containing species related to animals and humans) shows the presence of thiopeptide-related genes. These genes are linked to the production of thiopeptide, an antibiotic active against Gram-positive bacteria [42]. The presence of these sequences has been observed in nine of the twelve species of group 1, suggesting an advantage over other bacterial species. Another metabolic pathway linked to a particular species ecological niche is related to phthoxazolin (NRPS-T1PKS) [43,44]. This metabolite is a cellulose inhibitor, and its presence in *A. nitrofigilis*, associated with *Spartina alterniflora*, suggests a role in bacteria-plant interaction [45].

Due to the high presence of hypothetical proteins, it is complicated to evaluate the presence of complete pathways, however, the results show specialization of some *Arcobacteraceae* species with different environmental and host characteristics.

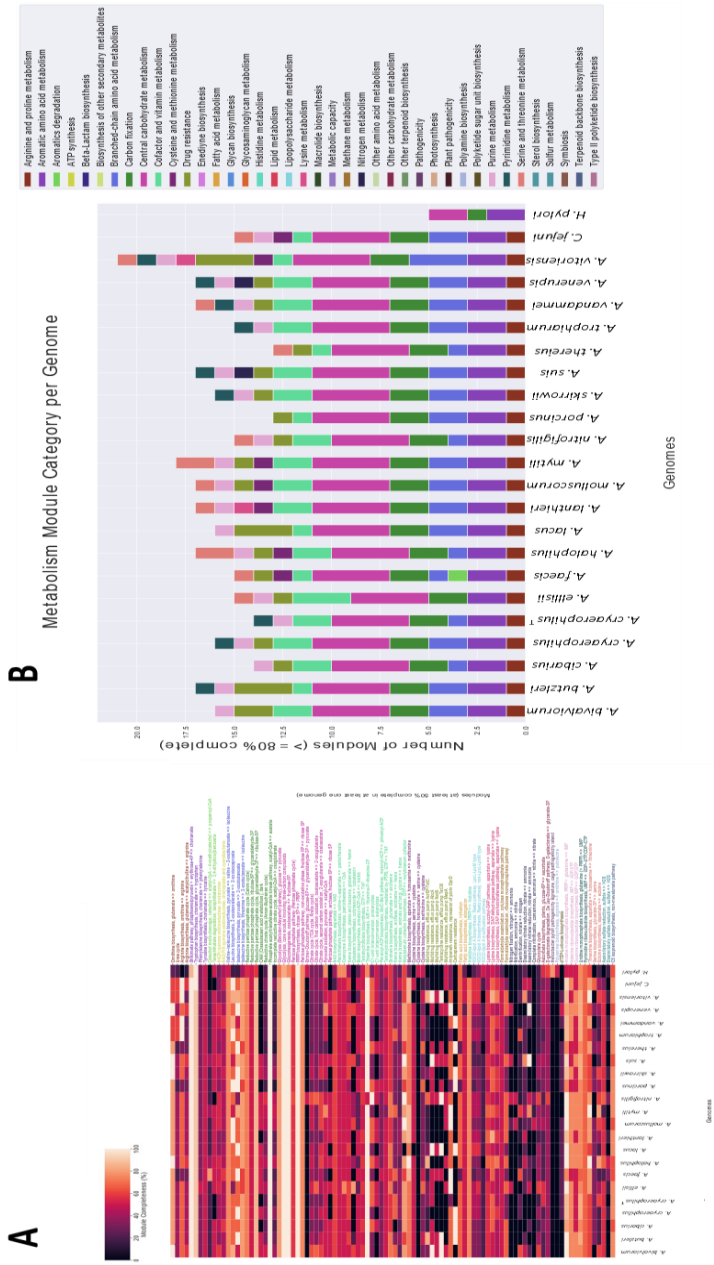


Figure 4.4. Genomes Annotated Pathway. The heatmap (A) shows the pathway modules with at least 50% of completeness in at least one genome. The bar chart (**B**) shows the pathway modules with at least 80% of completeness in at least one genome.

4.2.5 Putative virulence genes are not strongly correlated to different groups

In most of the *Arcobacter* pathogenicity studies performed so far, the presence of 9 putative virulence genes (PVGs), *hecA, hecB, ciaB, cadF, cj1349, irgA, pldA, mviN* and *tlyA*) are commonly assessed [46]. As already stated above, some *Arcobacteraceae* species are linked to infection in humans and animals, most of these species are included in group 1. The 21 genomes included in the present study harbored at least one virulence associated gene (Figure 4.5). The percentage of similarity of sequences compared with reference genomes PVGs sequences (RM 4018, query coverage ≥ 80%) shows that few traits are moderately associated to specific groups (Spearman's correlation *p-value* < *0.05*). The PVGs *cj1349* (fibronectin-binding protein) and *ciaB* (invasin) are positively correlated respectively to group 1, and the clinical-related species group (0.49, 0.51 respectively). The gene *mviN* (virulence factor) is negatively correlated to group 3 that includes species isolated from marine animals and aquatic environments (- 0.56). However, the absence of a strong association between PVGs and specifics groups suggests the necessity to consider new genes in the study of *Arcobacteraceae*, in particular when focused on their virulence mechanism.

In the present study, different Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) sequences have been detected. However, *A. butzleri, A. lacus* and *A. trophiarum* don't contain CRISPR or CAS sequences while the other species contain at least one of these sequences. A degeneration of CRISPR-CAS sequences has been linked to an increase of virulence and antibiotic resistance in different bacteria, including *C. jejuni* (Supplementary Table 4.1) [47]. A lack of CRISPR–CAS sequences

has been observed in the predominant species considered pathogenic *A. butzleri*, together with a species with a high genome similarity, *A. lacus*, although no much information on the pathogenicity of the last species has been reported. Moreover, also *A. trophiarum*, which displayed an *in vitro* colonization–invasion behavior similar to *A. butzleri*, doesn't show a presence of CRISPR-CAS sequences [48].

The Scoary comparison between the 5 groups proposed within *Arcobacteraceae*, and between strains characterized by different sources of isolation, allowed the detection of 47 group-correlated OGs only in group 1 (Scoary, Bonferroni's and Benjamini-Hochberg's methods corrected, $p\text{-value} < 0.05$), of which 8 orthogroups are positively related to group 1. Among these OGs Hemolysins-CBS-domain genes orthogroup is present in all 13 species belonging to group 1. Taking into account that part of those species is considered pathogenic, the presence of this gene hypothesizes the importance of hemolysis in the pathogenicity. The other genes related to group 1 are linked to different pathways, in particular to amino acid transport and metabolism (3 OGs positive correlated and 3 negative correlated), while energy production and conversion is the class with the largest negatively correlated OGs. The positively correlated amino acid transport metabolism COGs are linked to saccharopine dehydrogenase NADP binding domain, component of the transport system for branched-chain amino acids and EamA like transporter family, while PFAM Lysine exporter protein (LYSE YGGA), PFAM Aminotransferase class I and II and Glutamate synthase domain 2 are negatively correlated to group 1. In addition to these genes, other sequences are correlated to the species of group 1 but with a lower statistical significance (Scoary, Benjamini-Hochberg's method corrected $p\text{-value} < 0.05$; no Bonferroni's

correction). Orthogroups related to *motA tolQ, exbB, exbD* and *tonB* (TonB functions) are negatively related to group 1 and absent in all related species. However, another orthogroup codified for MotA/TolQ/ExbB proton channel family results positively correlated with group 1 suggesting a specific function of different orthologues related to different *Arcobacteraceae* species as already suggested for different *A. butzleri* strains [16]. Moreover, another positively correlated orthogroup is linked to β-lactamase activity, present in all species of group 1 indicating the presence of antibiotic resistance-related sequences [49].

Although the number of OGs containing hypothetical proteins remains high and does not allow the evaluation of entire pathways, these aspects lead to speculate about the importance of some sequences with the relevance of amino acid transport for group 1 assuming the possibility of an evolutionary adaptation of animals-related species through the loss of certain sequences linked to basic metabolism and the maintenance of these genes in other orthogroups.

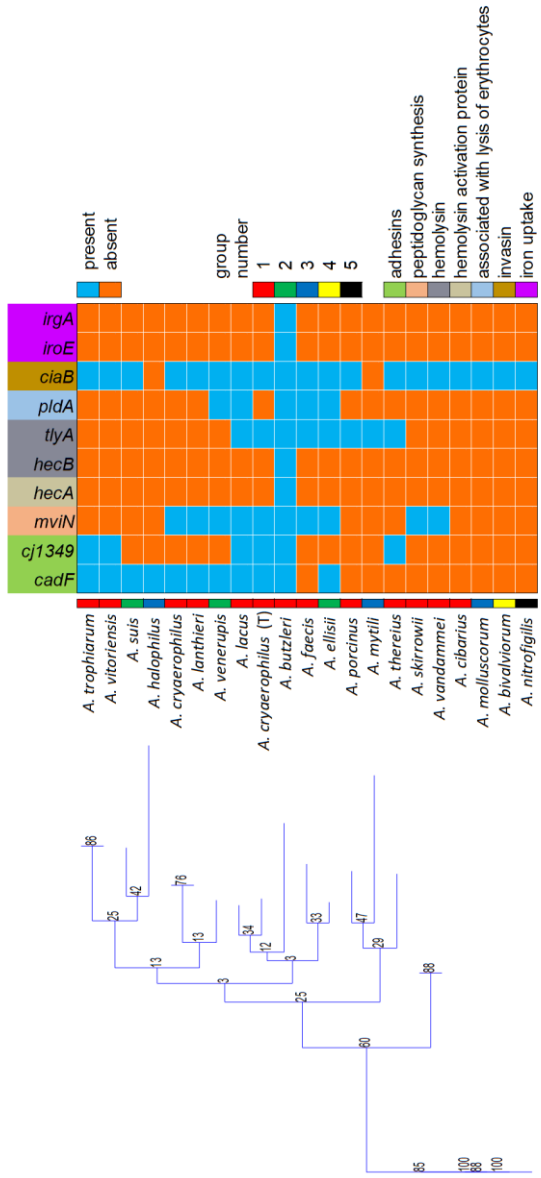


Figure 4.5. PVGs presence-absence dendrogram. The figure shows the dendrogram produced on the presence-absence binary matrix of nine genes currently considered virulence-related. The tree (Jaccard, Neighbor-joining) shows bootstrap 10.000 value while the different colors near species names indicate the species belonging group.

4.3. Conclusions

The species belonging to the *Arcobacteraceae* family are characterized by different sources of isolation and association with clinical relevance. This aspect is also observed in a differentiation in genome size and content. However, data show a large presence of cloud genes in the *Arcobacteraceae* pangenome. The *Arcobacteraceae* genome partitions (core, cloud and soft genome) don't differ from other species belonging to other bacterial genera like *Campylobacter* spp. suggesting the existence of only one bacterial genus "*Arcobacter*". This consideration is supported by species belonging to hypothesized genera (groups) during comparative genomics analysis (Figure 1-2).

The moderate correlation of some PVGs to species related to animals and human clinical cases suggests the need to find new gene candidates for the detection of pathogenic *Arcobacter* species . However, at present, evolutionary, and phylogenetic studies are still hampered by a wide presence of hypothetical protein. Another limit is the absence in the database of a large number of whole genomes raw reads, making difficult the comparison of different species. Although these limitations make the study of *Arcobacteraceae* challenging, the smaller genome size of species considered animal and human clinical-related led to the detection of the presence of certain gene classes and the loss of others. This suggests an evolutionary specialization of some species to animal hosts, leading to speculate about a link with the underestimation in clinical cases of *Arcobacter* spp. infections caused by their difficulty of *in vitro* cultivation. Moreover, the maintenance and presence of particular sequences demonstrate the importance of pathways for specific bacterial metabolism and in consequence for bacterial lifestyle linked to different environmental

niches, not only for the species considered pathogenic but also for those linked to specific environments. The study of 20 *Arcobacter* species type strains allowed the assessment of the *Arcobacteraceae* pangenome, with the detection of some specific sequences linked to ecological niches. The increase in the coming years of genomic sequences available and of information relating to *Arcobacter* spp. strains and species will lead to new information regarding *Arcobacteraceae* family favoring subsequent studies.

4.4. Material and methods

4.4.1. DNA extraction and genome sequencing

The different characteristics related to the environment or host from which the different strains have been isolated or their importance in the clinical field are indicated in figure 4.1. For whole-genome sequencing, high-quality DNA extracts were prepared using a Maxwell 16 tissue DNA purification kit (AS1030; Promega, Madison, WI, USA) and an automated Maxwell 16 DNA preparation instrument (AS2000; Promega). The final DNA extract was dissolved in 10mM Tris-HCL pH 8.5 and was treated with RNAse (2 mg/ml, 5µl per 100µl extract). DNA quality was checked by 1% (wt/vol) agarose gel electrophoresis, and DNA purity was evaluated using the QuantiFluor One double-stranded DNA system and the Quantus fluorometer (Promega). Paired-end 2 x 150-bp libraries were prepared at the Wellcome Trust Human Genome Center (Oxford, UK) using a NEBNext DNA library kit for Illumina (New England Biolabs, Ipswich, MA, USA) and sequenced on an Illumina HiSeq 4000 instrument.

4.4.2. Genomes retrieval and assembly

The raw Illumina paired-end accession numbers of 23 genomes raw sequences (of which 2 outgroups, *C. jejuni* and *H. pylori*), are indicated in table 1, a part of these genomes has been obtained from European nucleotide repository (https://www.ebi.ac.uk/ena/browser/home) as indicated in table 1. Sequencing reads were prepared for assembly by adapter trimming and read filtering using the Trimmomatic tool v0.39 [50]. Reads with phred scores below 30 were removed, and nonpaired reads were discarded. The fastq files quality has been checked with the software fastqc v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The genomes were assembled with Shovill v1.0.4 (https://github.com/tseemann/shovill) a pipeline that includes selection of the best kmer and annotation based on SPAdes v3.15.2 [51]. After sequences assembly, a quality check has been performed with Quast v5.0.2 [52]. The actual correspondence between genomes and different species has been evaluated using barrnap v0.9 (https://github.com/tseemann/barrnap) to extract 16S RNA gene sequences for a BLASTN 2.13.0+ comparison (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and GTDB-Tk v1.7.0 for an entire genome evaluation [53]. BLASTN 2.13.0+ has been employed for specific sequence comparisons.

4.4.3. Functional annotation and pathway annotation

The genomes have been annotated with the software Prokka v1.14.5 to obtain sequences and functions of the genes (functional annotation), and to obtain input gene sequence files (e.g. .gff, .fna, .faa) for subsequent bioinformatic analysis [54]. The pipeline KAAS v2.1

(https://www.genome.jp/kegg/kaas/), MicrobeAnnotator v2.0.5 (considering completeness of pathway at least 80% in one genome, kofamscan and the swissprot database) and emapper-2.1.6 (http://eggnog-mapper.embl.de/) have been applied to obtain data about gene pathways of aminoacidic and nucleotide sequences [55–57], while CRISPR-Cas++ v1.1.2 has been employed to detect CAS and CRISPR sequences [58]. The detection of pathways linked to secondary metabolites production has been performed with antiSMASH 6.0.1 [59].

4.4.4. Pangenome evaluation tools

To assess the pangenome characteristics of the species belonging to the *Arcobacteraceae* family, different pangenome analysis tools have been applied: Roary v3.13.0 and Panaroo v1.2.8 (the threshold for the analysis has been set at minimum 80% identity) have been used on .gff file from Prokka annotation to identify the core and accessory gene partitions jointly at the obtainment of a binary matrix relating to gene presence/absence in the species [31,60]. The bioinformatics tool PPanGGOLiN v1.1.136 has been applied with default options to identify additional data about pangenome partitions and genomic plasticity regions [32,61]. Graphical visualization of pangenomes partitions has been performed with Anvi'o v7.1 (mcl 8) [62–64]. To obtain data about *Arcobacteraceae* family, these analyses have been evaluated on the 21 genomes excluding the two outgroups. The orthogroups analysis has been performed with the use of OrthoFinder v2.5.4 [65], whereas the possible correlation between the presence and absence of particular orthogroups within different strains characteristics has been evaluated with Scoary v1.6.16 [66]. MegaX v10.1.7 has been used to obtain dendrograms from specific sequences,

while bcgTree v1.1.0 has been applied to obtain dendrograms from Prokka core amino acidic sequences [67,68].

4.4.5. Statistical analysis and visualization

The homogeneity tests have been checked respectively with Shapiro-Wilk's W and Modified Levene's tests (Brown-Forsythe test). Kruskal–Wallis (K-W) and Anova test were used to evaluate overall differences and variations between multiple groups, while Wilcoxon Rank sum test (WRS) and two-sample t-test have been performed to evaluate differences between two groups, for nonparametric (K-W, WRS) and parametric data (Anova, t-test). A *post hoc* analyses Dunn's and Tukey's tests were used for nonparametric and parametric data respectively. These statistical analyses have been performed with RStudio 2021.09.0 (R v3.6.1, https://www.r-project.org/). The Spearman's correlation test have been performed with Past4 v4.09 [69]. The dendrogram trees have been constructed, visualized and graphically curated with iTol online software (https://itol.embl.de/) [70].**4.5 Availability of data**

The raw sequence reads (SRA) of the genomes sequenced in this work have been deposited on NCBI (https://www.ncbi.nlm.nih.gov/) at bioproject number PRJNA808439. The sequence codes obtained from the ENA database (https://www.ebi.ac.uk/ena/browser/home) are indicated in Table 1 together with the SRA codes relating to the bioproject PRJNA808439.

4.6 References

[1] A. Pérez-Cataluña, N. Salas-Massó, A.L. Diéguez, S. Balboa, A. Lema, J.L. Romalde, M.J. Figueras, Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos, Front. Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.02077.

[2] T.P. Ramees, K. Dhama, K. Karthik, R.S. Rathore, A. Kumar, M. Saminathan, R. Tiwari, Y.S. Malik, R.K. Singh, *Arcobacter* : an emerging food-borne zoonotic pathogen, its public health concerns and advances in diagnosis and control – a comprehensive review, Vet. Q. 37 (2017) 136–161. https://doi.org/10.1080/01652176.2017.1323355.

[3] S.L.W. On, W.G. Miller, P.J. Biggs, A.J. Cornelius, P. Vandamme, *Aliarcobacter*, *Halarcobacter*, *Malaciobacter*, *Pseudarcobacter* and *Poseidonibacter* are later synonyms of *Arcobacter*: transfer of *Poseidonibacter parvus*, *Poseidonibacter antarcticus*, '*Halarcobacter arenosus*', and '*Aliarcobacter vitoriensis*' to *Arcobacter* as *Arcobacter parvus* comb. nov., *Arcobacter antarcticus* comb. nov., *Arcobacter arenosus* comb. nov. and *Arcobacter vitoriensis* comb. nov. , Int. J. Syst. Evol. Microbiol. 71 (2021). https://doi.org/10.1099/ijsem.0.005133.

[4] S.L.W. On, W.G. Miller, P.J. Biggs, A.J. Cornelius, P. Vandamme, A critical rebuttal of the proposed division of the genus *Arcobacter* into six genera using comparative genomic, phylogenetic, and phenotypic criteria, Syst. Appl. Microbiol. 43 (2020) 126108. https://doi.org/10.1016/j.syapm.2020.126108.

[5] L. Collado, J. Guarro, M.J. Figueras, Prevalence of *Arcobacter* in Meat and Shellfish, J. Food Prot. 72 (2009) 1102–1106. https://doi.org/10.4315/0362-028X-72.5.1102.

[6] M.J. Figueras, L. Collado, A. Levican, J. Perez, M.J. Solsona, C. Yustes, *Arcobacter molluscorum* sp. nov., a new species isolated from shellfish, Syst. Appl. Microbiol. 34 (2011) 105–109. https://doi.org/10.1016/j.syapm.2010.10.001.

[7] A. Levican, L. Collado, C. Aguilar, C. Yustes, A.L. Diéguez, J.L. Romalde, M.J. Figueras, Arcobacter bivalviorum sp. nov. and *Arcobacter venerupis* sp. nov., new species isolated from shellfish, Syst. Appl. Microbiol. 35 (2012) 133–138. https://doi.org/10.1016/j.syapm.2012.01.002.

[8] T.P. Ramees, R.S. Rathore, P. Suresh, B. Sailo, H.V. Mohan, A. Kumar, R.K. Singh, Genotyping and genetic diversity of *Arcobacter butzleri* and *Arcobacter cryaerophilus* isolated from different sources by using ERIC-PCR from India, Vet. Q. 34 (2014) 211–217.

https://doi.org/10.1080/01652176.2014.979511.

[9] D. Chieffi, F. Fanelli, V. Fusco, *Arcobacter butzleri* : Up-to-date taxonomy, ecology, and pathogenicity of an emerging pathogen, Compr. Rev. Food Sci. Food Saf. (2020) 1541-4337.12577. https://doi.org/10.1111/1541-4337.12577.

[10] 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.

[11] K. Houf, L.A. Devriese, L. De Zutter, J. Van Hoof, P. Vandamme, Development of a new protocol for the isolation and quantification of *Arcobacter* species from poultry products, Int. J. Food Microbiol. 71 (2001) 189–196. https://doi.org/10.1016/S0168-1605(01)00605-5.

[12] F. Giacometti, A. Lucchi, G. Manfreda, D. Florio, R.G. Zanoni, A. Serraino, Occurrence and genetic diversity of *Arcobacter butzleri* in an artisanal dairy plant in Italy, Appl. Environ. Microbiol. 79 (2013) 6665–6669. https://doi.org/10.1128/AEM.02404-13.

[13] M. Elmali, H.Y. Can, Occurence and antimicrobial resistance of *Arcobacter* species in food and slaughterhouse samples, Food Sci. Technol. 37 (2017) 280–285. https://doi.org/10.1590/1678-457X.19516.

[14] 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. https://doi.org/10.1016/j.meegid.2020.104213.

[15] F. Fanelli, A. Di Pinto, A. Mottola, G. Mule, D. Chieffi, F. Baruzzi, G. Tantillo, V. Fusco, Genomic Characterization of *Arcobacter butzleri* Isolated From Shellfish: Novel Insight Into Antibiotic Resistance and Virulence Determinants, Front. Microbiol. 10 (2019) 1–17. https://doi.org/10.3389/fmicb.2019.00670.

[16] D. Buzzanca, C. Botta, I. Ferrocino, V. Alessandria, K. Houf, K. Rantsiou, Functional pangenome analysis reveals high virulence plasticity of *Aliarcobacter butzleri* and affinity to human mucus, Genomics. 113 (2021) 2065–2076. https://doi.org/10.1016/j.ygeno.2021.05.001.

[17] L. Debruyne, K. Houf, L. Douidah, S. De Smet, P. Vandamme, Reassessment of the taxonomy of *Arcobacter cryaerophilus*, Syst. Appl. Microbiol. 33 (2010) 7–14. https://doi.org/10.1016/j.syapm.2009.10.001.

[18] P.-J. Kerkhof, S.L.W. On, K. Houf, *Arcobacter vandammei* sp. nov.,

isolated from the rectal mucus of a healthy pig, Int. J. Syst. Evol. Microbiol. 71 (2021). https://doi.org/10.1099/ijsem.0.005113.

[19] H.Y.K. Lam, M.J. Clark, R. Chen, R. Chen, G. Natsoulis, M. O'Huallachain, F.E. Dewey, L. Habegger, E.A. Ashley, M.B. Gerstein, A.J. Butte, H.P. Ji, M. Snyder, Performance comparison of whole-genome sequencing platforms, Nat. Biotechnol. 30 (2012) 78–82. https://doi.org/10.1038/nbt.2065.

[20] S. Suzuki, N. Ono, C. Furusawa, B.-W. Ying, T. Yomo, Comparison of Sequence Reads Obtained from Three Next-Generation Sequencing Platforms, PLoS One. 6 (2011) e19534. https://doi.org/10.1371/journal.pone.0019534.

[21] J.F. Vázquez-Castellanos, R. García-López, V. Pérez-Brocal, M. Pignatelli, A. Moya, Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut, BMC Genomics. 15 (2014) 37. https://doi.org/10.1186/1471-2164-15-37.

[22] J. Chun, A. Oren, A. Ventosa, H. Christensen, D.R. Arahal, M.S. da Costa, A.P. Rooney, H. Yi, X.-W. Xu, S. De Meyer, M.E. Trujillo, Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes, Int. J. Syst. Evol. Microbiol. 68 (2018) 461–466. https://doi.org/10.1099/ijsem.0.002516.

[23] Y. Liu, Q. Lai, Z. Shao, Genome-Based Analysis Reveals the Taxonomy and Diversity of the Family *Idiomarinaceae*, Front. Microbiol. 9 (2018) 1–15. https://doi.org/10.3389/fmicb.2018.02453.

[24] V. Bansal, C. Boucher, Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going?, IScience. 18 (2019) 37–41. https://doi.org/10.1016/j.isci.2019.06.035.

[25] S.L.W. On, D. Althaus, W.G. Miller, D. Lizamore, S.G.L. Wong, A.J. Mathai, V. Chelikani, G.P. Carter, *Arcobacter cryaerophilus* Isolated From New Zealand Mussels Harbor a Putative Virulence Plasmid, Front. Microbiol. 10 (2019). https://doi.org/10.3389/fmicb.2019.01802.

[26] P.-J. Kerkhof, A.-M. Van den Abeele, B. Strubbe, D. Vogelaers, P. Vandamme, K. Houf, Diagnostic approach for detection and identification of emerging enteric pathogens revisited: the (*Ali*)*arcobacter lanthieri* case, New Microbes New Infect. 39 (2021) 100829. https://doi.org/10.1016/j.nmni.2020.100829.

[27] G.G.R. Murray, J. Charlesworth, E.L. Miller, M.J. Casey, C.T. Lloyd, M. Gottschalk, A.W. (Dan) Tucker, J.J. Welch, L.A. Weinert, Genome Reduction Is Associated with Bacterial Pathogenicity across Different

Scales of Temporal and Ecological Divergence, Mol. Biol. Evol. 38 (2021) 1570–1579. https://doi.org/10.1093/molbev/msaa323.

[28] L.A. Weinert, J.J. Welch, Why Might Bacterial Pathogens Have Small Genomes?, Trends Ecol. Evol. 32 (2017) 936–947. https://doi.org/10.1016/j.tree.2017.09.006.

[29] F. Rovetto, A. Carlier, A.M. Van Den Abeele, K. Illeghems, F. Van Nieuwerburgh, L. Cocolin, K. Houf, Characterization of the emerging zoonotic pathogen *Arcobacter thereius* by whole genome sequencing and comparative genomics, PLoS One. 12 (2017) 1–27. https://doi.org/10.1371/journal. pone.0180493.

[30] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G.G. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: Rapid large-scale prokaryote pan genome analysis, Bioinformatics. 31 (2015) 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

[31] G. Tonkin-Hill, N. MacAlasdair, C. Ruis, A. Weimann, G. Horesh, J.A. Lees, R.A. Gladstone, S. Lo, C. Beaudoin, R.A. Floto, S.D.W. Frost, J. Corander, S.D. Bentley, J. Parkhill, Producing polished prokaryotic pangenomes with the Panaroo pipeline, Genome Biol. 21 (2020) 180. https://doi.org/10.1186/s13059-020-02090-4.

[32] 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, PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph, PLOS Comput. Biol. 16 (2020) e1007732. https://doi.org/10.1371/journal.pcbi.1007732.

[33] J.C. Golz, L. Epping, M.-T. Knüver, M. Borowiak, F. Hartkopf, C. Deneke, B. Malorny, T. Semmler, K. Stingl, Whole genome sequencing reveals extended natural transformation in *Campylobacter* impacting diagnostics and the pathogens adaptive potential, Sci. Rep. 10 (2020) 3686. https://doi.org/10.1038/s41598-020-60320-y.

[34] Z. Zhong, W. Zhang, Y. Song, W. Liu, H. Xu, X. Xi, B. Menghe, H. Zhang, Z. Sun, Comparative genomic analysis of the genus *Enterococcus*, Microbiol. Res. 196 (2017) 95–105. https://doi.org/10.1016/j.micres.2016.12.009.

[35] V. Bravo, A. Katz, L. Porte, T. Weitzel, C. Varela, N. Gonzalez-Escalona, C.J. Blondel, Genomic analysis of the diversity, antimicrobial resistance and virulence potential of clinical *Campylobacter jejuni* and *Campylobacter coli* strains from Chile, PLoS Negl. Trop. Dis. 15 (2021) e0009207. https://doi.org/10.1371/journal.pntd.0009207.

[36] G. Méric, K. Yahara, L. Mageiros, B. Pascoe, M.C.J. Maiden, K.A. Jolley, S.K. Sheppard, A Reference Pan-Genome Approach to Comparative Bacterial Genomics: Identification of Novel Epidemiological Markers in Pathogenic *Campylobacter*, PLoS One. 9 (2014) e92798. https://doi.org/10.1371/journal.pone.0092798.

[37] S. Wang, A. Meade, H.-M. Lam, H. Luo, Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in Rhizobiales, MSystems. 5 (2020) 1–19. https://doi.org/10.1128/mSystems.00438-20.

[38] S.W. Montminy, N. Khan, S. McGrath, M.J. Walkowicz, F. Sharp, J.E. Conlon, K. Fukase, S. Kusumoto, C. Sweet, K. Miyake, S. Akira, R.J. Cotter, J.D. Goguen, E. Lien, Virulence factors of Yersinia pestis are overcome by a strong lipopolysaccharide response, Nat. Immunol. 7 (2006) 1066–1073. https://doi.org/10.1038/ni1386.

[39] M. Kurokawa, S. Seno, H. Matsuda, B.-W. Ying, Correlation between genome reduction and bacterial growth, DNA Res. 23 (2016) 517–525. https://doi.org/10.1093/dnares/dsw035.

[40] C.A. Rowley, M.M. Kendall, To B12 or not to B12: Five questions on the role of cobalamin in host-microbial interactions, PLOS Pathog. 15 (2019) e1007479. https://doi.org/10.1371/journal.ppat.1007479.

[41] A.A. Richter, C.-N. Mais, L. Czech, K. Geyer, A. Hoeppner, S.H.J. Smits, T.J. Erb, G. Bange, E. Bremer, Biosynthesis of the Stress-Protectant and Chemical Chaperon Ectoine: Biochemistry of the Transaminase EctB, Front. Microbiol. 10 (2019) 1–20. https://doi.org/10.3389/fmicb.2019.02811.

[42] A.A. Vinogradov, H. Suga, Review Introduction to Thiopeptides : Biological Activity , Biosynthesis , and Strategies for Functional Reprogramming, Cell Chem. Biol. 27 (2020) 1032–1051. https://doi.org/10.1016/j.chembiol.2020.07.003.

[43] J. Rang, Y. Li, L. Cao, L. Shuai, Y. Liu, H. He, Q. Wan, Y. Luo, Z. Yu, Y. Zhang, Y. Sun, X. Ding, S. Hu, Q. Xie, L. Xia, Deletion of a hybrid NRPS-T1PKS biosynthetic gene cluster via Latour gene knockout system in *Saccharopolyspora pogona* and its effect on butenyl-spinosyn biosynthesis and growth development, Microb. Biotechnol. 14 (2021) 2369–2384. https://doi.org/10.1111/1751-7915.13694.

[44] Y. TANAKA, I. KANAYA, Y. TAKAHASHI, M. SHINOSE, H. TANAKA, S. OMURA, Phthoxazolin A, a specific inhibitor of cellulose biosynthesis from microbial origin. I. Discovery, taxonomy of producing microorganism, fermentation, and biological activity., J. Antibiot. (Tokyo). 46 (1993) 1208–1213.

https://doi.org/10.7164/antibiotics.46.1208.

[45] C.R. McCLUNG, D.G. PATRIQUIN, R.E. DAVIS, *Campylobacter nitrofigilis* sp. nov., a Nitrogen-Fixing Bacterium Associated with Roots of Spartina alterniflora Loisel, Int. J. Syst. Bacteriol. 33 (1983) 605–612. https://doi.org/10.1099/00207713-33-3-605.

[46] L. Douidah, L. de Zutter, J. Bare, P. De Vos, P. Vandamme, O. Vandenberg, A.-M. Van den Abeele, K. Houf, Occurrence of Putative Virulence Genes in *Arcobacter* Species Isolated from Humans and Animals, J. Clin. Microbiol. 50 (2012) 735–741. https://doi.org/10.1128/JCM.05872-11.

[47] R. Louwen, D. Horst-Kreft, A.G. Boer, L. Graaf, G. Knegt, M. Hamersma, A.P. Heikema, A.R. Timms, B.C. Jacobs, J.A. Wagenaar, H.P. Endtz, J. Oost, J.M. Wells, E.E.S. Nieuwenhuis, A.H.M. Vliet, P.T.J. Willemsen, P. Baarlen, A. Belkum, A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain–Barré syndrome, Eur. J. Clin. Microbiol. Infect. Dis. 32 (2013) 207–226. https://doi.org/10.1007/s10096-012-1733-4.

[48] A. Levican, A. Alkeskas, C. Gunter, S.J. Forsythe, M.J. Figueras, Adherence to and Invasion of Human Intestinal Cells by *Arcobacter* Species and Their Virulence Genotypes, Appl. Environ. Microbiol. 79 (2013) 4951–4957. https://doi.org/10.1128/AEM.01073-13.

[49] G.P. Hooff, J.J.A. van Kampen, R.J.W. Meesters, A. van Belkum, W.H.F. Goessens, T.M. Luider, Characterization of β-Lactamase Enzyme Activity in Bacterial Lysates using MALDI-Mass Spectrometry, J. Proteome Res. 11 (2012) 79–84. https://doi.org/10.1021/pr200858r.

[50] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics. 30 (2014) 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

[51] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. 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. https://doi.org/10.1089/cmb.2012.0021.

[52] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics. 29 (2013) 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

[53] P.-A. Chaumeil, A.J. Mussig, P. Hugenholtz, D.H. Parks, GTDB-Tk: a

toolkit to classify genomes with the Genome Taxonomy Database, Bioinformatics. 36 (2019) 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

[54] T. Seemann, Prokka: Rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

[55] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007) W182–W185. https://doi.org/10.1093/nar/gkm321.

[56] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (2019) D309–D314. https://doi.org/10.1093/nar/gky1085.

[57] C.A. Ruiz-Perez, R.E. Conrad, K.T. Konstantinidis, MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes, BMC Bioinformatics. 22 (2021) 11. https://doi.org/10.1186/s12859-020-03940-5.

[58] C. Pourcel, M. Touchon, N. Villeriot, J.-P. Vernadet, D. Couvin, C. Toffano-Nioche, G. Vergnaud, CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers, Nucleic Acids Res. 48 (2019) D535–D544. https://doi.org/10.1093/nar/gkz915.

[59] K. Blin, S. Shaw, A.M. Kloosterman, Z. Charlop-Powers, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 6.0: improving cluster detection and comparison capabilities, Nucleic Acids Res. 49 (2021) W29–W35. https://doi.org/10.1093/nar/gkab335.

[60] F. Sitto, F.U. Battistuzzi, Estimating Pangenomes with Roary, Mol. Biol. Evol. 37 (2020) 933–939. https://doi.org/10.1093/molbev/msz284.

[61] A. Bazin, G. Gautreau, C. Médigue, D. Vallenet, A. Calteau, panRGP: a pangenome-based method to predict genomic islands and explore their diversity, BioRxiv. 1 (2020) 2020.03.26.007484. https://doi.org/10.1101/2020.03.26.007484.

[62] A.M. Eren, Ö.C. Esen, C. Quince, J.H. Vineis, H.G. Morrison, M.L. Sogin, T.O. Delmont, Anvi'o: an advanced analysis and visualization platform for 'omics data, PeerJ. 3 (2015) e1319.

https://doi.org/10.7717/peerj.1319.

[63] E. Dumas, J. Geiselmann, R. Monte, P.T. Monteiro, M. Page, D. Ropers, Bacterial Molecular Networks, Springer New York, New York, NY, 2012. https://doi.org/10.1007/978-1-61779-361-5.

[64] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797. https://doi.org/10.1093/nar/gkh340.

[65] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (2019) 238. https://doi.org/10.1186/s13059-019-1832-y.

[66] O. Brynildsrud, J. Bohlin, L. Scheffer, V. Eldholm, Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary, Genome Biol. 17 (2016) 238. https://doi.org/10.1186/s13059-016-1108-8.

[67] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Mol. Biol. Evol. 35 (2018) 1547–1549. https://doi.org/10.1093/molbev/msy096.

[68] M.J. Ankenbrand, A. Keller, bcgTree: automatized phylogenetic tree building from bacterial core genomes, Genome. 59 (2016) 783–791. https://doi.org/10.1139/gen-2015-0175.

[69] and P.D.R. Øyvind Hammer, David A.T. Harper, PAST: paleontological statistics software package for education and data analysis, Palaeontol. Electron. 4 (2001) 1352–1357.

[70] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, Nucleic Acids Res. 47 (2019) W256–W259. https://doi.org/10.1093/nar/gkz239.

5. General Conclusions

5.1. Conclusions

The attention of the scientific community towards *Arcobacteraceae* family is significantly rising in the last years [1,2]. In this frame, the study initially focused on *A. butzleri* exploiting a comparative genomics approach, integrating the genomic data with physiological tests with human gut models with and without mucus.

The open pangenome of *A. butzleri* and the interchangeability of potential virulence genes was observed in the 32 strains analyzed. These aspects have been proposed as key genomic features for the host adaptation of these bacteria.

The variable virulome has been linked to strains phenotypes, in particular identifying the LPS assembling pathway as a potential strain-specific signature. The functional annotation of *A. butzleri* genomes showed the presence of putative virulence genes and antigen recognition markers. The *A. butzleri* strains from different hosts showed a similar colonization ability *in vitro*, without a marked invasiveness. However, part of the strains showed a different host colonization ability compared to other strains. These strains were chosen for subsequent transcriptomics studies in contact with host cells.

The expression of some genes currently considered virulence-associated by genomic analyses were confirmed by RNA-seq analysis. A fast response to environmental stimuli was observed in the strain LMG 11119 that showed a greater colonization. This aspect suggests an important role of environment-related gene expression response that could also play a role in adaptation under host colonization conditions.

The importance of iron metabolism during *A. butzleri* infections was suggested by overexpressed genes linked to iron transport in the strain characterized by higher colonization (LMG 11119).

Part of the DEGs were organic acid-related genes suggesting a signaling and/or substrate role of host derived organic acids during host-pathogen contact. This aspect was confirmed by chemical analysis (HPLC). The expression profile of genes related to organic acids metabolism was correlated with the consumption pattern of the respective acids during the contact of *A. butzleri* with host cells. The chemical analyzes further highlighted a host response to the bacterial presence. In this context a release of lactate by human cells when in contact with *A. butzleri* was observed.

The transcriptome data allowed to better understand the role of genes currently considered virulence-related from functional annotation studies. The studies presented in this thesis were performed on simplified *in vitro* models. Although the gut characteristics are more complex compared to the cellular models, these data can be used for subsequent targeted *in vitro* and *in vivo* studies. Moreover, the information obtained suggests new gene sequences for the detection and study of *A. butzleri*.

The information on *A. butzleri* suggest the need of studies expanded to other *Arcobacter* species. This necessity is made even more important by the lack of a unique opinion about *Arcobacteraceae* phylogeny and nomenclature [3,4].

The *Arcobacter* species are characterized by different sources of isolation and clinical relevance [1,3]. This aspect may be linked to a different genome size and gene content among species. The *Arcobacteraceae* genome partitions (core, cloud and soft genome) do not differ from other

species belonging to other bacterial genera like *Campylobacter* spp., suggesting the existence of only one bacterial genus "*Arcobacter*". This consideration is supported by species separation belonging to hypothesized genera (groups) during comparative genomics analysis.

The correlation of some PVGs to species related to animals and human clinical cases suggests the need to find new genes for the detection of pathogenic *Arcobacter* species and strains. The evolutionary, and phylogenetic studies are still difficult due to a wide presence of hypothetical proteins in *Arcobacter* spp. genomes. The absence in the database of many whole genomes raw reads make difficult the different species comparison. In this study a smaller genome size of species considered animal and human clinical-related compared to environment related species was observed. This aspect led to detect group-specific gene classes suggesting an evolutionary specialization of some species to animal hosts. The smaller genome size of clinical and animal related species leads to hypothesize a link between genome size and the pathogenicity of *Arcobacter* spp.. The presence of specific sequences demonstrates the importance of pathways for specific bacterial metabolisms (e.g., cellulose inhibition in *A. nitrofigilis*). This aspect suggests the existence of sequences related to various bacterial lifestyles in different environmental niches. The study of *Arcobacteraceae* pangenome allowed the detection of some specific sequences linked to ecological niches. The increase in the coming years of genomic sequences and on *Arcobacteraceae* information will be useful to better understand their environmental and clinical role.

The results presented in this thesis confirm the pathogenic potential of *A. butzleri* suggesting the presence of other species with virulence traits in

their genome. Even if the *Arcobacteraceae* family includes species from different environments and with different characteristics the pangenome data of this bacterial family suggest the existence of a single genus.

5.2. References

[1] T.P. Ramees, K. Dhama, K. Karthik, R.S. Rathore, A. Kumar, M. Saminathan, R. Tiwari, Y.S. Malik, R.K. Singh, *Arcobacter* : an emerging food-borne zoonotic pathogen, its public health concerns and advances in diagnosis and control – a comprehensive review, Vet. Q. 37 (2017) 136–161. https://doi.org/10.1080/01652176.2017.1323355.

[2] D. Chieffi, F. Fanelli, V. Fusco, *Arcobacter butzleri* : Up-to-date taxonomy, ecology, and pathogenicity of an emerging pathogen, Compr. Rev. Food Sci. Food Saf. (2020) 1541-4337.12577. https://doi.org/10.1111/1541-4337.12577.

[3] A. Pérez-Cataluña, N. Salas-Massó, A.L. Diéguez, S. Balboa, A. Lema, J.L. Romalde, M.J. Figueras, Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos, Front. Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.02077.

[4] S.L.W. On, W.G. Miller, P.J. Biggs, A.J. Cornelius, P. Vandamme, A critical rebuttal of the proposed division of the genus *Arcobacter* into six genera using comparative genomic, phylogenetic, and phenotypic criteria, Syst. Appl. Microbiol. 43 (2020) 126108. https://doi.org/10.1016/j.syapm.2020.126108.

6. Supplementary Materials

Supplementary Table 2.1. 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 logarithm$_{10}$ (log) and are relative to Mucus producing models (MP) and Not mucus producing models (NMP). Moreover, are indicated 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. ND indicates a not detectable bacterial load (< 100 CFU ml^{-1}).

		Mucus producing models (MP)						**Not Mucus producing models (NMP)**					
		log			**st. dev ±**						**st. dev ±**		
	***A. butzleri* Strain number and isolation source**	**T0**	**T1**	**T2**	**T0**	**T1**	**T2**	**T0**	**T1**	**T2**	**T0**	**T1**	**T2**
1_Human	stool sample	8.03	8.10	4.12	0.37	0.46	0.19	6.82	6.22	1.96	0.79	0.49	0.81
2_Human	stool sample	6.29	8.39	6.17	0.63	0.16	0.25	6.14	7.14	4.42	0.55	0.59	0.36
3_Human	stool sample	6.72	7.53	4.22	0.56	0.64	0.28	6.64	6.47	2.71	0.53	0.67	0.70
4_Dog	feces	7.23	7.19	2.81	0.96	0.88	1.40	6.99	6.56	2.63	0.56	0.46	0.69
5_Sheep	feces	7.99	8.65	5.30	0.54	0.82	0.59	8.06	7.34	4.43	0.30	0.12	0.26
6_Horse	feces	8.34	8.09	4.98	0.67	0.67	0.04	8.20	7.66	4.84	0.44	0.19	0.27
7_Bovine	feces	5.71	5.56	ND	0.39	0.28	NA	6.58	6.48	2.21	0.67	0.28	0.95
8_Chicken	skin	8.10	7.43	4.07	0.51	0.32	0.16	8.24	7.86	5.41	0.37	0.32	1.28
9_Pig	meat	7.38	7.90	3.73	0.70	0.97	0.13	7.87	6.04	4.01	0.49	1.51	0.16
10_Pig	rectum content, (rc1-13)	7.34	7.24	4.15	0.94	1.35	0.93	7.53	7.10	3.34	0.61	0.25	0.84
11_Pig	rectum content, (rc1-14)	8.43	6.29	2.69	0.40	0.71	1.37	8.41	7.02	4.01	0.28	0.24	0.15

12_Pig	rectum content, (rc2-10)	8.98	8.04	4.82	0.16	0.73	0.48	7.95	6.89	3.60	0.93	0.68	0.91
13_Pig	rectum content, (rc2-20)	7.41	8.03	4.08	0.98	0.55	0.27	7.38	6.80	3.09	0.59	0.47	0.78
14_Pig	duodenum content, (dc1-3AAN)	6.04	6.73	3.83	1.33	0.24	0.20	6.97	6.68	3.39	0.68	0.55	0.85
15_Pig	caecum content, (cm1-2AAN)	7.52	8.16	3.49	0.70	0.91	0.40	7.73	7.53	4.88	0.40	0.22	0.68
16_Pig	colon descendent mucus, (cdm1-1AAN)	7.81	7.72	5.11	0.27	0.44	0.86	8.23	7.02	3.33	0.38	0.07	0.84
17_Pig	colon descendent content, (cdc2-1AAN)	6.12	6.42	3.41	0.23	0.22	0.14	6.73	6.03	2.04	0.41	0.24	0.83
18_Pig	colon descendent content, (cdc2-2AAN)	7.61	8.24	5.06	0.93	0.68	0.21	8.06	7.37	4.47	0.60	0.20	0.15
19_Pig	rectum content , (rc1-2kAAN)	6.97	7.81	4.67	0.40	0.35	0.17	7.52	6.98	4.25	0.49	0.29	0.40
20_Pig	rectum content, (rc1-3AAN)	7.78	7.42	4.13	0.43	0.22	0.23	7.96	6.71	4.42	0.30	0.63	0.06
21_Pig	rectum mucus, (rm1-2AAN)	8.27	7.84	4.92	0.39	0.40	0.74	8.05	7.49	4.08	0.32	0.17	0.30
22_Pig	rectum content, (rc2-1AAN)	7.68	7.19	3.73	0.14	1.05	0.37	7.72	7.10	2.38	0.32	0.21	0.98
23_Pig	rectum mucus , (rm2-1AAN)	8.25	6.87	3.90	0.32	0.78	0.35	7.39	6.48	3.42	0.77	0.66	0.86
24_Human	stool sample	8.48	7.64	4.76	0.56	0.34	0.63	7.76	7.37	3.62	0.91	0.79	0.95
25_Human	stool sample	7.92	6.64	4.42	0.31	1.14	0.26	7.37	6.31	2.54	0.63	0.95	1.04
26_Human	stool sample	7.64	5.72	ND	0.44	0.81	1.00	7.52	6.08	1.42	0.38	0.50	0.87
27_Human	stool sample	7.80	6.58	3.99	0.57	0.86	0.26	8.32	6.89	3.56	0.45	0.39	0.14
28_Human	stool sample	7.24	6.52	ND	0.54	0.85	NA	7.97	6.49	ND	0.53	0.35	NA
29_Human	stool sample	7.26	8.13	4.47	0.63	1.16	0.66	7.01	6.56	3.81	0.42	0.55	0.10
30_Human	stool sample	7.54	6.87	4.25	0.29	0.06	0.36	7.53	6.88	3.35	0.52	0.53	0.84
31_Human	stool sample	7.05	5.78	ND	0.58	0.48	NA	7.56	5.51	1.46	0.51	0.34	0.95

32_Human	stool sample	6.95	6.93	3.85	0.71	0.28	0.32	6.77	6.39	3.11	0.40	0.26	0.81

		Mucus producing models						**Not Mucus producing models**					
		log			st. dev ±			log			st. dev ±		
		T0	T1	T2	T0	T1	T2	T0	T1	T2	T0	T1	T2
average	32 *A. butzleri* strains	7.50	7.30	4.25	0.55	0.62	0.46	7.53	6.80	3.43	0.52	0.44	0.64

Supplementary Table 2.2. 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), coverage ((read count * read length)/genome size), 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 classes are indicated with the number of sequences linked to their translation.

Origin:	Chicken skin	Cow faeces	Dog faeces	Horse faeces	Sheep faeces
Strain code:	8	7	4	6	5
Genomes size (Mbp)	2.31	2.25	2.18	2.18	2.14
coverage (X)	240.76	237.59	258.04	253.88	228.66
GC content (%)	26.89	26.82	26.9	26.9	26.92
Number of contings	44	28	25	41	32
total genes	2321	2262	2249	2204	2173
CDS	2276	2307	2205	2161	2128
tRNA	44	44	43	42	44
hypothetical proteins	584	620	547	513	498
Transposases	1	5	7	6	1
Prophage sequences (questionable)	0	0	0	0	0
Prophage sequences (incomplete)	1	4	3	2	4
CRISPR sequences	2	0	0	2	0
CAS sequence	gc1	0	0	gc2	0

Origin: Human faeces										
Strain code:	2	3	24	25	26	27	28	29	30	31	32

Genomes size (Mbp)	2.30	2.31	2.19	2.33	2.30	2.14	2.04	2.27	2.15	2.13	2.27
coverage (X)	195.32	205.69	534.34	417.31	439.34	439.66	457.27	491.39	542.66	486.87	518.43
GC content (%)	26.88	26.88	26.90	26.85	26.83	26.90	27.07	26.76	26.95	26.98	26.90
Number of contings	51	26	35	41	37	43	37	35	98	25	39
total genes	2278	2317	2169	2338	2301	2194	2101	2263	2150	2164	2325
CDS	2236	2271	2124	2294	2260	2148	2060	2222	2104	2120	2284
tRNA	41	45	44	43	40	45	40	40	45	43	40
hypothetical proteins	535	553	482	579	554	507	482	549	493	491	592
Transposases	1	0	0	5	0	8	1	2	2	3	13
Prophage sequences (questionable)	0	1	0	0	0	0	0	0	0	0	0
Prophage sequences (incomplete)	1	1	0	1	1	1	2	1	3	1	3
CRISPR sequences	2	0	3	0	1	1	1	3	0	0	2
CAS sequence	0	0	0	0	gc1	0	0	gc1, gc1	0	0	0

Supplementary Table 2.3. 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 **Figure 2.5.**

Gene or protein name	Locus tag	Protein description	Protein code
Virulence genes			
cj1349 (hp)	01173	Fibrinogen/fibronectin binding protein	S0ERZ0
cadF (oprF)	00304	Outer membrane porin F	P13794
porA	01532	Major outer membrane protein	P80672
similar to PorA	01513 (12)	Campylobacter major outer membrane protein	PF05538.5
pgaB	02101	Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase	P75906
ciaB (hp)	01320	Invasion protein CiaB	A0A516T5H0
irgA (cirA)	01765	Iron-regulated outer membrane protein	P17315
TonB domain	00673	Gram-negative bacterial TonB protein	PF03544.8
transport protein TonB domain	01740	transport protein TonB	PRK10819
ccdB	01833 (1)	Toxin ccdB	P62554
TonB receptor domain	00451	TonB dependent receptor	PF00593.18
TonB receptor domain	00164	TonB-dependent Receptor Plug Domain	PF07715.9
exbB	01366	Biopolymer transport protein ExbB	P0ABU7
exbD	01365	Biopolymer transport protein ExbD	P0ABV2
tdhA	00164	TonB-dependent heme receptor A precursor	Q7VNU1
epsE	01417	Type II secretion system protein E	P37093
epsF	01416	Type II secretion system protein F	P45780
xcpQ	02073	Type II secretion system protein D precursor	P35818
VirB6 protein	01499 (4)	TrbL/VirB6 plasmid conjugal transfer protein	PF046100.8
VirB10 protein	02277 (18)	Type IV secretion system protein virB10	Q9RPX5
VirB8 protein	02280 (18)	VirB8 protein	PF043335.7

virB4	02285 (18)	Type IV secretion system protein virB4	Q9R2W4
VirB3-like protein	02286 (18)	Type IV secretory pathway, VirB3-like family protein	PF05101.7
VirB2 family protein	02287 (18)	TrbC/VIRB2 family protein	PF04956.7
pldA	00819	Putative phospholipase A1 precursor	Q9K0U7
tlyA	02071	16S/23S rRNA (cytidine-2'-O)-methyltransferase TlyA	Q50760
mviN (murJ)	00802	Lipid II flippase MurJ (MviN)	P0AF16
HlyD domain	00646	HlyD family secretion protein	PF12700.1
hlyB	00557	Alpha-hemolysin translocation ATP-binding protein HlyB	Q47258
hecA (shlA)	00258 (5)	Hemolysin	P15320
hecA (hpmA)	00739	Hemolysin	P16466
hecB (shlB)	00740	Hemolysin transporter protein ShlB	P15321
cyaA	00751	Bifunctional hemolysin/adenylate cyclase precursor	P0DKX7
inlJ	00615	Internalin-J precursor	Q8Y3L4
inlA	00476 (29)	Internalin-A precursor	P0DJM0
phoP1	01355	Transcriptional regulatory protein PhoP	P23836
phoP2	02038	Virulence transcriptional regulatory protein	P0DM78
phoP3	02078	Alkaline phosphatase synthesis transcriptional regulatory protein PhoP	P13792
phoQ	00642	Sensor protein PhoQ	P23837
mprA	00641	Response regulator MprA	A0R3I8
mprB	00593	Signal transduction histidine-protein kinase/phosphatase MprB	Q7U0X3
bvgS	00652	Virulence sensor protein BvgS	P16575
resA	00790	Thiol-disulfide oxidoreductase ResA	Q81SZ9
DSBA-like domain protein	00882	DSBA-like thioredoxin domain protein	PF01323.14
fixL	00243	Sensor protein FixL	P10955
zraS	00184	Sensor protein ZraS	P37461
pdtaS	00514	Putative sensor histidine kinase PdtaS	O05846
ttrS	01409	Tetrathionate sensor histidine kinase TtrS	Q8ZPP6
iroE (besA)	00944	Alpha/beta hydrolase	O32102
pld	00355	Phospholipase D	Q92G53

STM3117	01571	Virulence protein	Q8ZM36
Exopolysaccharide production			
epsD	02102	Putative glycosyltransferase EpsD	P71053
epsH	01308 (27)	Putative glycosyltransferase EpsH	P71057
epsM	00892 (2)	UDP-N-acetylbacillosamine N-acetyltransferase	P71063
epsN	01040 (14)	Putative pyridoxal phosphate-dependent aminotransferase EpsN	Q795J3
epsJ	02099	Uncharacterized glycosyltransferase EpsJ	P71059
Host mucus interaction			
mltA	01902	Membrane-bound lytic murein transglycosylase A precursor	P0A935
mltD	00703	Membrane-bound lytic murein transglycosylase D	P0AEZ7
bdlA	00427	Biofilm dispersion protein	Q9I3S1
tabA	01795 (1)	Toxin-antitoxin biofilm protein tabA	P0AF96
rmlD	02118 (1)	dTDP-4-dehydrorhamnose reductase	O66251
rmlA	02120 (1)	Glucose-1-phosphate thymidylyltransferase 1	P37744
kfoC	02180 (2)	Chondroitin synthase	Q8L0V4
wbbL	02123 (1)	N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase	Q7D5T2
O-antigen ligase	01723	O-antigen ligase	A6TFL7
cotSA	02099 (2)	Spore coat protein SA	P46915
pga	01730	Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase	P75906
sunS	02010 (1)	SPBc2 prophage-derived glycosyltransferase sunS	O31986
rfaC	02088	Lipopolysaccharide heptosyltransferase 1	P24173
RimI domain	02146 (2)	ribosomal-protein-alanine N-acetyltransferase	PRK09491
Capsule biosynthesis domain	00898 (2)	Capsule polysaccharide biosynthesis protein	PF05159.8
Arylesterase	01709 (1)	Arylesterase precursor	Q07792
ureG	02010	Urease accessory protein UreG	Q9RP19
ureF	02208	Urease accessory protein UreF	Q79VJ2
ureE	02209	Urease accessory protein UreE	P50049
ureC	02210	Urease subunit alpha	Q07397
ureA	02211	Urease subunit alpha	P14916

ureD1	02212	Urease accessory protein UreD	Q79VJ0
Flagellum and motility			
Flagellar assembly protein H	01958	Flagellar assembly protein H	PRK06669
Flagellar basal body rod modification protein	01956	Flagellar basal body rod modification protein	PRK06655
Flagellar basal body-associated protein FliL	00029	Flagellar basal body-associated protein FliL	PRK07021
Flagellar biosynthesis protein FliR	01940	Flagellar biosynthesis protein FliR	PRK05701
Flagellar hook-associated protein FlgL	01936	Flagellar hook-associated protein FlgL	PRK08412
Flagellar hook-basal body protein FliE	01943	Flagellar hook-basal body protein FliE	PRK03907
Flagellar hook-length control protein FliK	00943	Flagellar hook-length control protein FliK	PF02120.10
flgB	01961	Flagellar basal body rod protein FlgB	P24500
flgC	01942	Flagellar basal-body rod protein FlgC	P0A1I7
flgE	01954	Flagellar hook protein FlgE	P50610
flgG	01962	Flagellar basal-body rod protein FlgG	P0A1J3
flgH	00031	Flagellar L-ring protein precursor	Q9PPM0
flgI	00024	Flagellar P-ring protein precursor	Q9PMJ8
flgK	00032	Flagellar hook-associated protein 1	P0A1J5
flhA	01935	Flagellar biosynthesis protein FlhA	O06758
flhB	00942	Flagellar biosynthetic protein FlhB	O67813
flhF	01946	Flagellar biosynthesis protein FlhF	Q01960
fliD	00033	Flagellar hook-associated protein 2	P24216
fliF	01960	Flagellar M-ring protein	P15928
fliG	01959	Flagellar motor switch protein FliG	O25119
fliI	01937	Flagellum-specific ATP synthase	O07025
fliM	00023	Flagellar motor switch protein FliM	P23453
fliN	01950	Flagellar motor switch protein FliN	P15070
fliQ	00026	Flagellar biosynthetic protein FliQ	P74931
fliS	00034	Flagellar protein FliS	P26609
ylxH	01945	Flagellum site-determining protein YlxH	P40742
hag	000565	Flagellin	Q05203
Flagellar filament 33 kDa core protein	01840	Flagellar filament 33 kDa core protein	P21989
flaA	02138 (6)	Flagellin A	P0A0S1
flaB2	01434	Flagellar filament 33 kDa core protein	P21989

flaB	00056 (17)	Flagellin B	Q07911
flagellar motor switch protein	00972	Flagellar motor switch protein	PRK06782
Flagellin N-methylase	00828	Flagellin N-methylase	PF03692.9
fliP	00683	Flagellar biosynthetic protein FliP precursor	P54700
fliW1	00058 (14)	Flagellar assembly factor FliW 1	O25769
fliW2	00062	Flagellar assembly factor FliW 2	O25929
motB	00222	Motility protein B	P56427
swrC	02513 (18)	Swarming motility protein SwrC	O31501
Chemotaxys			
cheB	00490	Chemotaxis response regulator protein-glutamate methylesterase	P07330
cheR	00492	Chemotaxis protein methyltransferase	P07801
cheA	00493	Chemotaxis protein CheA	P07363
cheY	00494	Chemotaxis protein CheY	P0A2D5
cheD	00491	Chemoreceptor glutamine deamidase CheD	Q9X005
cheV	00442	Chemotaxis protein CheV	P37599
cheW	00249	Chemotaxis protein CheW	P0A964
mcp4	00331	Methyl-accepting chemotaxis protein 4	Q9X1E2
mcpB	00187 (16)	Methyl-accepting chemotaxis protein McpB	P39215
pctA	02146	Methyl-accepting chemotaxis protein PctA	G3XD24
pomA	00223	Chemotaxis protein PomA	O06873
tap	00353	Methyl-accepting chemotaxis protein IV	P07018
tar	00146	Methyl-accepting chemotaxis protein II	P02941
tsr	00843	Methyl-accepting chemotaxis protein I	P02942
Antibiotic resistance			
arnA	00946 (10)	Bifunctional polymyxin resistance protein ArnA	O52325
arnB	01705	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase	Q8ZNF3
bcr	01657	Bicyclomycin resistance protein	P28246
bepD	02202	Efflux pump periplasmic linker BepD precursor	Q8G2M7
bepE	02203	Efflux pump membrane transporter BepE	Q8G2M6
bla	00853	Beta-lactamase OXA-15 precursor	Q51574
cat3	01822	Chloramphenicol acetyltransferase 3	P00484

eptA	01087	Phosphoethanolamine transferase EptA	P30845
fsr	00488	Fosmidomycin resistance protein	P52067
hcpC	00383	Putative beta-lactamase HcpC precursor	O25728
ileS	00342	Isoleucine--tRNA ligase	P00956
macA	00647	Macrolide export protein MacA	P75830
macB	00264	Macrolide export ATP-binding/permease protein MacB	P75831
mdtB	00283	Multidrug resistance protein MdtB	B7NCB1
mdtE	00686	Multidrug resistance protein MdtE precursor	P37636
mexA	01660	Multidrug resistance protein MexA precursor	P52477
mexB	01661	Multidrug resistance protein MexB	P52002
oprM	01873	Outer membrane protein OprM precursor	Q51487
pbp	00682	Beta-lactam-inducible penicillin-binding protein	P07944
relE	01278 (1)	mRNA interferase RelE	P0C077
rlmN	01162	putative dual-specificity RNA methyltransferase RlmN	Q7A600
sttH	00086	Streptothricin hydrolase	Q1MW86
tetA	00158	Tetracycline resistance protein, class C	P02981
uppP	01998	Undecaprenyl-diphosphatase	P60932
wbpD	01701	Group B chloramphenicol acetyltransferase	G3XD01

Supplementary Table 2.4. 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 gene cluster assembling (*) are indicated with asterisks.

Reference strains	Reference locus	Reference conting	Start	End	Length	Strand	Gene	Product	Pathway involvement	Protein code
Structure A (14,15)										
14	U_02031	19	9272	10138	866	-1		Glycosyl transferases group 1		PF00534.14
14	U_02032	19	10384	10752	368	+1	*gmhB*	D-glycero-alpha-D-manno-heptose-1%2C7-bisphosphate 7-phosphatase		Q9AGY5
14	U_02033	19	10761	11768	1007	+1	*hldD*	ADP-L-glycero-D-manno-heptose-6-epimerase		P67910
14	U_02034	19	11769	13205	1436	+1	*hldE*	Bifunctional protein HldE		P76658
14	U_02035	19	13186	13755	569	+1	*gmhA1*	Phosphoheptose isomerase 1		Q9PNE6
14	U_02036	19	13843	15741	1898	+1	*ltaS2*	Lipoteichoic acid synthase 2		O34952
14	U_02037	19	15754	16698	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
14	U_02038	19	16691	17752	1061	-1		hypothetical protein		
14	U_02039	19	17745	18665	920	-1		lipopolysaccharide core biosynthesis protein		PRK10422
14	U_02040	19	18662	19411	749	-1	*sunS*	SPBc2 prophage-derived glycosyltransferase SunS	*	O31986
14	U_02041	19	19408	20130	722	-1	*pgaB*	Poly-beta-1%2C6-N-acetyl-D-glucosamine N-deacetylase precursor		P75906
14	U_02042	19	20123	20872	749	-1		hypothetical protein		
14	U_02043	19	20985	22034	1049	+1	*cotSA*	Spore coat protein SA		P46915
14	U_02044	19	22457	23371	914	-1	*wbbL*	N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase		Q7D5T2
14	U_02045	19	23371	24657	1286	-1		O-Antigen ligase	**	PF04932.9
14	U_02046	19	24692	25567	875	-1		hypothetical protein		
14	U_02047	19	25568	26413	845	-1	*kfoC*	Chondroitin synthase	*	Q8L0V4
14	U_02048	19	26416	27453	1037	-1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
14	U_02049	19	27468	28346	878	-1	*rmlD*	dTDP-4-dehydrorhamnose reductase	*	O66251
14	U_02050	19	28339	28914	575	-1	*rfbC*	dTDP-4-dehydrorhamnose 3%2C5-epimerase	*	P26394
14	U_02051	19	28917	29807	890	-1	*rmlA1*	Glucose-1-phosphate thymidylyltransferase 1		P37744
14	U_02052	19	29837	30851	1016	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
14	U_02053	19	30851	31623	773	-1		hypothetical protein		
14	U_02054	19	31622	32542	914	-1	*htrB*	Lipid A biosynthesis lauroyl acyltransferase		P0ACV0
14	U_02055	19	32535	33536	1001	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173

14	U_0205 6	19	3353 6	3500 5	1469	-1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
14	U_0205 7	19	3500 5	3525 9	254	-1	*fdx*	Ferredoxin		P00208
14	U_0205 8	19	3530 3	3605 2	749	-1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
14	U_0205 9	19	3605 5	3743 1	1376	-1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
14	U_0206 0	19	3743 7	4187 6	4439	-1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831

Structure B (6)

6	H_0205 7	17	181	1299	1118	+1		hypothetical protein		
6	H_0205 8	17	1292	2551	1259	+1		O-Antigen ligase	**	PF04932.9
6	H_0205 9	17	2555	3181	626	-1		putative lipopolysaccharide biosynthesis O-acetyl transferase WbbJ		PRK09677
6	H_0206 0	17	3191	4240	1049	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
6	H_0206 1	17	4353	5102	749	+1		hypothetical protein		
6	H_0206 2	17	5095	5817	722	+1	*pgaB*	Poly-beta-1%2C6-N-acetyl-D-glucosamine N-deacetylase precursor		P75906
6	H_0206 3	17	5814	6563	749	+1	*sunS*	SPBc2 prophage-derived glycosyltransferase SunS	*	O31986
6	H_0206 4	17	6560	7480	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
6	H_0206 5	17	7473	8534	1061	+1		hypothetical protein		
6	H_0206 6	17	8527	9471	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
6	H_0206 7	17	9484	11382	1898	-1	*ltaS2*	Lipoteichoic acid synthase 2		O34952
6	H_0206 8	17	11473	12042	569	-1	*qmhA1*	Phosphoheptose isomerase 1		Q9PNE6
6	H_0206 9	17	12023	13459	1436	-1	*hldE*	Bifunctional protein HldE		P76658
6	H_0207 0	17	13460	14467	1007	-1	*hldD*	ADP-L-glycero-D-manno-heptose-6-epimerase		P67910
6	H_0207 1	17	14476	14772	296	-1	*gmhB*	D-glycero-alpha-D-manno-heptose-1%2C7-bisphosphate 7-phosphatase		Q9AGY5
6	H_0207 2	17	15091	15957	866	+1		Glycosyl transferases group 1		PF00534.14

Structure C (2)

2	S_0209 6	22	181	1299	1118	+1		hypothetical protein		
2	S_0209 7	22	1292	2551	1259	+1		O-Antigen ligase	**	PF04932.9
2	S_0209 8	22	2555	3181	626	-1	*lacA*	Galactoside O-acetyltransferase	*	P07464
2	S_0209 9	22	3191	4240	1049	-1	*cotSA*	Spore coat protein SA		P46915
2	S_0210 0	22	4353	5102	749	+1		hypothetical protein		

2	S_0210 1	22	5095	5817	722	+1	*pgaB*	Poly-beta-1%2C6-N-acetyl-D-glucosamine N-deacetylase precursor		P75906
2	S_0210 2	22	5814	6563	749	+1	*sunS*	SPBc2 prophage-derived glycosyltransferase SunS	*	O31986
2	S_0210 3	22	6560	7480	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
2	S_0210 4	22	7473	8534	1061	+1		hypothetical protein		
2	S_0210 5	22	8527	9471	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
2	S_0210 6	22	9484	11382	1898	-1		phosphoglycerol transferase I		PRK12363
2	S_0210 7	22	11470	12039	569	-1	*gmhA1*	Phosphoheptose isomerase 1		Q9PNE6
2	S_0210 8	22	12020	13456	1436	-1	*hldE*	Bifunctional protein HldE		P76658
2	S_0210 9	22	13457	14464	1007	-1	*hldD*	ADP-L-glycero-D-manno-heptose-6-epimerase		P67910
2	S_0211 0	22	14473	14841	368	-1	*gmhB*	D-glycero-alpha-D-manno-heptose-1%2C7-bisphosphate 7-phosphatase		Q9AGY5
2	S_0211 1	22	15088	15954	866	+1		Glycosyl transferases group 1		PF00534.14

Structure D (1,8,28)

1	R_0200 4	18	181	1299	1118	+1		hypothetical protein		
1	R_0200 5	18	1292	2551	1259	+1		O-Antigen ligase	**	PF04932.9
1	R_0200 6	18	2555	3181	626	-1	*lacA*	Galactoside O-acetyltransferase	*	P07464
1	R_0200 7	18	3191	4240	1049	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
1	R_0200 8	18	4353	5102	749	+1		hypothetical protein		
1	R_0200 9	18	5095	5817	722	+1	*pgaB*	Poly-beta-1%2C6-N-acetyl-D-glucosamine N-deacetylase precursor		P75906
1	R_0201 0	18	5814	6563	749	+1	*sunS*	SPBc2 prophage-derived glycosyltransferase SunS	*	O31986
1	R_0201 1	18	6560	7480	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
1	R_0201 2	18	7473	8534	1061	+1		hypothetical protein		
1	R_0201 3	18	8527	9471	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
1	R_0201 4	18	9484	11382	1898	-1		phosphoglycerol transferase I		PRK12363
1	R_0201 5	18	11470	12039	569	-1	*gmhA1*	Phosphoheptose isomerase 1		Q9PNE6
1	R_0201 6	18	12020	13456	1436	-1	*hldE*	Bifunctional protein HldE		P76658
1	R_0201 7	18	13457	14464	1007	-1	*hldD*	ADP-L-glycero-D-manno-heptose-6-epimerase		P67910
1	R_0201 8	18	14473	14759	296	-1	*gmhB*	D-glycero-alpha-D-manno-heptose-1%2C7-bisphosphate 7-phosphatase		Q9AGY5

1	R_0201 9	18	1508 8	1595 4	866	+1		Glycosyl transferases group 1		PF00534 .14
					Structure E (17)					
17	X_0068 9	6	5035 6	5052 3	167	+1		hypothetical protein		
17	X_0069 0	6	5053 4	5248 9	1955	+1	*tap*	Methyl-accepting chemotaxis protein IV		P07018
17	X_0069 1	6	5268 8	5347 3	785	+1		hypothetical protein		
17	X_0069 2	6	5347 7	5427 1	794	+1		hypothetical protein		
17	X_0069 3	6	5446 6	5890 5	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
17	X_0069 4	6	5891 1	6028 7	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
17	X_0069 5	6	6029 0	6103 9	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
17	X_0069 6	6	6108 3	6133 7	254	+1	*fdx*	Ferredoxin		P00208
17	X_0069 7	6	6133 7	6280 6	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
17	X_0069 8	6	6280 3	6380 7	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
17	X_0069 9	6	6380 0	6471 4	914	+1		Phosphatidylinositol mannoside acyltransferase		O06203
17	X_0070 0	6	6471 1	6548 4	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
17	X_0070 1	6	6548 6	6583 9	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
17	X_0070 2	6	6585 3	6785 3	2000	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
17	X_0070 3	6	6785 3	6815 2	299	+1	*rmlA*	Glucose-1-phosphate thymidylyltransferase		P26393
17	X_0070 4	6	6818 0	6951 7	1337	+1		O-Antigen ligase	**	PF04932 .9
17	X_0070 5	6	6951 4	7063 8	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
17	X_0070 6	6	7063 5	7122 5	590	+1		hypothetical protein		PRK10345
17	X_0070 7	6	7123 5	7224 8	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
17	X_0070 8	6	7227 0	7334 0	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
17	X_0070 9	6	7342 6	7416 9	743	+1		Glycosyl transferase family 2		PF00535 .20
17	X_0071 0	6	7415 9	7525 9	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
17	X_0071 1	6	7526 5	7672 2	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
17	X_0071 2	6	7673 1	7762 1	890	+1		hypothetical protein		
17	X_0071 3	6	7763 0	7818 7	557	+1		hypothetical protein		PRK10345
17	X_0071 4	6	7820 5	7985 1	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
17	X_0071 5	6	7998 6	8073 5	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723

17	X_0071 6	6	8073 5	8165 5	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
17	X_0071 7	6	8164 8	8271 2	1064	+1		hypothetical protein		
17	X_0071 8	6	8270 5	8364 6	941	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
17	X_0071 9	6	8363 8	8420 4	566	-1	*gmh A1*	Phosphoheptose isomerase 1		Q9PNE6
Structure F (9)										
9	T_0207 9	20	87	365	278	+1		hypothetical protein		
9	T_0208 0	20	375	1454	1079	+1		hypothetical protein		
9	T_0208 1	20	1479	2936	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
9	T_0208 2	20	2944	4014	1070	+1		hypothetical protein		
9	T_0208 3	20	3998	5266	1268	+1		O-Antigen ligase	**	PF04932.9
9	T_0208 4	20	5270	5674	404	+1	*fdtA*	TDP-4-oxo-6-deoxy-alpha-D-glucose-3%2C4-oxoisomerase		Q6T1W8
9	T_0208 5	20	5664	6170	506	+1	*fdtC*	dTDP-3-amino-3%2C6-dideoxy-alpha-D-galactopyranose 3-N-acetyltransferase		Q6T1W7
9	T_0208 6	20	6231	6839	608	+1	*lacA*	Galactoside O-acetyltransferase	*	P07464
9	T_0208 7	20	6836	7936	1100	+1	*fdtB*	dTDP-3-amino-3%2C6-dideoxy-alpha-D-galactopyranose transaminase		Q6T1W6
9	T_0208 8	20	7923	8750	827	+1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
9	T_0208 9	20	8756	9385	629	-1	*lacA*	Galactoside O-acetyltransferase	*	P07464
9	T_0209 0	20	9389	1044 1	1052	-1	*mfps A*	Mannosylfructose-phosphate synthase		A7TZT2
9	T_0209 1	20	1055 4	1130 3	749	+1		hypothetical protein		
9	T_0209 2	20	1129 6	1201 8	722	+1	*pgaB*	Poly-beta-1%2C6-N-acetyl-D-glucosamine N-deacetylase precursor		P75906
9	T_0209 3	20	1201 5	1276 4	749	+1	*sunS*	SPBc2 prophage-derived glycosyltransferase SunS	*	O31986
9	T_0209 4	20	1276 1	1368 1	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
9	T_0209 5	20	1367 4	1473 5	1061	+1		hypothetical protein		
9	T_0209 6	20	1472 8	1567 2	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
9	T_0209 7	20	1568 5	1758 3	1898	-1		phosphoglycerol transferase I		PRK12363
9	T_0209 8	20	1767 1	1824 0	569	-1	*gmh A1*	Phosphoheptose isomerase 1		Q9PNE6
Structure G (32)										
32	a_0140 9	7	1177 0	1320 6	1436	+1	*hldE*	Bifunctional protein HldE		P76658
32	a_0141 0	7	1318 7	1375 6	569	+1	*gmh A1*	Phosphoheptose isomerase 1		Q9PNE6
32	a_0141 1	7	1384 4	1574 2	1898	+1		phosphoglycerol transferase I		PRK12363
32	a_0141 2	7	1575 5	1669 9	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
32	a_0141 3	7	1669 2	1775 6	1064	-1		hypothetical protein		

32	a_0141 4	7	1775 7	1837 7	620	-1		hypothetical protein		PRK103 45
32	a_0141 5	7	1838 0	1929 7	917	-1		lipopolysaccharide core biosynthesis protein		PRK104 22
32	a_0141 6	7	1929 7	2004 6	749	-1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23
32	a_0141 7	7	2018 1	2182 7	1646	+1	*eptB*	Phosphoethanolamine transferase EptB		P37661
32	a_0141 8	7	2184 5	2240 2	557	-1		hypothetical protein		PRK103 45
32	a_0141 9	7	2241 7	2301 0	593	-1		hypothetical protein		PRK103 45
32	a_0142 0	7	2301 3	2411 3	1100	-1	*epsD*	Putative glycosyltransferase EpsD		P71053
32	a_0142 1	7	2410 3	2484 6	743	-1		Glycosyl transferase family 2		PF00535 .20
32	a_0142 2	7	2493 2	2600 2	1070	+1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
32	a_0142 3	7	2601 5	2707 9	1064	+1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
32	a_0142 4	7	2707 1	2839 3	1322	-1		O-Antigen ligase	**	PF04932 .9
32	a_0142 5	7	2839 5	2922 8	833	-1	*pssM*	Exopolysaccharide glucosyl ketal-pyruvate-transferase		Q9FCP2
32	a_0142 6	7	2924 3	3123 4	1991	-1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
32	a_0142 7	7	3124 8	3160 1	353	-1	*dgkA*	Diacylglycerol kinase		P0ABN1
32	a_0142 8	7	3160 3	3237 3	770	-1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK096 05
32	a_0142 9	7	3237 0	3328 4	914	-1		Phosphatidylinositol mannoside acyltransferase		A0QWG 5
32	a_0143 0	7	3327 7	3428 1	1004	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
32	a_0143 1	7	3427 8	3574 7	1469	-1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
32	a_0143 2	7	3574 7	3600 1	254	-1	*fdx*	Ferredoxin		P00208
32	a_0143 3	7	3604 5	3679 4	749	-1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK126 76
32	a_0143 4	7	3679 7	3817 3	1376	-1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
32	a_0143 5	7	3817 9	4261 8	4439	-1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
32	a_0143 6	7	4281 3	4360 7	794	-1		hypothetical protein		
32	a_0143 7	7	4360 0	4447 2	872	-1		pyrroloquinoline quinone biosynthesis protein PqqE		PRK053 01
32	a_0143 8	7	4459 5	4655 0	1955	-1	*tap*	Methyl-accepting chemotaxis protein IV		P07018

32	a_0143 9	7	4656 1	4672 8	167	-1		hypothetical protein		
					Structure H (3)					
3	F_0208 1	11	6731	7516	785	+1		hypothetical protein		
3	F_0208 2	11	7520	8314	794	+1		hypothetical protein		
3	F_0208 3	11	8509	1294 8	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
3	F_0208 4	11	1295 4	1433 0	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
3	F_0208 5	11	1433 3	1508 2	749	+1		bifunctional inositol-1 monophosphatase/fructo se-1%2C6- bisphosphatase		PRK126 76
3	F_0208 6	11	1512 6	1538 0	254	+1	*fdx*	Ferredoxin		P00208
3	F_0208 7	11	1538 0	1684 9	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
3	F_0208 8	11	1684 6	1785 0	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
3	F_0208 9	11	1784 3	1875 7	914	+1		Phosphatidylinositol mannoside acyltransferase		A0QWG 5
3	F_0209 0	11	1875 4	1952 7	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK096 05
3	F_0209 1	11	1952 9	1988 2	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
3	F_0209 2	11	1989 6	2189 9	2003	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
3	F_0209 3	11	2189 9	2278 0	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
3	F_0209 4	11	2277 0	2379 8	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
3	F_0209 5	11	2380 0	2487 3	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
3	F_0209 6	11	2488 5	2622 2	1337	+1		O-Antigen ligase	**	PF04932 .9
3	F_0209 7	11	2621 9	2734 3	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
3	F_0209 8	11	2734 0	2793 0	590	+1		hypothetical protein		PRK103 45
3	F_0209 9	11	2794 0	2895 3	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
3	F_0210 0	11	2897 6	3004 6	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
3	F_0210 1	11	3013 2	3087 5	743	+1		Glycosyl transferase family 2		PF00535 .20
3	F_0210 2	11	3086 5	3196 5	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
3	F_0210 3	11	3197 1	3342 8	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
3	F_0210 4	11	3343 7	3433 0	893	+1		hypothetical protein		
3	F_0210 5	11	3433 9	3489 6	557	+1		hypothetical protein		PRK103 45
3	F_0210 6	11	3491 4	3656 0	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
3	F_0210 7	11	3669 5	3744 4	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23

3	F_0210 8	11	3744 4	3836 4	920	+1		lipopolysaccharide core biosynthesis protein		PRK104 22
3	F_0210 9	11	3835 7	3941 8	1061	+1		hypothetical protein		
3	F_0211 0	11	3941 1	4034 9	938	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
3	F_0211 1	11	4036 8	4226 6	1898	-1		phosphoglycerol transferase I		PRK123 63

Structure I (10,11,13,23)

10	I_0201 5	13	2204 3	2285 8	815	+1		hypothetical protein		
10	I_0201 6	13	2285 1	2364 5	794	+1		hypothetical protein		
10	I_0201 7	13	2384 0	2827 9	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
10	I_0201 8	13	2828 5	2966 1	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
10	I_0201 9	13	2966 4	3041 3	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
10	I_0202 0	13	3045 7	3071 1	254	+1	*fdx*	Ferredoxin		P00208
10	I_0202 1	13	3071 1	3218 0	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
10	I_0202 2	13	3217 7	3318 1	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
10	I_0202 3	13	3317 4	3408 8	914	+1		lipid A biosynthesis lauroyl acyltransferase		PRK08419
10	I_0202 4	13	3408 5	3485 8	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
10	I_0202 5	13	3486 0	3521 3	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
10	I_0202 6	13	3522 7	3722 7	2000	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
10	I_0202 7	13	3722 7	3810 8	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
10	I_0202 8	13	3809 8	3912 6	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
10	I_0202 9	13	3912 8	4020 1	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
10	I_0203 0	13	4021 3	4155 0	1337	+1		O-Antigen ligase	**	PF04932.9
10	I_0203 1	13	4154 7	4267 1	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
10	I_0203 2	13	4266 8	4325 8	590	+1		hypothetical protein		PRK10345
10	I_0203 3	13	4326 8	4428 1	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
10	I_0203 4	13	4430 4	4537 4	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
10	I_0203 5	13	4546 0	4620 3	743	+1		Glycosyl transferase family 2		PF00535.20
10	I_0203 6	13	4619 3	4729 3	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
10	I_0203 7	13	4729 9	4875 6	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
10	I_0203 8	13	4876 5	4965 5	890	+1		hypothetical protein		

10	I_0203 9	13	4966 4	5022 1	557	+1		hypothetical protein		PRK10345
10	I_0204 0	13	5023 9	5188 5	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
10	I_0204 1	13	5202 0	5276 9	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
10	I_0204 2	13	5276 9	5368 9	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
10	I_0204 3	13	5368 2	5474 6	1064	+1		hypothetical protein		
10	I_0204 4	13	5473 9	5568 3	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
10	I_0204 5	13	5569 6	5759 4	1898	-1		phosphoglycerol transferase I		PRK12363

Structure J (5)

5	O_0164 2	9	4176 1	4255 8	797	+1		hypothetical protein		
5	O_0164 3	9	4255 1	4334 5	794	+1		hypothetical protein		
5	O_0164 4	9	4354 0	4797 9	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
5	O_0164 5	9	4798 5	4936 1	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
5	O_0164 6	9	4936 4	5011 3	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
5	O_0164 7	9	5015 7	5041 1	254	+1	*fdx*	Ferredoxin		P00208
5	O_0164 8	9	5041 1	5188 0	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
5	O_0164 9	9	5187 7	5288 1	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
5	O_0165 0	9	5287 4	5378 8	914	+1		Phosphatidylinositol mannoside acyltransferase		A0QWG5
5	O_0165 1	9	5378 5	5455 8	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
5	O_0165 2	9	5456 0	5491 3	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
5	O_0165 3	9	5492 7	5692 7	2000	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
5	O_0165 4	9	5692 7	5780 8	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
5	O_0165 5	9	5779 8	5882 6	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
5	O_0165 6	9	5882 8	5990 1	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
5	O_0165 7	9	5991 3	6125 0	1337	+1		O-Antigen ligase	**	PF04932.9
5	O_0165 8	9	6124 7	6237 1	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
5	O_0165 9	9	6236 8	6295 8	590	+1		hypothetical protein		PRK10345
5	O_0166 0	9	6296 8	6398 1	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
5	O_0166 1	9	6400 4	6507 4	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272

5	O_0166 2	9	6516 0	6590 3	743	+1		Glycosyl transferase family 2		PF00535 .20
5	O_0166 3	9	6589 3	6699 3	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
5	O_0166 4	9	6699 9	6845 6	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
5	O_0166 5	9	6846 5	6935 5	890	+1		hypothetical protein		
5	O_0166 6	9	6936 4	6992 1	557	+1		hypothetical protein		PRK10345
5	O_0166 7	9	6993 9	7158 5	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
5	O_0166 8	9	7172 0	7246 9	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
5	O_0166 9	9	7246 9	7338 9	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
5	O_0167 0	9	7338 2	7444 3	1061	+1		hypothetical protein		
5	O_0167 1	9	7443 6	7538 0	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
5	O_0167 2	9	7539 3	7729 1	1898	-1		phosphoglycerol transferase I		PRK12363

Structure K (24)

24	1_0198 0	16	288	1085	797	+1		hypothetical protein		
24	1_0198 1	16	1078	1872	794	+1		hypothetical protein		
24	1_0198 2	16	2067	6506	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
24	1_0198 3	16	6512	7888	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
24	1_0198 4	16	7891	8640	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
24	1_0198 5	16	8684	8938	254	+1	*fdx*	Ferredoxin		P00208
24	1_0198 6	16	8938	1040 7	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
24	1_0198 7	16	1040 4	1140 8	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
24	1_0198 8	16	1140 1	1231 5	914	+1	*msbB*	Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase	*	P24205
24	1_0198 9	16	1231 2	1308 5	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
24	1_0199 0	16	1308 7	1344 0	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
24	1_0199 1	16	1345 4	1546 0	2006	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
24	1_0199 2	16	1545 7	1633 8	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
24	1_0199 3	16	1632 8	1735 6	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
24	1_0199 4	16	1735 8	1843 1	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
24	1_0199 5	16	1844 3	1978 0	1337	+1		O-Antigen ligase	**	PF04932 .9
24	1_0199 6	16	1977 7	2090 1	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267

24	1_0199 7	16	2089 8	2148 8	590	+1		hypothetical protein		PRK10345
24	1_0199 8	16	2149 8	2251 1	1013	-1	*epsJ*	putative glycosyltransferase EpsJ N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	P71059
24	1_0199 9	16	2253 4	2360 4	1070	-1	*pglJ*		*	Q0P9C7
24	1_0200 0	16	2369 0	2443 3	743	+1		Glycosyl transferase family 2		PF00535.20
24	1_0200 1	16	2442 3	2552 3	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
24	1_0200 2	16	2552 9	2698 6	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
24	1_0200 3	16	2699 5	2788 8	893	+1		hypothetical protein		
24	1_0200 4	16	2789 7	2845 4	557	+1		hypothetical protein		PRK10345
24	1_0200 5	16	2847 2	3011 8	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
24	1_0200 6	16	3025 4	3100 3	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
24	1_0200 7	16	3100 3	3192 3	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
24	1_0200 8	16	3191 6	3297 7	1061	+1		hypothetical protein		
24	1_0200 9	16	3297 0	3391 4	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
24	1_0201 0	16	3392 7	3582 5	1898	-1	*ltaS2*	Lipoteichoic acid synthase 2		O34952

Structure L (25)

25	2_0083 4	4	8485 8	8564 3	785	+1		hypothetical protein		
25	2_0083 5	4	8564 7	8644 1	794	+1		hypothetical protein		
25	2_0083 6	4	8663 6	9107 5	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
25	2_0083 7	4	9108 1	9245 7	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
25	2_0083 8	4	9246 0	9320 9	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
25	2_0083 9	4	9325 3	9350 7	254	+1	*fdx*	Ferredoxin		P00208
25	2_0084 0	4	9350 7	9497 6	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
25	2_0084 1	4	9497 3	9597 7	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
25	2_0084 2	4	9597 0	9688 4	914	+1		Phosphatidylinositol mannoside acyltransferase		A0QWG5
25	2_0084 3	4	9688 1	9765 4	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
25	2_0084 4	4	9765 6	9800 9	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
25	2_0084 5	4	9802 3	1000 29	2006	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
25	2_0084 6	4	1000 26	1009 07	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
25	2_0084 7	4	1008 97	1019 25	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391

25	2_0084 8	4	1019 27	1030 00	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
25	2_0084 9	4	1030 12	1043 49	1337	+1		O-Antigen ligase	**	PF04932 .9
25	2_0085 0	4	1043 46	1054 70	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
25	2_0085 1	4	1054 67	1060 57	590	+1		hypothetical protein		PRK103 45
25	2_0085 2	4	1060 67	1070 80	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
25	2_0085 3	4	1071 03	1081 73	1070	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
25	2_0085 4	4	1082 59	1090 02	743	+1		Glycosyl transferase family 2		PF00535 .20
25	2_0085 5	4	1089 92	1100 92	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
25	2_0085 6	4	1100 98	1115 55	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
25	2_0085 7	4	1115 64	1124 57	893	+1		hypothetical protein		
25	2_0085 8	4	1124 66	1130 23	557	+1		hypothetical protein		PRK103 45
25	2_0085 9	4	1130 41	1146 87	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
25	2_0086 0	4	1148 22	1155 71	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23
25	2_0086 1	4	1155 71	1164 91	920	+1		lipopolysaccharide core biosynthesis protein		PRK104 22
25	2_0086 2	4	1164 84	1175 45	1061	+1		hypothetical protein		
25	2_0086 3	4	1175 38	1184 82	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
25	2_0086 4	4	1184 95	1203 93	1898	-1		phosphoglycerol transferase I		PRK123 63

Structure M (27)

27	4_0170 2	14	1384 6	1574 4	1898	+1	*ltaS2*	Lipoteichoic acid synthase 2		O34952
27	4_0170 3	14	1575 7	1670 1	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
27	4_0170 4	14	1669 4	1775 8	1064	-1		hypothetical protein		
27	4_0170 5	14	1775 1	1867 1	920	-1		lipopolysaccharide core biosynthesis protein		PRK104 22
27	4_0170 6	14	1867 1	1942 0	749	-1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23
27	4_0170 7	14	1955 5	2120 1	1646	+1	*eptB*	Phosphoethanolamine transferase EptB		P37661
27	4_0170 8	14	2121 9	2177 6	557	-1		hypothetical protein		PRK103 45
27	4_0170 9	14	2178 5	2267 5	890	-1		hypothetical protein		
27	4_0171 0	14	2268 4	2414 1	1457	-1	*patA*	Peptidoglycan O-acetyltransferase		O25526
27	4_0171 1	14	2414 7	2524 7	1100	-1	*epsD*	Putative glycosyltransferase EpsD		P71053

27	4_0171 2	14	2523 7	2598 0	743	-1		Glycosyl transferase family 2		PF00535 .20
27	4_0171 3	14	2606 6	2713 6	1070	+1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
27	4_0171 4	14	2715 9	2817 2	1013	+1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
27	4_0171 5	14	2818 2	2877 2	590	-1		hypothetical protein		PRK103 45
27	4_0171 6	14	2876 9	2989 3	1124	-1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
27	4_0171 7	14	2989 0	3122 7	1337	-1		O-Antigen ligase	**	PF04932 .9
27	4_0171 8	14	3123 9	3231 2	1073	-1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
27	4_0171 9	14	3231 4	3334 2	1028	-1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
27	4_0172 0	14	3333 2	3421 3	881	-1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
27	4_0172 1	14	3421 3	3621 6	2003	-1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
27	4_0172 2	14	3623 0	3658 3	353	-1	*dgkA*	Diacylglycerol kinase		P0ABN1
27	4_0172 3	14	3658 5	3735 8	773	-1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK096 05
27	4_0172 4	14	3735 5	3826 9	914	-1		Phosphatidylinositol mannoside acyltransferase		A0QWG 5
27	4_0172 5	14	3826 2	3926 6	1004	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
27	4_0172 6	14	3926 3	4073 2	1469	-1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
27	4_0172 7	14	4073 2	4098 6	254	-1	*fdx*	Ferredoxin		P00208
27	4_0172 8	14	4103 0	4177 9	749	-1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK126 76
27	4_0172 9	14	4178 2	4315 8	1376	-1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
27	4_0173 0	14	4316 4	4760 3	4439	-1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
27	4_0173 1	14	4779 8	4859 2	794	-1		hypothetical protein		
27	4_0173 2	14	4859 6	4938 1	785	-1		hypothetical protein		

Structure N (29)

29	6_0207 4	13	1384 7	1574 5	1898	+1		phosphoglycerol transferase I		PRK123 63
29	6_0207 5	13	1575 8	1670 2	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
29	6_0207 6	13	1669 5	1775 6	1061	-1		hypothetical protein		
29	6_0207 7	13	1774 9	1866 9	920	-1		lipopolysaccharide core biosynthesis protein		PRK104 22
29	6_0207 8	13	1866 9	1941 8	749	-1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23

29	6_0207 9	13	1955 3	2119 9	1646	+1	*eptB*	Phosphoethanolamine transferase EptB		P37661
29	6_0208 0	13	2121 6	2177 3	557	-1		hypothetical protein		PRK10345
29	6_0208 1	13	2178 2	2267 2	890	-1		hypothetical protein		
29	6_0208 2	13	2268 1	2413 8	1457	-1	*patA*	Peptidoglycan O-acetyltransferase		O25526
29	6_0208 3	13	2414 4	2524 4	1100	-1	*epsD*	Putative glycosyltransferase EpsD		P71053
29	6_0208 4	13	2523 4	2597 7	743	-1		Glycosyl transferase family 2		PF00535.20
29	6_0208 5	13	2606 3	2713 3	1070	+1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
29	6_0208 6	13	2715 6	2816 9	1013	+1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
29	6_0208 7	13	2817 9	2876 9	590	-1		hypothetical protein		PRK10345
29	6_0208 8	13	2876 6	2989 0	1124	-1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
29	6_0208 9	13	2988 7	3122 4	1337	-1		O-Antigen ligase	**	PF04932.9
29	6_0209 0	13	3123 6	3230 9	1073	-1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
29	6_0209 1	13	3231 1	3333 9	1028	-1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
29	6_0209 2	13	3332 9	3421 0	881	-1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
29	6_0209 3	13	3421 0	3621 0	2000	-1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
29	6_0209 4	13	3622 4	3657 7	353	-1	*dgkA*	Diacylglycerol kinase		P0ABN1
29	6_0209 5	13	3657 9	3735 2	773	-1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
29	6_0209 6	13	3734 9	3826 3	914	-1	*msbB*	Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase	*	P24205
29	6_0209 7	13	3825 6	3926 0	1004	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
29	6_0209 8	13	3925 7	4072 6	1469	-1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
29	6_0209 9	13	4072 6	4098 0	254	-1	*fdx*	Ferredoxin		P00208
29	6_0210 0	13	4102 4	4177 3	749	-1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
29	6_0210 1	13	4177 6	4315 2	1376	-1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
29	6_0210 2	13	4315 8	4759 7	4439	-1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
29	6_0210 3	13	4779 2	4858 6	794	-1		hypothetical protein		
29	6_0210 4	13	4859 0	4937 5	785	-1		hypothetical protein		

Structure O (30)

30	7_0149 9	9	1384 4	1574 2	1898	+1		phosphoglycerol transferase I		PRK12363
30	7_0150 0	9	1575 5	1669 9	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692

30	7_0150 1	9	1669 2	1775 6	1064	-1		hypothetical protein		
30	7_0150 2	9	1774 9	1866 9	920	-1		lipopolysaccharide core biosynthesis protein		PRK10422
30	7_0150 3	9	1866 9	1941 8	749	-1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
30	7_0150 4	9	1955 3	2119 9	1646	+1	*eptB*	Phosphoethanolamine transferase EptB		P37661
30	7_0150 5	9	2121 7	2177 4	557	-1		hypothetical protein		PRK10345
30	7_0150 6	9	2178 3	2267 3	890	-1		hypothetical protein		
30	7_0150 7	9	2268 2	2413 9	1457	-1	*patA*	Peptidoglycan O-acetyltransferase		O25526
30	7_0150 8	9	2414 5	2524 5	1100	-1	*epsD*	Putative glycosyltransferase EpsD		P71053
30	7_0150 9	9	2523 5	2597 8	743	-1		Glycosyl transferase family 2		PF00535.20
30	7_0151 0	9	2606 4	2713 4	1070	+1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
30	7_0151 1	9	2715 7	2817 0	1013	+1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
30	7_0151 2	9	2818 0	2877 0	590	-1		hypothetical protein		PRK10345
30	7_0151 3	9	2876 7	2989 1	1124	-1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
30	7_0151 4	9	2988 8	3122 5	1337	-1		O-Antigen ligase	**	PF04932.9
30	7_0151 5	9	3123 7	3231 0	1073	-1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
30	7_0151 6	9	3231 2	3334 0	1028	-1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
30	7_0151 7	9	3333 0	3421 1	881	-1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
30	7_0151 8	9	3421 1	3621 1	2000	-1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
30	7_0151 9	9	3622 5	3657 8	353	-1	*dgkA*	Diacylglycerol kinase		P0ABN1
30	7_0152 0	9	3658 0	3735 3	773	-1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
30	7_0152 1	9	3735 0	3826 4	914	-1		Phosphatidylinositol mannoside acyltransferase		O06203
30	7_0152 2	9	3825 7	3926 1	1004	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
30	7_0152 3	9	3925 8	4072 7	1469	-1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
30	7_0152 4	9	4072 7	4098 1	254	-1	*fdx*	Ferredoxin		P00208
30	7_0152 5	9	4102 5	4177 4	749	-1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
30	7_0152 6	9	4177 7	4315 3	1376	-1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
30	7_0152 7	9	4315 9	4759 8	4439	-1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
30	7_0152 8	9	4779 3	4858 7	794	-1		hypothetical protein		
30	7_0152 9	9	4859 1	4937 6	785	-1		hypothetical protein		

Structure P (31)

31	8_01445	7	7424 3	7502 8	785	+1		hypothetical protein		
31	8_01446	7	7503 2	7582 6	794	+1		hypothetical protein		
31	8_01447	7	7602 1	8046 0	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
31	8_01448	7	8046 6	8184 2	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
31	8_01449	7	8184 5	8259 4	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
31	8_01450	7	8263 8	8289 2	254	+1	*fdx*	Ferredoxin		P00208
31	8_01451	7	8289 2	8436 1	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
31	8_01452	7	8435 8	8536 2	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
31	8_01453	7	8535 5	8626 9	914	+1		Phosphatidylinositol mannoside acyltransferase		A0QWG5
31	8_01454	7	8626 6	8703 9	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
31	8_01455	7	8704 1	8739 4	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
31	8_01456	7	8740 8	8941 1	2003	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
31	8_01457	7	8941 1	9029 2	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
31	8_01458	7	9028 2	9131 0	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
31	8_01459	7	9131 2	9238 5	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
31	8_01460	7	9239 7	9373 4	1337	+1		O-Antigen ligase	**	PF04932.9
31	8_01461	7	9373 1	9485 5	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
31	8_01462	7	9485 2	9544 2	590	+1		hypothetical protein		PRK10345
31	8_01463	7	9545 2	9646 5	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
31	8_01464	7	9648 8	9755 8	1070	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
31	8_01465	7	9764 4	9838 7	743	+1		Glycosyl transferase family 2		PF00535.20
31	8_01466	7	9837 7	9947 7	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
31	8_01467	7	9948 3	1009 40	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
31	8_01468	7	1009 49	1018 39	890	+1		hypothetical protein		
31	8_01469	7	1018 48	1024 05	557	+1		hypothetical protein		PRK10345
31	8_01470	7	1024 23	1040 69	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
31	8_01471	7	1042 04	1049 53	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723

	8_0147		1049	1058				lipopolysaccharide core		PRK104
31	2	7	53	73	920	+1		biosynthesis protein		22
	8_0147		1058	1069						
31	3	7	66	30	1064	+1		hypothetical protein		
	8_0147		1068	1078				ADP-heptose--LPS		
31	4	7	96	67	971	+1	*rfaF*	heptosyltransferase 2	*	P37692
	8_0147		1078	1097				Lipoteichoic acid		
31	5	7	80	78	1898	-1	*ltaS2*	synthase 2		O34952

Structure Q (16)

	W_018		1697	1775						
16	18	22	3	8	785	+1		hypothetical protein		
	W_018		1776	1855						
16	19	22	2	6	794	+1		hypothetical protein		
								Glutamate synthase		
	W_018		1875	2319				[NADPH] large chain		
16	20	22	1	0	4439	+1	*gltB*	precursor		P09831
	W_018		2319	2457				Glutamate synthase		
16	21	22	6	2	1376	+1	*gltD*	[NADPH] small chain		P09832
								bifunctional inositol-1		
								monophosphatase/fructo		
	W_018		2457	2532				se-1%2C6-		PRK126
16	22	22	5	4	749	+1		bisphosphatase		76
	W_018		2536	2562						
16	23	22	8	2	254	+1	*fdx*	Ferredoxin		P00208
								Guanosine-5'-		
								triphosphate%2C3'-		
	W_018		2562	2709				diphosphate		
16	24	22	2	1	1469	+1	*gppA*	pyrophosphatase		P25552
	W_018		2708	2809				Lipopolysaccharide		
16	25	22	8	2	1004	+1	*rfaC*	heptosyltransferase 1	*	P24173
	W_018		2808	2899				lipid A biosynthesis		PRK084
16	26	22	5	9	914	+1		lauroyl acyltransferase		19
	W_018		2899	2977						
16	27	22	6	8	782	+1		hypothetical protein		
	W_018		2976	3012						
16	28	22	8	1	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
	W_018		3013	3213				Lipoteichoic acid		
16	29	22	5	5	2000	+1	*ltaS1*	synthase 1		Q797B3
	W_018		3213	3311						
16	30	22	7	7	980	+1		hypothetical protein		
	W_018		3312	3444				GlcNAc-PI de-N-		PF02585
16	31	22	1	3	1322	+1		acetylase		.11
								N-acetylgalactosamine-		
								N%2CN'-		
								diacetylbacillosaminyl-		
								diphospho-undecaprenol		
								4-alpha-N-		
	W_018		3444	3551				acetylgalactosaminyltra		
16	32	22	0	3	1073	+1	*pglJ*	nsferase	*	Q0P9C7
	W_018		3551	3675						PF04932
16	33	22	0	7	1247	+1		O-Antigen ligase	**	.9
	W_018		3679	3692						
16	34	22	4	2	128	+1		hypothetical protein		
								N%2CN'-		
	W_018		3697	3800				diacetyllegionaminic		
16	35	22	2	9	1037	+1	*neuB*	acid synthase		Q5ZXH9

Structure R (12,20)

								pyrroloquinoline		
	M_019		4019	4106				quinone biosynthesis		PRK053
12	46	8	1	3	872	+1		protein PqqE		01
	M_019		4105	4185						
12	47	8	6	0	794	+1		hypothetical protein		
								Glutamate synthase		
	M_019		4204	4648				[NADPH] large chain		
12	48	8	5	4	4439	+1	*gltB*	precursor		P09831
	M_019		4649	4786				Glutamate synthase		
12	49	8	0	6	1376	+1	*gltD*	[NADPH] small chain		P09832
								bifunctional inositol-1		
								monophosphatase/fructo		
	M_019		4786	4861				se-1%2C6-		PRK126
12	50	8	9	8	749	+1		bisphosphatase		76
	M_019		4866	4891						
12	51	8	2	6	254	+1	*fdx*	Ferredoxin		P00208

12	M_0195	8	4891	5038	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
12	M_0195	8	5038	5138	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
12	M_0195	8	5137	5229	914	+1		Phosphatidylinositol mannoside acyltransferase		A0QWG5
12	M_0195	8	5229	5306	773	+1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
12	M_0195	8	5306	5341	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
12	M_0195	8	5343	5543	2000	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
12	M_0195	8	5543	5631	881	+1	*rmlA2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
12	M_0195	8	5630	5733	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
12	M_0195	8	5733	5840	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
12	M_0195	8	5841	5975	1337	+1		O-Antigen ligase	**	PF04932.9
12	M_0195	8	5975	6087	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
12	M_0195	8	6087	6146	590	+1		hypothetical protein		PRK10345
12	M_0195	8	6147	6248	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
12	M_0195	8	6250	6357	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
12	M_0195	8	6366	6440	743	+1		Glycosyl transferase family 2		PF00535.20
12	M_0195	8	6439	6549	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
12	M_0195	8	6550	6696	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
12	M_0195	8	6697	6786	890	+1		hypothetical protein		
12	M_0195	8	6786	6842	557	+1		hypothetical protein		PRK10345
12	M_0195	8	6844	7009	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
12	M_0195	8	7022	7097	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
12	M_0195	8	7097	7189	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
12	M_0195	8	7188	7295	1064	+1		hypothetical protein		
12	M_0195	8	7294	7388	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
12	M_0195	8	7390	7579	1898	-1		phosphoglycerol transferase I		PRK12363

Structure S (22)

22	D_01312	5	1391	1581	1898	+1		phosphoglycerol transferase I		PRK12363
22	D_01313	5	1582	1676	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692

22	D_0131 4	5	1676 1	1782 5	1064	-1		hypothetical protein		
22	D_0131 5	5	1781 8	1873 8	920	-1		lipopolysaccharide core biosynthesis protein		PRK10422
22	D_0131 6	5	1873 8	1948 7	749	-1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
22	D_0131 7	5	1962 2	2126 8	1646	+1	*eptB*	Phosphoethanolamine transferase EptB		P37661
22	D_0131 8	5	2128 6	2184 3	557	-1		hypothetical protein		PRK10345
22	D_0131 9	5	2185 2	2274 2	890	-1		hypothetical protein		
22	D_0132 0	5	2275 1	2420 8	1457	-1	*patA*	Peptidoglycan O-acetyltransferase		O25526
22	D_0132 1	5	2421 4	2531 4	1100	-1	*epsD*	Putative glycosyltransferase EpsD		P71053
22	D_0132 2	5	2530 4	2604 7	743	-1		Glycosyl transferase family 2		PF00535.20
22	D_0132 3	5	2613 3	2720 3	1070	+1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
22	D_0132 4	5	2722 6	2823 9	1013	+1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
22	D_0132 5	5	2824 9	2883 9	590	-1		hypothetical protein		PRK10345
22	D_0132 6	5	2883 6	2996 0	1124	-1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
22	D_0132 7	5	2995 7	3129 4	1337	-1		O-Antigen ligase	**	PF04932.9
22	D_0132 8	5	3130 6	3237 9	1073	-1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
22	D_0132 9	5	3237 6	3341 0	1034	-1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
22	D_0133 0	5	3340 0	3428 1	881	-1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
22	D_0133 1	5	3428 1	3628 4	2003	-1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
22	D_0133 2	5	3629 8	3665 1	353	-1	*dgkA*	Diacylglycerol kinase		P0ABN1
22	D_0133 3	5	3665 3	3742 6	773	-1		bifunctional UGMP family protein/serine/threonine protein kinase		PRK09605
22	D_0133 4	5	3742 3	3833 7	914	-1		Phosphatidylinositol mannoside acyltransferase		A0QWG5
22	D_0133 5	5	3833 0	3933 4	1004	-1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
22	D_0133 6	5	3933 1	4080 0	1469	-1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
22	D_0133 7	5	4080 0	4105 4	254	-1	*fdx*	Ferredoxin		P00208
22	D_0133 8	5	4109 8	4184 7	749	-1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
22	D_0133 9	5	4185 0	4322 6	1376	-1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
22	D_0134 0	5	4323 2	4767 1	4439	-1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831

22	D_0134 1	5	4786 6	4866 0	794	-1		hypothetical protein		
22	D_0134 2	5	4865 3	4952 5	872	-1		pyrroloquinoline quinone biosynthesis protein PqqE		PRK053 01

Structure T (3)

3	F_0170 8	7	3818 1	3935 3	1172	+1	*tilS*	tRNA(Ile)-lysidine synthase		
3	F_0170 9	7	3935 0	3996 7	617	+1	*hisH 1*	Imidazole glycerol phosphate synthase subunit HisH 1		Q0P8U2
3	F_0171 0	7	3997 0	4073 1	761	+1	*hisF*	Imidazole glycerol phosphate synthase subunit HisF		C3MBC 1
3	F_0171 1	7	4072 8	4224 2	1514	+1		hypothetical protein		
3	F_0171 2	7	4224 2	4341 7	1175	+1		Glycosyl transferases group 1		PF00534 .14
3	F_0171 3	7	4350 2	4520 2	1700	+1	*hepA*	Heterocyst differentiation ATP-binding protein HepA		P22638
3	F_0171 4	7	4520 6	4721 2	2006	+1	*oatA*	O-acetyltransferase OatA		Q7A3D6
3	F_0171 5	7	4720 9	4833 9	1130	+1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
3	F_0171 6	7	4832 7	5022 5	1898	+1	*asnB*	Asparagine synthetase [glutamine-hydrolyzing] 1		P54420
3	F_0171 7	7	5021 8	5132 4	1106	+1	*pglA*	N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol alpha-1%2C3-N-acetylgalactosaminyltransferase		Q0P9C9
3	F_0171 8	7	5138 3	5215 3	770	+1	*tagO*	putative undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase		O34753
3	F_0171 9	7	5240 2	5364 0	1238	+1	*wzxC*	Lipopolysaccharide biosynthesis protein WzxC		P77377
3	F_0172 0	7	5364 0	5476 7	1127	+1	*tilS*	tRNA(Ile)-lysidine synthase		
3	F_0172 1	7	5476 4	5529 7	533	+1	*vatD*	Streptogramin A acetyltransferase		P50870
3	F_0172 2	7	5529 4	5643 9	1145	+1		hypothetical protein		
3	F_0172 3	7	5643 2	5763 4	1202	+1		O-Antigen ligase	**	PF04932 .9
3	F_0172 4	7	5764 0	5880 3	1163	+1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
3	F_0172 5	7	5883 2	5993 8	1106	+1		hypothetical protein		
3	F_0172 6	7	5994 7	6055 2	605	+1	*hisH 1*	Imidazole glycerol phosphate synthase subunit HisH 1		Q0P8U2
3	F_0172 7	7	6055 2	6131 9	767	+1	*hisF*	Imidazole glycerol phosphate synthase subunit HisF		C3MBC 1
3	F_0172 8	7	6131 9	6242 8	1109	+1		Glycosyl transferases group 1		PF00534 .14
3	F_0172 9	7	6242 5	6349 5	1070	+1	*wbpI*	UDP-2%2C3-diacetamido-2%2C3-dideoxy-D-glucuronate 2-epimerase		G3XD61

3	F_0173 0	7	6348 8	6448 6	998	+1	*pgaC*	Poly-beta-1%2C6-N-acetyl-D-glucosamine synthase		P75905
3	F_0173 1	7	6448 3	6544 5	962	+1	*tagO*	putative undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase		O34753
3	F_0173 2	7	6544 5	6601 7	572	+1	*rmlC*	dTDP-4-dehydrorhamnose 3%2C5-epimerase		Q9HU21
3	F_0173 3	7	6602 7	6776 9	1742	+1	*pglF*	UDP-N-acetyl-alpha-D-glucosamine C6 dehydratase		Q0P9D4
3	F_0173 4	7	6783 1	6822 9	398	+1		hypothetical protein		
3	F_0173 5	7	6823 3	6908 7	854	-1	*ligA*	DNA ligase		P44121
3	F_0173 6	7	6915 2	6970 9	557	+1	*apt*	Adenine phosphoribosyltransferase		P69503
3	F_0173 7	7	6972 6	7093 4	1208	+1	*trpB*	Tryptophan synthase beta chain		Q81TL8
3	F_0173 8	7	7095 2	7165 3	701	+1	*yohD*	Inner membrane protein YohD		P33366

Structure U (10,11,13,23)

10	I_0091 5	4	1263 09	1270 10	701	-1	*yohD*	Inner membrane protein YohD		P33366
10	I_0091 6	4	1270 28	1282 36	1208	-1	*trpB*	Tryptophan synthase beta chain		Q81TL8
10	I_0091 7	4	1282 53	1288 10	557	-1	*apt*	Adenine phosphoribosyltransferase		P69503
10	I_0091 8	4	1288 75	1297 29	854	+1	*ligA*	DNA ligase		P44121
10	I_0091 9	4	1297 33	1301 31	398	-1		hypothetical protein		
10	I_0092 0	4	1301 93	1319 35	1742	-1	*pglF*	UDP-N-acetyl-alpha-D-glucosamine C6 dehydratase		Q0P9D4
10	I_0092 1	4	1319 45	1325 17	572	-1	*rmlC*	dTDP-4-dehydrorhamnose 3%2C5-epimerase		Q9HU21
10	I_0092 2	4	1325 17	1334 79	962	-1	*tagO*	putative undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase		O34753
10	I_0092 3	4	1334 76	1346 75	1199	-1		putative glycosyl transferase		PRK10307
10	I_0092 4	4	1346 68	1357 38	1070	-1	*wbpI*	UDP-2%2C3-diacetamido-2%2C3-dideoxy-D-glucuronate 2-epimerase		G3XD61
10	I_0092 5	4	1357 35	1368 47	1112	-1		beta-1%2C6-galactofuranosyltransferase		PRK09814
10	I_0092 6	4	1368 37	1376 13	776	-1	*hisF*	Imidazole glycerol phosphate synthase subunit HisF		C3MBC1
10	I_0092 7	4	1376 13	1382 18	605	-1	*hisH 1*	Imidazole glycerol phosphate synthase subunit HisH 1		Q0P8U2
10	I_0092 8	4	1382 27	1393 33	1106	-1		hypothetical protein		
10	I_0092 9	4	1393 62	1405 25	1163	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
10	I_0093 0	4	1405 31	1417 33	1202	-1		O-Antigen ligase	**	PF04932.9
10	I_0093 1	4	1417 26	1428 71	1145	-1		hypothetical protein		
10	I_0093 2	4	1428 68	1434 01	533	-1	*vatD*	Streptogramin A acetyltransferase		P50870
10	I_0093 3	4	1433 98	1445 01	1103	-1	*tilS*	tRNA(Ile)-lysidine synthase		

10	I_00934	4	144503	145108	605	-1		Putative acetyltransferase		Q5HCZ5
10	I_00935	4	145105	146373	1268	-1	*tuaB*	Teichuronic acid biosynthesis protein TuaB		O32273
10	I_00936	4	146370	147365	995	-1	*tagO*	putative undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase		O34753
10	I_00937	4	147424	148530	1106	-1	*pglA*	N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol alpha-1%2C3-N-acetylgalactosaminyltransferase		Q0P9C9
10	I_00938	4	148523	149650	1127	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
10	I_00939	4	149647	151653	2006	-1	*oatA*	O-acetyltransferase OatA		Q7A3D6
10	I_00940	4	151657	153360	1703	-1	*hepA*	Heterocyst differentiation ATP-binding protein HepA		P22638
10	I_00941	4	153353	153559	206	-1		hypothetical protein		
10	I_00942	4	153569	154705	1136	-1		hypothetical protein		
10	I_00943	4	154702	155463	761	-1	*hisF*	Imidazole glycerol phosphate synthase subunit HisF		C3MBC1
10	I_00944	4	155466	156083	617	-1	*hisH 1*	Imidazole glycerol phosphate synthase subunit HisH 1		Q0P8U2
10	I_00945	4	156080	157240	1160	-1	*tilS*	tRNA(Ile)-lysidine synthase		

Structure V (18,19,21)

18	Y_02045	22	147	1523	1376	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
18	Y_02046	22	1523	2404	881	+1	*rmlA 2*	Glucose-1-phosphate thymidylyltransferase 2		P61887
18	Y_02047	22	2394	3422	1028	+1	*rfbB*	dTDP-glucose 4%2C6-dehydratase	*	P26391
18	Y_02048	22	3424	4497	1073	+1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3
18	Y_02049	22	4509	5846	1337	+1		O-Antigen ligase	**	PF04932.9
18	Y_02050	22	5843	6967	1124	+1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
18	Y_02051	22	6964	7554	590	+1		hypothetical protein		PRK10345
18	Y_02052	22	7563	8576	1013	-1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
18	Y_02053	22	8599	9669	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
18	Y_02054	22	9755	10498	743	+1		Glycosyl transferase family 2		PF00535.20
18	Y_02055	22	10488	11588	1100	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
18	Y_02056	22	11594	13051	1457	+1	*patA*	Peptidoglycan O-acetyltransferase		O25526
18	Y_02057	22	13060	13950	890	+1		hypothetical protein		

18	Y_0205 8	22	1395 9	1451 6	557	+1		hypothetical protein		PRK10345
18	Y_0205 9	22	1453 4	1618 0	1646	-1	*eptB*	Phosphoethanolamine transferase EptB		P37661
18	Y_0206 0	22	1631 5	1706 4	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
18	Y_0206 1	22	1706 4	1798 4	920	+1		lipopolysaccharide core biosynthesis protein		PRK10422
18	Y_0206 2	22	1797 7	1904 1	1064	+1		hypothetical protein		
18	Y_0206 3	22	1903 4	1997 8	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
18	Y_0206 4	22	1999 1	2188 9	1898	-1		phosphoglycerol transferase I		PRK12363

Structure W (7)

7	Q_0139 5	7	6076 3	6133 5	572	+1	*yigZ*	IMPACT family member YigZ		P27862
7	Q_0139 6	7	6136 5	6280 1	1436	+1	*xcpQ*	Type II secretion system protein D precursor		P35818
7	Q_0139 7	7	6331 5	6526 7	1952	+1	*tap*	Methyl-accepting chemotaxis protein IV		P07018
7	Q_0139 8	7	6546 6	6626 3	797	+1		hypothetical protein		
7	Q_0139 9	7	6625 6	6705 0	794	+1		hypothetical protein		
7	Q_0140 0	7	6724 5	7168 4	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
7	Q_0140 1	7	7169 0	7306 6	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
7	Q_0140 2	7	7306 9	7381 8	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
7	Q_0140 3	7	7386 2	7411 6	254	+1	*fdx*	Ferredoxin		P00208
7	Q_0140 4	7	7411 6	7558 5	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
7	Q_0140 5	7	7558 2	7658 6	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
7	Q_0140 6	7	7657 9	7749 3	914	+1		Phosphatidylinositol mannoside acyltransferase		O06203
7	Q_0140 7	7	7749 0	7826 0	770	+1		hypothetical protein		
7	Q_0140 8	7	7826 2	7861 5	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
7	Q_0140 9	7	7862 9	8062 0	1991	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
7	Q_0141 0	7	8064 0	8184 8	1208	+1		O-Antigen ligase	**	PF04932.9
7	Q_0141 1	7	8184 1	8291 7	1076	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase	*	Q0P9C7
7	Q_0141 2	7	8293 3	8400 3	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
7	Q_0141 3	7	8408 9	8483 2	743	+1		Glycosyl transferase family 2		PF00535.20
7	Q_0141 4	7	8482 2	8591 6	1094	+1	*epsD*	Putative glycosyltransferase EpsD		P71053

7	Q_0141 5	7	8591 3	8698 0	1067	-1		lipopolysaccharide core biosynthesis protein		PRK10422
7	Q_0141 6	7	8709 0	8768 3	593	+1		hypothetical protein		PRK10345
7	Q_0141 7	7	8772 3	8847 2	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK01723
7	Q_0141 8	7	8847 2	8938 9	917	+1		lipopolysaccharide core biosynthesis protein		PRK10422
7	Q_0141 9	7	8939 2	9001 2	620	+1		hypothetical protein		PRK10345
7	Q_0142 0	7	9001 3	9107 7	1064	+1		hypothetical protein		
7	Q_0142 1	7	9107 0	9201 4	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
7	Q_0142 2	7	9202 7	9392 5	1898	-1		phosphoglycerol transferase I		PRK12363
7	Q_0142 3	7	9401 3	9458 2	569	-1	*gmh A1*	Phosphoheptose isomerase 1		Q9PNE6
7	Q_0142 4	7	9456 3	9599 9	1436	-1	*hldE*	Bifunctional protein HldE		P76658
7	Q_0142 5	7	9600 0	9700 7	1007	-1	*hldD*	ADP-L-glycero-D-manno-heptose-6-epimerase		P67910

Structure Y (4)

4	G_0181 4	9	3964 9	4022 1	572	+1	*yigZ*	IMPACT family member YigZ		P27862
4	G_0181 5	9	4025 1	4168 7	1436	+1	*xcpQ*	Type II secretion system protein D precursor		P35818
4	G_0181 6	9	4219 1	4414 3	1952	+1	*tar*	Methyl-accepting chemotaxis protein II		P07017
4	G_0181 7	9	4426 6	4513 8	872	+1		pyrroloquinoline quinone biosynthesis protein PqqE		PRK05301
4	G_0181 8	9	4513 1	4592 5	794	+1		hypothetical protein		
4	G_0181 9	9	4612 0	5055 9	4439	+1	*gltB*	Glutamate synthase [NADPH] large chain precursor		P09831
4	G_0182 0	9	5056 5	5194 1	1376	+1	*gltD*	Glutamate synthase [NADPH] small chain		P09832
4	G_0182 1	9	5194 4	5269 3	749	+1		bifunctional inositol-1 monophosphatase/fructose-1%2C6-bisphosphatase		PRK12676
4	G_0182 2	9	5273 7	5299 1	254	+1	*fdx*	Ferredoxin		P00208
4	G_0182 3	9	5299 1	5446 0	1469	+1	*gppA*	Guanosine-5'-triphosphate%2C3'-diphosphate pyrophosphatase		P25552
4	G_0182 4	9	5445 7	5546 1	1004	+1	*rfaC*	Lipopolysaccharide heptosyltransferase 1	*	P24173
4	G_0182 5	9	5545 4	5636 8	914	+1		Phosphatidylinositol mannoside acyltransferase		O06203
4	G_0182 6	9	5636 5	5713 5	770	+1		hypothetical protein		
4	G_0182 7	9	5713 7	5749 0	353	+1	*dgkA*	Diacylglycerol kinase		P0ABN1
4	G_0182 8	9	5750 4	5949 5	1991	+1	*ltaS1*	Lipoteichoic acid synthase 1		Q797B3
4	G_0182 9	9	5951 5	6072 3	1208	+1		O-Antigen ligase	**	PF04932.9
4	G_0183 0	9	6071 6	6179 2	1076	-1	*pglJ*	N-acetylgalactosamine-N%2CN'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-	*	Q0P9C7

								acetylgalactosaminyltransferase		
4	G_0183 1	9	6180 8	6287 8	1070	-1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
4	G_0183 2	9	6296 4	6370 7	743	+1		Glycosyl transferase family 2		PF00535 .20
4	G_0183 3	9	6369 7	6479 1	1094	+1	*epsD*	Putative glycosyltransferase EpsD		P71053
4	G_0183 4	9	6478 8	6585 5	1067	-1		lipopolysaccharide core biosynthesis protein		PRK104 22
4	G_0183 5	9	6596 5	6655 8	593	+1		hypothetical protein		PRK103 45
4	G_0183 6	9	6659 8	6734 7	749	+1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23
4	G_0183 7	9	6734 7	6826 4	917	+1		lipopolysaccharide core biosynthesis protein		PRK104 22
4	G_0183 8	9	6826 7	6888 7	620	+1		hypothetical protein		PRK103 45
4	G_0183 9	9	6888 8	6995 2	1064	+1		hypothetical protein		
4	G_0184 0	9	6994 5	7088 9	944	+1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
4	G_0184 1	9	7090 2	7280 0	1898	-1		phosphoglycerol transferase I		PRK123 63
4	G_0184 2	9	7288 8	7345 7	569	-1	*gmhA1*	Phosphoheptose isomerase 1		Q9PNE6
4	G_0184 3	9	7343 8	7487 4	1436	-1	*hldE*	Bifunctional protein HldE		P76658
4	G_0184 4	9	7487 5	7588 2	1007	-1	*hldD*	ADP-L-glycero-D-manno-heptose-6-epimerase		P67910

Structure Z (26)

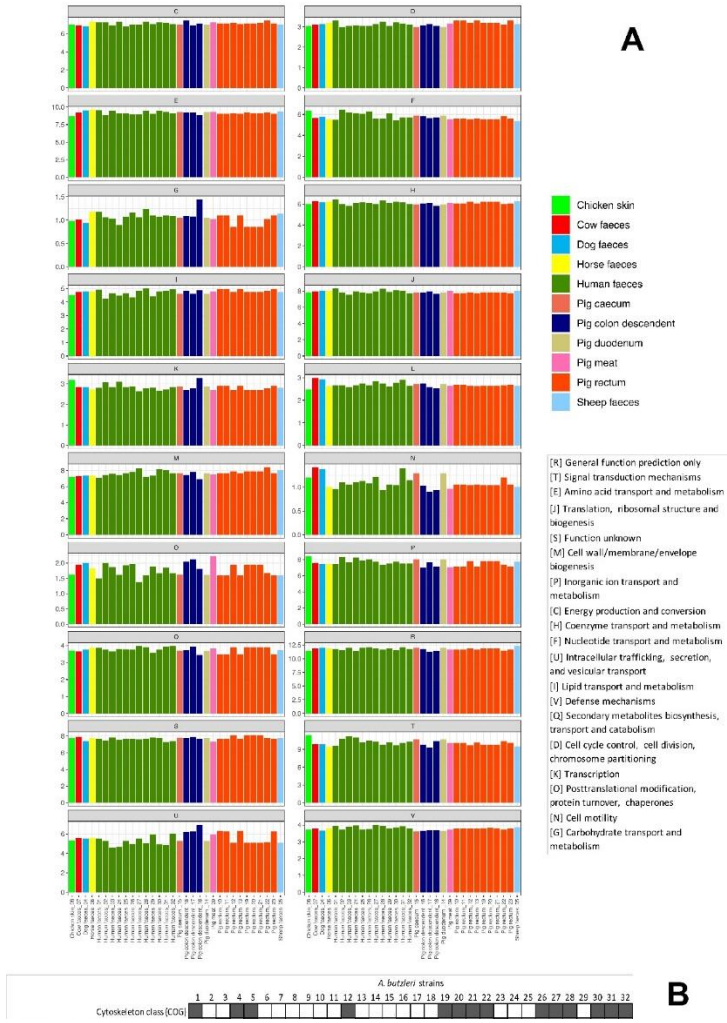
26	3_0217 0	18	1388 9	1578 7	1898	+1		phosphoglycerol transferase I		PRK123 63
26	3_0217 1	18	1580 0	1674 4	944	-1	*rfaF*	ADP-heptose--LPS heptosyltransferase 2	*	P37692
26	3_0217 2	18	1673 7	1779 8	1061	-1		hypothetical protein		
26	3_0217 3	18	1779 1	1871 1	920	-1		lipopolysaccharide core biosynthesis protein		PRK104 22
26	3_0217 4	18	1871 1	1946 0	749	-1		3-deoxy-D-manno-octulosonic-acid kinase		PRK017 23
26	3_0217 5	18	1959 6	2124 2	1646	+1	*eptB*	Phosphoethanolamine transferase EptB		P37661
26	3_0217 6	18	2126 0	2181 7	557	-1		hypothetical protein		PRK103 45
26	3_0217 7	18	2182 6	2271 9	893	-1		hypothetical protein		
26	3_0217 8	18	2272 8	2418 5	1457	-1	*patA*	Peptidoglycan O-acetyltransferase		O25526
26	3_0217 9	18	2419 1	2529 1	1100	-1	*epsD*	Putative glycosyltransferase EpsD		P71053
26	3_0218 0	18	2528 1	2602 4	743	-1		Glycosyl transferase family 2		PF00535 .20

26	3_02181	18	26110	27180	1070	+1	*tuaC*	Putative teichuronic acid biosynthesis glycosyltransferase TuaC		O32272
26	3_02182	18	27203	28216	1013	+1	*epsJ*	putative glycosyltransferase EpsJ	*	P71059
26	3_02183	18	28226	28816	590	-1		hypothetical protein		PRK10345
26	3_02184	18	28813	29937	1124	-1	*tuaH*	Putative teichuronic acid biosynthesis glycosyltransferase TuaH		O32267
26	3_02185	18	29934	31271	1337	-1		O-Antigen ligase	**	PF04932.9
26	3_02186	18	31283	32356	1073	-1	*vioA*	dTDP-4-amino-4%2C6-dideoxy-D-glucose transaminase		Q6U1I3

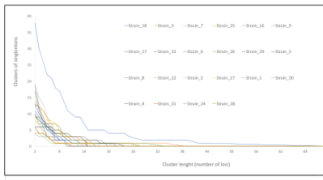
Supplementary Table 2.5. genes MLST codes of the strains object of study. In the last column is indicated the nearest sequence type code-(nearest ST). Some genes sequences resulted in new alleles, these genes are indicated with an asterisk.

Strains code	*asp A*	*atp A*	*gln A*	*glt A*	*gly A*	*pg m*	*tkt*	nearests STs
1	20	7	20	15	186	8	14	78
2	209	15	15	48	638	74	86	646
3	1	1	1	1	550	1	1	1
4	23	7	114	19	90	101	165	344
5	6	5	38	7	144	93	13	36
6	20*	22	4	17	67*	37	14	777,79
7	20	7	1	19	146	101	165	340
8	234	15	26	164	375	260	176	460
9	255	5	1	30	103	53	66	594
10	5*	34	11	30	56*	2	55	52,429,138,477,317,33,369,531,507,361,31,76,259,493,27,694
11	5*	34	11	30	56*	2	55	33,694,477,76,507,317,369,531,138,493,259,31,27,429,361,52
12	37	5	11	17	551	16	40	139
13	5*	34*	11*	30*	56*	2	55	507,138,361,694,477,259,531,369,76,317,33,429,493,31,52,27
14	378*	61	128*	144	375*	5*	183	742,752,652,403,523,808,404,474,743
15	14*	61	128*	144	746	5*	183	729,403,808,474,730,404,652,743,649,530
16	13	4	40	19	677*	102	158	419,401,440
17	153	4	40	64*	140*	11	24	349
18	46	19	9	23	25	229*	204*	162
19	37	5	11	17	127	16	40	139
20	37	5	11	17	194*	16	40	139
21	37	5	11	17	127	16	40	139
22	362	2	7*	30	72*	19*	2	507
23	5*	34	11	30	56*	2	55	27,76,369,317,477,361,52,493,33,429,694,31,507,138,531,259
24	380	62	26	240	767	43	329	825
25	39	33	33	39	57	109	38	148
26	41	222	26	12	67*	39	24	320,73,468,152,377,761,141,746,739

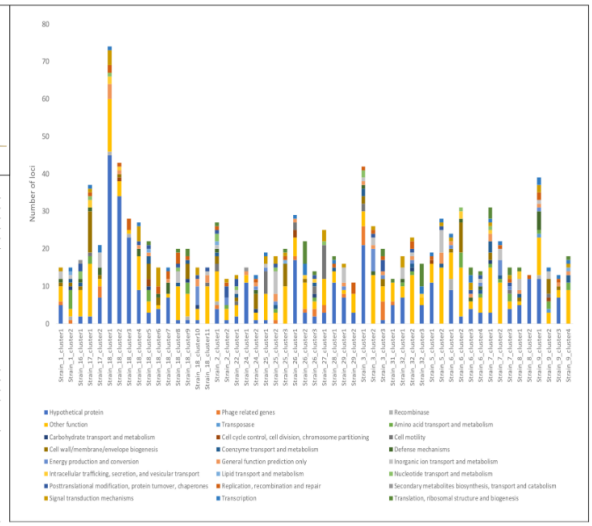
27	30	5	9	30	120	35	4	586
28	25	7	20	15	186	8	14	78
29	26	17	17	2	36	327*	75	630,147,508,468,112,73,74,13,526,113,577
30	6	34	1	12	120	50	14	45
31	28	5	7	7	167	266	6	116
32	75	25	49	23	523	16	33*	198,643,494,485,140,492,679



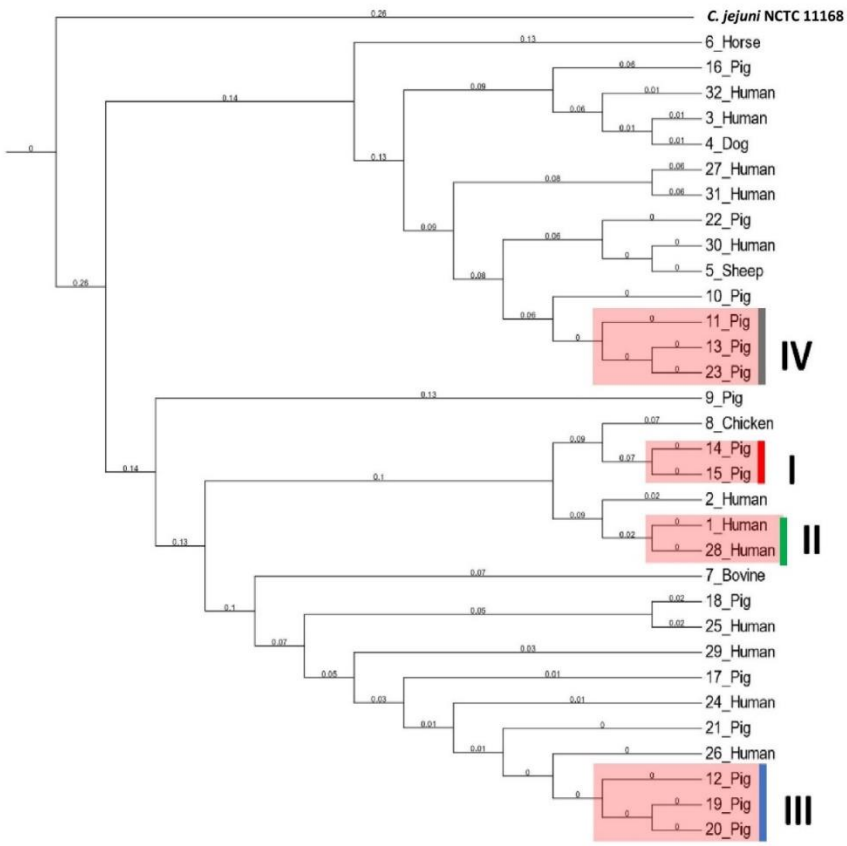
Supplementary Figure 2.1. Bar-plots (**A**) displaying the distribution of COG classes in the each of the 32 annotated genomes (% of putative proteins assigned to a class compared to the putative proteins). Coding keys of colors and class codes are shown in the caption. Heatmap (**B**) showing the presence (grey) / absence (white) matrix of genes involved in the cytoskeleton function.



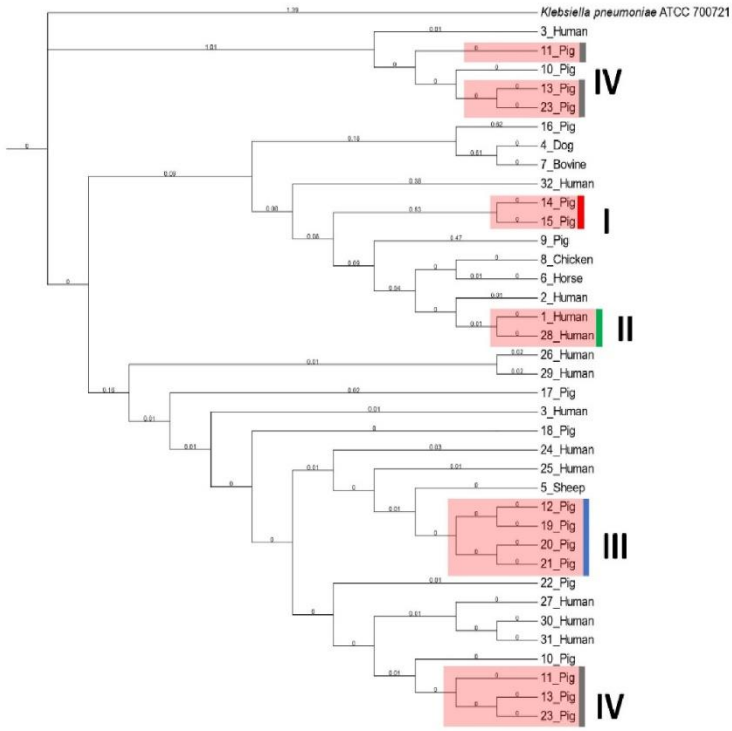
Source	Strains	No. of singletons	No. of singletons cluters (> 2 loci)	No. of singletons cluters (> 10 loci)
Sheep faeces	Strain_5	87	8	2
Chicken skin	Strain_8	90	10	2
Cow faeces	Strain_7	160	19	3
Dog faeces	Strain_4	72	6	0
Horse faeces	Strain_6	129	11	4
Human faeces	Strain_1	80	10	2
	Strain_2	110	9	2
	Strain_3	168	14	3
	Strain_24	55	4	2
	Strain_25	150	19	3
	Strain_26	125	9	3
	Strain_27	86	8	1
	Strain_28	43	4	1
	Strain_29	122	17	2
	Strain_30	77	6	0
	Strain_31	64	6	1
	Strain_32	131	13	3
Pig meat	Strain_9	146	10	4
Pig duodenum	Strain_14	2	0	0
Pig caecum	Strain_15	2	0	0
Pig colon descendent	Strain_16	149	19	1
	Strain_17	136	13	2
	Strain_18	476	38	11
Pig rectum	Strain_10	0	0	0
	Strain_11	0	0	0
	Strain_12	1	0	0
	Strain_13	0	0	0
	Strain_19	0	0	0
	Strain_20	0	0	0
	Strain_21	0	0	0
	Strain_22	88	10	1
	Strain_23	0	0	0



Supplementary figure 2.2. Singletons distribution along the genomes and composition of the main clusters of singletons (> 10 loci) identified in the accessory genome.



Supplementary figure 2.3. UPGMA phylogenetic analysis of *porA*. The groups of strains from isolated are indicated from the panes, the numbers (I–IV) indicate the different groups of strains.



Supplementary figure 2.4. UPGMA phylogenetic analysis of O-antigen ligase. The strains grouped by source of isolation are indicated from the panes, the numbers (I–IV) indicate the different groups of strains. In the case of the O-antigen ligase dendrogram some strains are repeated, this aspect is linked to the presence of several gene copies. The O-antigen ligase sequence of *Klebsiella pneumoniae* ATCC 700721 (used from Prokka for the functional annotation) has been used as outgroup.

Supplementary Table 3.1. *A. butzleri in vitro* test log CFU cm^{-2} values. The data in this table represent bacteria load of bacterial inoculum (T0), bacteria detected after PBS washing (T1) and bacteria detected after gentamicin application (T2). T1 and T2 loads have been detected at 30' and 90'. In table are shown the corresponding standard errors while "ND" indicates a not detectable load.

strain	**LMG 11119**	**LMG 10828^T**	**31**
T0			
average (log)	6.58	6.54	6.51
standard error	0.15	0.09	0.63
T1 30'			
average (log)	5.42	3.82	2.25
standard error	0.04	0.22	1.14
T1 90'			
average (log)	6.21	3.76	3.90
standard error	0.53	0.20	0.67
T2 30'			
average (log)	3.64	ND	1.91
standard error	0.03	ND	0.99
T2 90'			
average (log)	3.68	2.68	1.71
standard error	0.26	0.13	0.86

Supplementary Table 3.2. *A. butzleri* DMEM DEGs from the comparison with Arcobacter agar. The table shows logFC (< -1.5, > 1.5) logCPM, *p value* (< 0.05) and FDR (< 0.05) values of differentially expressed genes linked to currently considered putative virulence genes. The column "gene" shows the protein name of the DEGs and the relative locus tag.

gene	logFC	logCPM	P value	FDR

Strain LMG 11119				
01097 Virulence sensor protein BvgS precursor	2.25	3.29	1.31E-09	1.83E-09
00898 Capsule polysaccharide biosynthesis protein	2.18	5.09	2.95E-18	5.45E-18
00852 Chemotaxis response regulator protein-glutamate methylesterase	2.75	5.95	1.07E-32	3.54E-32
00851 Chemoreceptor glutamine deamidase CheD	3.41	6.44	8.45E-51	7.37E-50
01001 Methyl-accepting chemotaxis protein IV	1.56	4.71	1.45E-10	2.09E-10
00850 Chemotaxis protein methyltransferase	3.71	6.32	2.77E-49	2.14E-48
00201 Chemotaxis protein CheW	2.32	8.70	2.80E-29	7.80E-29
00848 Chemotaxis protein CheY	1.66	8.83	5.95E-16	1.01E-15
01806 Chemotaxis protein CheY	3.79	7.90	8.55E-47	5.77E-46
00338 hypothetical protein	2.40	5.59	1.40E-24	3.26E-24
00605 hypothetical protein	2.03	6.65	1.44E-22	3.11E-22
00883 Colicin V production protein	4.38	7.06	3.87E-70	1.33E-68
02099 Spore coat protein SA	4.75	4.95	2.16E-45	1.36E-44
00429 Type II secretion system protein F	4.04	5.29	8.52E-39	3.78E-38
00892 Putative acetyltransferase EpsM	3.89	4.83	2.40E-35	8.91E-35
00380 Biopolymer transport protein ExbB	3.75	7.51	4.15E-52	3.81E-51
00920 Biopolymer transport protein ExbB	1.78	4.45	1.97E-10	2.82E-10
00155 Biopolymer transport protein ExbD	4.74	6.05	9.52E-67	2.36E-65
00379 Biopolymer transport protein ExbD	4.01	5.87	2.23E-41	1.14E-40
00919 Biopolymer transport protein ExbD	1.80	5.99	1.75E-15	2.91E-15
02238 Sensor protein FixL	2.99	4.42	1.66E-19	3.18E-19
01554 Flagellum-specific ATP synthase	2.48	7.19	4.44E-25	1.06E-24
01686 Flagellar hook-associated protein 1	1.75	9.19	7.08E-16	1.19E-15
01531 Flagellar M-ring protein	3.81	5.73	1.83E-43	1.05E-42
01649 Flagellin N-methylase	3.62	5.51	6.18E-42	3.29E-41
01807 flagellar motor switch protein	3.84	6.97	2.31E-50	1.99E-49

01533 flagellar assembly protein H	2.37	5.15	1.08E-20	2.16E-20
01685 Flagellar L-ring protein precursor	3.12	6.66	3.79E-43	2.15E-42
01530 Flagellar basal body rod protein FlgB	2.98	6.43	1.64E-32	5.34E-32
01549 Flagellar basal-body rod protein FlgC	2.46	8.69	3.06E-30	8.95E-30
01529 Flagellar basal-body rod protein FlgG	2.13	7.36	1.48E-22	3.19E-22
01556 Flagellar biosynthesis protein FlhA	4.16	5.30	4.77E-42	2.57E-41
01552 Flagellar biosynthetic protein FlhB	3.19	6.01	2.13E-46	1.41E-45
01834 Flagellar biosynthetic protein FlhB	2.63	4.29	6.50E-17	1.14E-16
01545 Flagellar biosynthesis protein FlhF	3.89	5.75	6.01E-47	4.11E-46
01532 Flagellar motor switch protein FliG	3.17	5.65	2.40E-35	8.91E-35
01550 Flagellar hook-length control protein FliK	3.60	6.42	4.31E-45	2.64E-44
01833 Flagellar hook-length control protein FliK	1.87	6.14	3.38E-17	6.02E-17
01683 flagellar basal body-associated protein FliL	2.38	7.89	2.60E-21	5.34E-21
01677 Flagellar motor switch protein FliM	3.01	6.97	1.07E-40	5.34E-40
01541 Flagellar motor switch protein FliN	1.52	7.82	1.12E-12	1.70E-12
01066 Flagellar biosynthetic protein FliP precursor	3.54	5.48	4.89E-42	2.63E-41
01680 Flagellar biosynthetic protein FliQ	4.78	6.82	1.18E-82	1.92E-80
01551 flagellar biosynthesis protein FliR	3.71	4.96	2.34E-31	7.19E-31
01688 Flagellar protein FliS	3.47	8.32	5.67E-32	1.80E-31
00015 Flagellar assembly factor FliW 2	2.26	8.15	2.19E-16	3.76E-16
01103 HlyD family secretion protein	2.47	1.27	1.32E-02	1.48E-02
00729 Internalin-J precursor	1.85	6.25	1.27E-20	2.55E-20
00943 Colicin I receptor precursor	1.98	1.70	2.97E-03	3.41E-03
02180 Chondroitin synthase	2.59	4.72	5.80E-19	1.08E-18
00285 Methyl-accepting chemotaxis protein 4	3.95	5.04	3.88E-20	7.59E-20
00174 Motility protein B	3.24	6.32	1.34E-41	7.00E-41
00229 Response regulator MprA	3.48	4.92	1.17E-31	3.65E-31

00773 Response regulator MprA	2.85	8.14	5.54E-16	9.37E-16
00783 Response regulator MprA	2.97	3.90	9.01E-17	1.57E-16
01108 Response regulator MprA	2.86	1.33	1.03E-03	1.19E-03
01367 Response regulator MprA	1.84	8.87	1.43E-16	2.47E-16
01799 Response regulator MprA	3.60	2.96	2.23E-12	3.36E-12
02088 Transcriptional repressor MprA	2.25	5.06	6.01E-09	8.17E-09
01624 Lipid A export ATP-binding/permease protein MsbA	3.43	5.54	1.11E-36	4.47E-36
01623 putative peptidoglycan biosynthesis protein MurJ	3.35	2.96	1.34E-12	2.03E-12
02097 O-Antigen ligase	3.34	2.96	4.65E-13	7.11E-13
00648 Outer membrane porin F precursor	5.57	6.37	6.07E-88	1.54E-85
00828 putative sensor histidine kinase pdtaS	2.35	2.88	2.26E-07	2.92E-07
01219 putative sensor histidine kinase pdtaS	2.93	1.20	3.56E-03	4.08E-03
02101 Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase precursor	4.21	4.57	3.05E-30	8.92E-30
02220 Virulence transcriptional regulatory protein PhoP	1.93	6.25	3.92E-06	4.89E-06
02256 Transcriptional regulatory protein PhoP	4.70	4.16	5.84E-32	1.85E-31
01107 Sensor protein PhoQ	3.49	1.47	1.62E-04	1.93E-04
01640 Putative phospholipase A1 precursor	2.62	4.82	1.66E-24	3.86E-24
00175 Chemotaxis protein PomA	2.29	7.60	3.21E-29	8.93E-29
02172 Lipopolysaccharide heptosyltransferase 1	3.33	4.75	5.41E-31	1.63E-30
02175 Lipopolysaccharide heptosyltransferase 1	2.37	5.13	2.54E-23	5.61E-23
02179 dTDP-glucose 4,6-dehydratase	1.56	5.80	2.61E-14	4.17E-14
01523 Ribosome maturation factor RimM	4.94	5.22	3.83E-48	2.80E-47
01691 Ribosomal protein S12 methylthiotransferase RimO	2.66	6.76	1.64E-39	7.51E-39
01756 Ribosome maturation factor RimP	4.21	5.85	4.57E-46	2.95E-45
02176 Glucose-1-phosphate thymidylyltransferase 1	1.66	7.52	7.86E-16	1.32E-15
02177 dTDP-4-dehydrorhamnose 3,5-epimerase	2.47	6.64	2.64E-31	8.05E-31
02178 dTDP-4-dehydrorhamnose reductase	2.92	5.02	1.05E-23	2.35E-23

02102 SPBc2 prophage-derived glycosyltransferase SunS	4.24	5.25	6.86E-42	3.63E-41
02069 hypothetical protein	3.49	4.63	1.00E-24	2.35E-24
01226 Methyl-accepting chemotaxis protein IV	2.97	7.65	1.44E-45	9.08E-45
01311 Methyl-accepting chemotaxis protein IV	2.60	3.40	9.01E-11	1.30E-10
01564 Methyl-accepting chemotaxis protein IV	2.04	1.61	7.89E-03	8.93E-03
01224 Methyl-accepting chemotaxis protein II	2.78	4.09	4.10E-19	7.75E-19
01655 Methyl-accepting chemotaxis protein II	2.81	5.68	4.54E-38	1.94E-37
01657 Methyl-accepting chemotaxis protein II	1.85	3.47	6.37E-08	8.38E-08
01854 Methyl-accepting chemotaxis protein II	2.03	4.80	4.75E-17	8.40E-17
02058 Methyl-accepting chemotaxis protein II	2.59	4.20	2.19E-15	3.63E-15
00460 Methyl-accepting chemotaxis protein II	2.65	4.25	6.78E-13	1.03E-12
01871 16S/23S rRNA (cytidine-2'-O)-methyltransferase TlyA	3.33	5.95	1.54E-29	4.33E-29
00156 colicin uptake protein TolQ	4.38	7.13	2.61E-53	2.59E-52
00117 TonB-dependent heme receptor A precursor	2.92	2.00	7.66E-06	9.46E-06
01076 Gram-negative bacterial tonB protein	4.12	4.56	5.67E-27	1.44E-26
01309 TonB dependent receptor	1.79	2.26	1.06E-03	1.24E-03
00918 transport protein TonB	2.01	4.70	6.52E-14	1.02E-13
00422 Tetrathionate sensor histidine kinase TtrS	2.52	4.38	1.66E-18	3.08E-18
02257 Tetrathionate sensor histidine kinase TtrS	3.50	1.30	1.01E-03	1.18E-03
02182 N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase	4.13	3.80	4.37E-20	8.54E-20
01869 Type II secretion system protein D precursor	2.68	4.81	3.27E-18	6.03E-18
01546 Flagellum site-determining protein YlxH	3.26	6.03	6.50E-28	1.73E-27
02239 Transcriptional regulatory protein ZraR	2.66	4.83	7.90E-17	1.38E-16
Strain LMG10828^{T}				
00490 Chemotaxis response regulator glutamate methylesterase	1.77	6.85	1.23E-11	3.91E-11
02099 Putative glycosyltransferase EpsJ	2.31	5.12	2.51E-15	1.03E-14
01366 Biopolymer transport protein ExbB	2.62	9.31	3.04E-29	4.09E-28

01365 Biopolymer transport protein ExbD	2.65	7.70	1.01E-27	1.17E-26
01208 Sensor protein FixL	1.67	3.15	2.78E-05	5.41E-05
01958 Flagellar assembly protein H	2.13	5.87	8.21E-17	3.79E-16
00565 Flagellin	-1.67	9.33	9.03E-12	2.9E-11
01960 Flagellar M-ring protein	2.26	6.46	1.63E-19	9.15E-19
00972 flagellar motor switch protein	1.68	7.64	1.07E-12	3.71E-12
01961 Flagellar basal body rod protein FlgB	1.64	7.54	4.05E-12	1.35E-11
01935 Flagellar biosynthesis protein FlhA	1.62	5.29	1.24E-07	2.97E-07
01959 Flagellar motor switch protein FliG	2.03	6.46	2.28E-16	1E-15
00683 Flagellar biosynthetic protein FliP precursor	3.01	6.87	7.93E-32	1.52E-30
00026 Flagellar biosynthetic protein FliQ	1.58	7.32	9.72E-09	2.53E-08
01940 flagellar biosynthesis protein FliR	1.75	5.88	7.05E-11	2.11E-10
00673 Gram-negative bacterial TonB protein	1.78	5.47	9.31E-10	2.57E-09
00740 Hemolysin transporter protein ShlB precursor	1.96	2.90	5.16E-06	1.09E-05
00452 Methyl-accepting chemotaxis protein IV	3.22	4.85	3.03E-20	1.8E-19
00353 Methyl-accepting chemotaxis protein IV	2.15	6.91	7.89E-20	4.51E-19
00331 Methyl-accepting chemotaxis protein 4	2.91	6.13	1.41E-16	6.38E-16
00834 Methyl-accepting chemotaxis protein II	1.95	5.65	1.38E-12	4.73E-12
02234 Methyl-accepting chemotaxis protein II	1.55	4.28	1.12E-07	2.69E-07
00802 putative peptidoglycan biosynthesis protein MurJ	2.02	3.90	1.87E-10	5.42E-10
02096 O-Antigen ligase	1.53	4.35	4.82E-06	1.01E-05
02078 Transcriptional regulatory protein PhoP	1.51	5.41	8.03E-06	1.66E-05
02038 Virulence transcriptional regulatory protein PhoP	1.60	7.43	2.19E-06	4.71E-06
00642 Sensor protein PhoQ	1.60	2.67	0.00039	0.000663
02180 cryptic beta-D-galactosidase subunit beta	1.74	6.35	1.49E-10	4.35E-10
00164 TonB-dependent heme receptor A precursor	1.85	2.12	0.0007	0.001159
01740 transport protein TonB	1.84	7.28	5.13E-07	1.16E-06

02079 Tetrathionate sensor histidine kinase TtrS	1.53	1.98	0.010922	0.015705
02212 Urease accessory protein UreD	1.70	2.83	8.34E-05	0.000155
02209 Urease accessory protein UreE	1.55	4.00	1.96E-05	3.88E-05
02208 Urease accessory protein UreF	2.06	2.70	1.88E-05	3.74E-05
01307 Sensor protein ZraS	1.68	5.92	1.39E-05	2.81E-05
Strain 31				
00607 Chemotaxis protein methyltransferase	1.60	7.37	1.95E-07	4.22E-07
00609 Chemotaxis response regulator glutamate methylesterase	2.76	7.00	1.02E-19	1.03E-18
00608 Chemoreceptor glutamine deamidase CheD	2.06	6.65	2.25E-10	6.46E-10
01893 CheW-like domain protein	2.05	3.20	2.54E-06	4.99E-06
01607 Chemotaxis protein CheY	2.04	8.49	1.82E-11	5.87E-11
01466 Putative glycosyltransferase EpsD	1.91	4.26	1.28E-07	2.81E-07
00986 Biopolymer transport protein ExbB	2.68	9.78	2.11E-11	6.76E-11
00233 Biopolymer transport protein ExbD	2.09	5.31	7.58E-10	2.06E-09
00987 Biopolymer transport protein ExbD	2.76	8.20	3.95E-11	1.22E-10
01606 flagellar motor switch protein	2.62	7.55	6.07E-20	6.39E-19
01946 flagellar assembly protein H	2.50	6.49	2.39E-19	2.28E-18
01731 Flagellar L-ring protein precursor	1.92	6.76	1.03E-10	3.07E-10
01948 Flagellar M-ring protein	2.49	6.73	3.21E-20	3.68E-19
00550 Flagellin	-1.82	9.61	5.96E-07	1.24E-06
01949 Flagellar basal body rod protein FlgB	1.84	7.71	2.92E-11	9.18E-11
01923 Flagellar biosynthesis protein FlhA	2.40	6.65	1.15E-16	7.09E-16
01332 Flagellar biosynthetic protein FlhB	1.83	5.49	2.92E-09	7.44E-09
01934 Flagellar biosynthesis protein FlhF	2.61	6.39	1.26E-18	1.08E-17
01947 Flagellar motor switch protein FliG	2.41	6.87	5.47E-19	4.97E-18
01929 Flagellar hook-length control protein FliK	1.63	7.08	5.28E-07	1.1E-06
01723 Flagellar motor switch protein FliM	1.52	5.07	1.09E-06	2.2E-06

00445 Flagellar biosynthetic protein FliP precursor	2.55	7.15	4.52E-15	2.2E-14
01726 Flagellar biosynthetic protein FliQ	1.82	7.18	6.39E-09	1.57E-08
01928 flagellar biosynthesis protein FliR	2.23	5.80	1.25E-12	4.5E-12
01914 Methyl-accepting chemotaxis protein IV	1.91	3.40	3.56E-05	6.29E-05
00106 Methyl-accepting chemotaxis protein 4	2.19	5.55	7.53E-12	2.52E-11
02131 Methyl-accepting chemotaxis protein II	2.14	4.24	2.83E-09	7.22E-09
00007 Methyl-accepting chemotaxis protein IV	2.17	5.08	2.43E-11	7.76E-11
00214 Motility protein B	2.15	6.89	4.95E-15	2.38E-14
00343 putative peptidoglycan biosynthesis protein MurJ	2.84	4.56	1.96E-14	8.72E-14
00132 Outer membrane porin F precursor	-1.53	11.68	7.01E-05	0.00012
00815 Outer membrane porin F precursor	1.58	6.56	8.77E-07	1.79E-06
00554 Response regulator MprA	1.82	3.47	1.31E-05	2.39E-05
01378 Response regulator MprA	1.89	7.55	6.23E-06	1.17E-05
02043 cryptic beta-D-galactosidase subunit beta	2.69	6.71	3.81E-17	2.47E-16
01440 TlyA	2.03	6.72	1.99E-09	5.17E-09
00455 Gram-negative bacterial tonB protein	2.25	5.31	4.23E-10	1.18E-09
01819 transport protein TonB	1.63	8.32	5.18E-06	9.85E-06
01043 Tetrathionate sensor histidine kinase TtrS	1.60	3.81	7.94E-06	1.48E-05
02069 Urease subunit alpha	1.57	2.61	0.002277	0.00337
02067 Urease accessory protein UreE	1.92	3.73	5.56E-07	1.16E-06
02066 Urease accessory protein UreF	2.44	2.49	8.54E-06	1.58E-05
02065 Urease accessory protein UreG	1.95	3.69	8.4E-07	1.72E-06
00887 Sensor protein ZraS	2.04	3.33	8.94E-06	1.65E-05

Supplementary Table 3.3. *A. butzleri* DEGs after 30' of contact with host cells. The table shows logFC (< -1.5, > 1.5), *p value* (< 0.05) values of differentially expressed genes linked to the currently considered putative virulence genes. The column under the strains names shows the protein name of the DEGs and the relative locus tag.

LMG 11119	**log FC**	**p value**	**31**	**log FC**	**p value**	**LMG 10828^T**	**log FC**	**p value**
00920 Biopolymer transport protein ExbB	2.1	7.19E-13	00234 hypothetical protein	1.84	5.28E-04	00201 Acetaldehyde dehydrogenase 2	2.58	5.55E-21
01311 Methyl-accepting chemotaxis protein IV	3.10	3.49E-19	00235 Acetaldehyde dehydrogenase 2	4.01	2.85E-08	01089 Cytochrome c-type protein SHP precursor	2.35	7.13E-09
00918 transport protein TonB	1.77	1.38E-16	00363 Cytochrome c	3.37	5.18E-07	00640 hypothetical protein	-1.15	1.78E-04
00082 NADH-quinone oxidoreductase subunit I	1.54	1.57E-15	00744 hypothetical protein	-3.86	9.33E-06	00624 hypothetical protein	-2.20	1.01E-11
00131 Putative electron transport protein YccM	1.61	2.43E-10	00745 Phosphate import ATP-binding protein PstB	-3.35	1.23E-05	01903 hypothetical protein	-1.76	8.27E-04
00143 hypothetical protein	1.50	5.06E-14	00747 Phosphate transport system permease protein PstC	-1.89	9.43E-04	00325 Inner membrane protein YjcH	3.15	1.50E-32
00145 30S ribosomal protein S15	1.73	1.74E-19	00750 Phosphate-binding protein PstS precursor	-2.59	6.19E-09	00324 Cation/acetate symporter ActP	2.90	1.12E-20
00153 Acetaldehyde dehydrogenase 2	-2.46	3.31E-39	01598 Integral membrane protein TerC family protein	-1.78	1.45E-03	00315 Acetate kinase	2.21	9.31E-17
00183 hypothetical protein	1.51	1.08E-12	00112 Inner membrane protein YjcH	2.39	2.82E-06	00316 Phosphate acetyltransferase	2.16	1.03E-13
00352 Antibiotic biosynthesis monooxygenase	1.66	7.29E-09	00113 Cation/acetate symporter ActP	1.81	3.29E-04	00187 Putative regulator of ribonuclease activity	2.46	3.04E-12
00475 Helix-turn-helix domain protein	1.75	1.48E-16	00265 Cytochrome c-type protein NrfH	5.13	2.54E-12	02298 Transglutaminase-like superfamily protein	6.80	1.23E-11

		2.	00371 hypothetical	2.	6.	02076 hypothetical	-	7.
	1.	25	protein	2	20	protein	6.	68
00610 hypothetical	7	E-		8	E-		9	E-
protein	3	03			07		1	06
00613 RNA		9.	00411 hypothetical	1.	3.	01448 hypothetical	-	9.
recognition motif.	1.	43	protein	5	27	protein	5.	09
(a.k.a. RRM, RBD, or	6	E-		2	E-		6	E-
RNP domain)	5	20			04		4	04
		3.	00434 OstA-like	-	2.	02291 Formate	6.	2.
	1.	97	protein	1.	64	hydrogenlyase	7	21
00693 fec operon	6	E-		9	E-	complex iron-sulfur	0	E-
regulator FecR	0	06		4	04	subunit		20
		1.	00550 Flagellin	2.	2.	01172 hypothetical	1.	1.
	1.	22		1	11	protein	5	42
00881 hypothetical	8	E-		3	E-		1	E-
protein	8	16			05			12
		4.	00854 Cytochrome	5.	9.	02292 Receptor	6.	1.
	1.	28	c-type protein SHP	0	18	family ligand	2	66
00992 CheW-like	6	E-	precursor	4	E-	binding region	8	E-
domain protein	2	06			19			06
		1.	01106 Maf-like	-	5.	02290 Twin-	6.	5.
	1.	94	protein	9.	48	arginine leader-	3	97
01028 Aerotaxis	8	E-		3	E-	binding protein	2	E-
receptor	7	10		8	07	DmsD		08
		5.	01250 ATP-	3.	4.	00588 Cytochrome	-	2.
	2.	25	dependent RNA	8	67	bd-I ubiquinol	1.	47
01289 Cysteine	3	E-	helicase RhlE	7	E-	oxidase subunit 1	9	E-
desulfurase	8	34			10		6	04
		2.	01713 hypothetical	-	3.	00980 hypothetical	-	2.
	2.	13	protein	1.	51	protein	5.	06
01290 NifU-like	2	E-		6	E-		9	E-
protein	9	35		0	04		8	04
		1.	01795 2-	1.	2.	02293 Sensor	5.	8.
	2.	22	methylcitrate	9	23	protein FixL	7	64
01409 30S ribosomal	2	E-	synthase	6	E-		7	E-
protein S21	8	25			05			07
		1.	02039 hypothetical	-	2.	02295 Natural	5.	4.
	1.	41	protein	8.	21	resistance-associated	5	85
	7	E-		8	E-	macrophage protein	3	E-
01436 recombinase A	2	17		0	06			06
	-	1.	02141 Copper	-	3.	01598 Periplasmic	2.	8.
	1.	94	chaperone CopZ	8.	24	serine endoprotease	1	44
01666 hypothetical	8	E-		1	E-	DegP precursor	1	E-
protein	4	05		1	05			18
	-	1.	02153 Cell wall-	2.	1.	00100 Aconitate	2.	9.
	1.	10	associated hydrolase	3	29	hydratase	2	91
01668 Phosphatidate	6	E-		2	E-		0	E-
cytidylyltransferase	1	06			05			12
		2.	00025 YceI-like	4.	1.	00264 Macrolide	6.	3.
	1.	47	domain protein	5	75	export ATP-	7	77
01778 2-aminoadipate	8	E-		9	E-	binding/permease	0	E-
transaminase	6	08			15	protein MacB		32
		1.	00172 Macrolide	4.	1.	00982 Helix-turn-	-	2.
	1.	04	export ATP-	9	63	helix domain protein	7.	46
01784 SnoaL-like	8	E-	binding/permease	2	E-		1	E-
polyketide cyclase	4	23	protein MacB		08		4	06
		1.	00173 Lipoprotein-	4.	5.	01339 HicB family	-	1.
01849 Hydrogenase	1.	43	releasing system	1	63	protein	1.	29
expression/formation	5	E-	ATP-binding protein	7	E-		6	E-
protein HypD	6	08	LolD		04		6	09

		2.	00231 D-cysteine	-	9.	02294	5.	9.
01850 Hydrogenase	1.	01	desulfhydrase	6.	68	Transcriptional	1	30
isoenzymes formation	7	E-		7	E-	regulatory protein	0	E-
protein HypC	0	10		5	04	ZraR		06
01851		2.	00266 Cytochrome	4.	2.	01422 Modification	-	4.
Hydrogenase/urease	1.	50	c-552 precursor	4	42	methylase DpnIIA	6.	93
nickel incorporation	8	E-		8	E-		2	E-
protein HypB	8	20			13		2	05
		4.	00374 hypothetical	2.	1.	01765 Colicin I	1.	4.
	1.	39	protein	3	43	receptor precursor	8	81
01860 hypothetical	5	E-		6	E-		6	E-
protein	8	17			04			03
	-	7.	00522 hypothetical	2.	2.	00727 hypothetical	-	3.
	2.	29	protein	3	71	protein	2.	60
01866 hypothetical	1	E-		3	E-		6	E-
protein	1	04			04		0	12
		1.	00555 Outer	2.	6.	00101 3-	2.	6.
	1.	27	membrane porin F	2	90	methylitaconate	5	93
02019 hypothetical	6	E-	precursor	1	E-	isomerase	6	E-
protein	1	11			05			10
		4.	00568 3,4-	7.	2.	00262 hypothetical	5.	3.
02032 putative HTH-	1.	03	dihydroxy-2-	4	14	protein	4	05
type transcriptional	6	E-	butanone 4-	2	E-		7	E-
regulator YxaF	9	15	phosphate synthase		06			14
		2.	00596 Ribosome-	1.	2.	00260 hypothetical	4.	9.
	1.	34	associated factor Y	8	88	protein	7	53
02035 DNA	5	E-		2	E-		6	E-
polymerase IV	1	11			05			17
		4.	00645 Divalent-	-	1.	00586 hypothetical	-	1.
	1.	21	cation tolerance	8.	28	protein	2.	73
02079 Radical SAM	8	E-	protein CutA	3	E-		6	E-
superfamily protein	4	13		3	05		6	03
		3.	00823 hypothetical	3.	6.	01338 hypothetical	-	1.
	1.	24	protein	2	79	protein	1.	65
	8	E-		7	E-		6	E-
02184 YciI-like protein	0	23			08		7	10
	-	1.	00894 hypothetical	-	3.	02077 hypothetical	-	3.
	5.	17	protein	8.	03	protein	6.	87
02255 hypothetical	4	E-		8	E-		4	E-
protein	2	04		7	06		0	05
	-	1.	00902 Imelysin	-	7.	02289 hypothetical	5.	2.
	1.	20		2.	26	protein	5	26
	5	E-		1	E-		0	E-
00777 Flagellin	5	17		1	04			30
		1.	01056 hypothetical	2.	3.	00998 hypothetical	-	2.
	1.	68	protein	8	93	protein	2.	39
00277 Inner membrane	7	E-		3	E-		1	E-
protein YjcH	3	05			04		8	05
		1.	01077 hypothetical	8.	6.	00263 Lipoprotein-	6.	2.
	2.	91	protein	9	48	releasing system	5	88
00239 Helix-turn-helix	0	E-		5	E-	ATP-binding protein	5	E-
domain protein	0	02			07	LolD		21
00240			01105 Phage	6.	3.	01762 Catecholate	2.	5.
Quinohemoprotein		9.	integrase family	1	23	siderophore receptor	4	40
amine dehydrogenase	1.	18	protein	9	E-	Fiu precursor	9	E-
A, alpha subunit, haem	8	E-			04			06
binding	7	04						

00242 Quinohemoprotein amine dehydrogenase subunit gamma	2.1	9.74 E-04	01313 Methyl-accepting chemotaxis protein II	-1.6	2.42 E-04	02310 hypothetical protein	3.4	5.95 E-07
00274 DNA polymerase III PolC-type	1.5	6.08 E-04	01347 tRNA-Glu(ttc)	-8.3	1.67 E-05	00778 Cytochrome c	2.5	2.87 E-23
00470 transfer-messenger RNA, SsrA	-2.1	7.48 E-05	01401 hypothetical protein	-8.2	2.87 E-05	01458 hypothetical protein	-6.1	5.67 E-05
00490 Helix-turn-helix domain protein	-1.5	1.06 E-02	01444 Methyl-accepting chemotaxis protein IV	-2.6	7.13 E-04	00558 Outer membrane porin F precursor	1.6	3.05 E-08
00557 Septum site-determining protein MinD	1.5	1.76 E-04	01461 Putative teichuronic acid biosynthesis glycosyltransferase TuaH	-7.3	1.09 E-04	00983 hypothetical protein	-3.2	4.61 E-03
00692 putative RNA polymerase sigma factor FecI	1.9	2.31 E-06	01467 Peptidoglycan O-acetyltransferase	-5.9	5.96 E-05	00471 Outer membrane lipoprotein Blc precursor	-1.9	6.73 E-16
00694 Ferric-pseudobactin BN7/BN8 receptor precursor	1.5	3.62 E-03	01497 Von Willebrand factor type A domain protein	-2.5	4.03 E-04	02287 Formate dehydrogenase H	4.4	1.65 E-21
00829 hypothetical protein	-2.8	3.93 E-03	01582 Flagellar filament 33 kDa core protein	-3.0	1.11 E-10	01454 hypothetical protein	-5.9	2.98 E-03
00935 hypothetical protein	2.2	2.80 E-13	01610 hypothetical protein	-8.0	2.92 E-05	01597 Transcriptional regulatory protein BaeR	1.5	2.37 E-07
00940 Catecholate siderophore receptor Fiu precursor	1.8	1.54 E-05	01794 Methylisocitrate lyase	1.8	1.27 E-04	00261 hypothetical protein	4.0	1.76 E-08
01101 Cadherin domain protein	-1.5	5.75 E-03	01820 Biopolymer transport protein ExbD	-1.7	2.52 E-04	00989 hypothetical protein	-5.8	6.17 E-04
01210 Ferrichrome receptor FcuA precursor	1.9	3.74 E-05	01903 Transposase DDE domain protein	-7.0	4.86 E-04	00555 Oxygen regulatory protein NreC	1.5	1.40 E-04
01568 bifunctional aldehyde dehydrogenase/enoyl-CoA hydratase	1.7	2.01 E-14	02009 UDP-N-acetyl-D-glucosamine 6-dehydrogenase	-7.8	1.19 E-08	01483 Tyrosine recombinase XerD	-2.0	1.54 E-04
01570 (3S)-malyl-CoA thioesterase	1.8	3.85 E-21	02082 Transcriptional activator protein CopR	-6.8	7.44 E-04	00616 Putative NAD(P)H nitroreductase YfkO	-1.5	5.98 E-12
01602 Cytochrome c	-3.5	2.51	02163 Methyl-accepting chemotaxis protein I	-2.6	3.83	00011 Zinc-responsive	-2.5	7.51

	27	E-63		E-07	transcriptional regulator	32	E-09
01665 zinc-responsive transcriptional regulator	-1.67	5.58E-07			00992 hypothetical protein	-5.23	3.06E-03
01667 CDP-alcohol phosphatidyltransferase	-1.84	3.30E-08			00472 protoporphyrinogen oxidase	-1.86	1.34E-14
01795 Iron uptake protein A1 precursor	1.73	8.87E-20			00619 Putative monooxygenase YcnE	-2.41	4.97E-09
01919 SkfA peptide export ATP-binding protein SkfE	2.48	1.78E-18			00620 Modulator of drug activity B	-2.07	4.92E-15
01920 hypothetical protein	1.58	1.30E-04			00618 NADPH dehydrogenase	-2.28	5.58E-20
01921 Ferrous iron transport protein A	1.67	2.12E-05			00284 Cobalt-zinc-cadmium resistance protein CzcB	-2.28	9.22E-06
02252 Cell wall-associated hydrolase	-2.68	1.09E-06			02150 hypothetical protein	-1.18	1.21E-03
					02271 hypothetical protein	-6.73	4.09E-06
					00617 EamA-like transporter family protein	-2.21	9.32E-19
					00425 Aerotaxis receptor	-1.96	2.26E-12
					00991 hypothetical protein	-5.18	4.60E-03
					00630 Sodium:sulfate symporter transmembrane region	-1.56	9.45E-04
					00259 hypothetical protein	3.75	2.18E-06
					01510 hypothetical protein	-6.67	4.80E-06

00631 General stress protein 14	-1.64	1.20E-06
00015 1-acyl-sn-glycerol-3-phosphate acyltransferase	-1.86	5.12E-08
00012 hypothetical protein	-2.69	6.06E-09
01485 hypothetical protein	-6.53	2.20E-05
00014 Phosphatidate cytidylyltransferase	-1.77	1.80E-13
01487 hypothetical protein	-6.14	9.38E-05
00573 Hemin transport system permease protein HmuU	-2.39	4.28E-03
01488 TraM recognition site of TraD and TraG	-5.67	8.88E-04
00426 Methyl-accepting chemotaxis protein IV	-1.97	3.73E-10
01596 Sensor protein BasS	2.14	2.93E-13
02034 Tyrosine recombinase XerC	-2.09	1.44E-04
02288 Putative formate dehydrogenase	3.45	5.83E-20
01087 Phosphoethanolamine transferase EptA	3.22	4.59E-16
01611 hypothetical protein	-1.53	6.49E-04
01307 Sensor protein ZraS	-1.55	2.79E-04

	01088 PAP2 superfamily protein	3.59	7.83E-23
	01486 hypothetical protein	-5.6	9.62E-04
	01141 Lactate utilization protein C	1.63	2.99E-06
	01008 Prophage CP4-57 integrase	-1.52	1.74E-07
	01887 hypothetical protein	-1.54	1.59E-03
	01603 Prophage CP4-57 integrase	-1.97	2.12E-09
	01609 hypothetical protein	-2.30	4.47E-03
	02100 Putative teichuronic acid biosynthesis glycosyltransferase TuaC	-1.62	2.48E-06
	00961 tRNA-Glu(ttc)	-8.90	2.96E-14
	00290 Quinohemoprotein amine dehydrogenase subunit gamma	-6.09	1.05E-04
	01610 hypothetical protein	-3.91	1.99E-04
	02237 Regulatory protein PchR	-1.88	7.55E-10
	00645 Chaperone protein DnaJ	-2.77	6.36E-04
	01613 hypothetical protein	-1.61	3.85E-04
	01607 hypothetical protein	-5.	1.65

		6	E-03
	01608 hypothetical protein	-2.5	3.98E-03
	01138 Glycolate permease GlcA	1.62	1.43E-07
	02238 Ferrichrome receptor FcuA precursor	-3.07	5.70E-04

Supplementary Table 3.4. *A. butzleri* DEGs after 90' of contact with host cells. The table shows logFC (< -1.5, > 1.5), *p value* (< 0.05) values of differentially expressed genes linked to the currently considered putative virulence genes. The column under the strains names shows the protein name of the DEGs and the relative locus tag.

LMG 11119	log FC	p value	31	log FC	p value	LMG 10828T	log FC	p value
00920 Biopolymer transport protein ExbB	3.13	8.2E-50	00164 Ribosomal RNA large subunit methyltransferase H	-1.92	3.98E-03	00201 Acetaldehyde dehydrogenase 2	2.70	9E-07
00918 transport protein TonB	2.72	1.6E-39	00235 Acetaldehyde dehydrogenase 2	3.61	5.06E-07	01089 Cytochrome c-type protein SHP precursor	2.22	2E-05
00082 NADH-quinone oxidoreductase subunit I	2.12	2.4E-31	00334 hypothetical protein	-10.24	9.77E-09	01900 heat shock protein GrpE	-1.51	2E-05
00145 30S ribosomal protein S15	2.72	2.6E-46	00363 Cytochrome c	3.15	2.06E-06	01901 Chaperone protein DnaK	-1.82	3E-08
00153 Acetaldehyde dehydrogenase 2	-2.72	2.1E-35	00399 hypothetical protein	-8.62	7.13E-05	00624 hypothetical protein	-3.25	3E-14
00610 hypothetical protein	2.40	2.0E-04	00435 putative GTP-binding protein EngB	-1.80	2.56E-03	01171 Cation efflux system protein CusA	1.92	4E-03
00693 fec operon regulator FecR	2.51	5.2E-18	00447 hypothetical protein	-8.10	1.45E-04	01902 Membrane-bound lytic murein transglycosylase A precursor	-2.31	6E-13
01289 Cysteine desulfurase	1.73	1.7E-	00466 50S ribosomal protein L36	-1.56	1.75E-04	01009 tRNA-Leu(caa)	-2.03	6E-09

			14									
01290	NifU-like protein	1.87	2.4E-19	00470	DNA-directed RNA polymerase subunit alpha	-1.51	5.05E-04	00906	Hydrogenase 2 maturation protease	-1.51	3E-03
01409	30S ribosomal protein S21	2.76	4.8E-42	00471	50S ribosomal protein L17	-1.56	8.91E-05	01318	Chaperone protein HtpG	-2.23	3E-10
01666	hypothetical protein	-1.82	1.1E-05	00520	Formyltetrahydrofolate deformylase	-1.62	1.53E-03	00904	Quinone-reactive Ni/Fe-hydrogenase large chain	-2.80	1E-11
01668	Phosphatidate cytidylyltransferase	-1.60	4.5E-07	00745	Phosphate import ATP-binding protein PstB	-3.11	2.81E-03	00905	Quinone-reactive Ni/Fe-hydrogenase B-type cytochrome subunit	-2.05	4E-07
01784	SnoaL-like polyketide cyclase	2.06	5.2E-28	00747	Phosphate transport system permease protein PstC	-2.33	2.22E-04	00903	Quinone-reactive Ni/Fe-hydrogenase small chain precursor	-3.15	3E-18
01850	Hydrogenase isoenzymes formation protein HypC	1.57	3.5E-07	00750	Phosphate-binding protein PstS precursor	-1.64	2.04E-03	00325	Inner membrane protein YjcH	3.53	4E-28
01860	hypothetical protein	1.61	6.2E-14	00960	tRNA modification GTPase MnmE	-1.56	4.29E-03	00324	Cation/acetate symporter ActP	3.80	2E-28
02184	YciI-like protein	1.89	5.3E-22	00972	GTPase Obg	-1.61	1.30E-03	00315	Acetate kinase	2.41	1E-14
00777	Flagellin	-1.91	1.2E-24	00974	Methionyl-tRNA formyltransferase	-1.68	5.65E-03	00316	Phosphate acetyltransferase	2.53	2E-14
00277	Inner membrane protein YjcH	2.82	1.6E-	00994	Aminodeoxyfutalosine nucleosidase	-2.20	1.98E-04	01287	hypothetical protein	1.50	2E-06

		10								
00274 DNA polymerase III PolC-type	2.40	6.5E-06	01195 ribosome-binding factor A	-1.53	2.52E-04	00187 Putative regulator of ribonuclease activity	3.00	2E-12		
00470 transfer-messenger RNA, SsrA	-2.11	2.0E-26	01300 hypothetical protein	-8.32	1.04E-04	02298 Transglutaminase-like superfamily protein	5.13	5E-06		
00490 Helix-turn-helix domain protein	-1.64	4.7E-03	01425 tRNA-Met(cat)	-11.81	1.65E-05	01290 hypothetical protein	1.95	2E-07		
00557 Septum site-determining protein MinD	1.58	7.1E-16	01455 Diacylglycerol kinase	-10.48	3.58E-10	02291 formate hydrogenlyase complex iron-sulfur subunit	5.81	5E-10		
00692 putative RNA polymerase sigma factor FecI	2.84	3.5E-20	01511 hypothetical protein	-9.65	1.09E-06	02290 twin-arginine leader-binding protein DmsD	5.17	1E-03		
00694 Ferric-pseudobactin BN7/BN8 receptor precursor	3.36	7.8E-17	01598 Integral membrane protein TerC family protein	-2.36	6.65E-06	00587 hypothetical protein	-7.21	6E-07		
00935 hypothetical protein	3.19	4.3E-34	01889 Translation initiation factor IF-1	-1.91	4.62E-04	00638 putative RNA polymerase sigma factor FecI	1.53	8E-04		
00940 Catecholate siderophore receptor Fiu precursor	3.77	3.2E-18	01956 Ribosome maturation factor RimM	-1.94	1.38E-03	01771 Carbon starvation protein A	1.81	3E-09		
01210 Ferrichrome receptor FcuA precursor	3.65	1.2E-17	01958 30S ribosomal protein S16	-2.01	1.53E-05	02293 Sensor protein FixL	5.03	7E-04		
01602 Cytochrome c	-4.41	6.1E-	01961 3-deoxy-D-manno-octulosonic acid transferase	-2.00	1.75E-03	02143 Iron-regulated protein A precursor	1.57	2E-04		

		62							
01665 zinc-responsive transcriptional regulator	-1.85	1.6E-09	02033 50S ribosomal protein L25	-1.79	5.04E-05	01357 Amino-acid carrier protein AlsT	1.75	1E-04	
01667 CDP-alcohol phosphatidyltransferase	-1.78	2.5E-09	02099 tRNA-Phe(gaa)	-1.54	1.26E-04	02295 Natural resistance-associated macrophage protein	4.27	8E-04	
01795 Iron uptake protein A1 precursor	2.56	6.2E-39	02112 Bifunctional DNA-directed RNA polymerase subunit beta-beta'	-1.52	1.28E-03	00767 hypothetical protein	-1.55	7E-06	
01919 SkfA peptide export ATP-binding protein SkfE	4.02	2.7E-59	00009 UDP-2,3-diacylglucosamine hydrolase	-8.02	1.95E-04	02242 Helix-hairpin-helix motif protein	1.96	9E-09	
01920 hypothetical protein	2.87	4.8E-14	00025 YceI-like domain protein	4.35	2.12E-14	01460 hypothetical protein	2.41	4E-04	
01921 Ferrous iron transport protein A	3.23	6.0E-16	00030 hypothetical protein	4.40	1.21E-03	01598 Periplasmic serine endoprotease DegP precursor	2.32	2E-11	
02252 Cell wall-associated hydrolase	-2.90	1.2E-09	00038 hypothetical protein	-6.42	4.83E-03	00264 Macrolide export ATP-binding/permease protein MacB	4.99	9E-11	
00919 Biopolymer transport protein ExbD	2.29	2.5E-24	00075 hypothetical protein	-2.16	1.59E-03	00982 Helix-turn-helix domain protein	-7.15	3E-06	
00943 Colicin I receptor precursor	2.40	1.8E-08	00089 hypothetical protein	-7.39	7.69E-04	01437 hypothetical protein	3.70	3E-05	
00042 hypothetical protein	1.61	4.9E-	00090 hypothetical protein	-8.00	2.44E-04	01339 HicB family protein	-1.85	1E-08	

00067 hypothetical protein	1.68	165.1E-18	00112 Inner membrane protein YjcH	2.39	3.46E-06	01770 hypothetical protein	1.62	6E-05
00068 50S ribosomal protein L32	1.73	2.7E-20	00113 Cation/acetate symporter ActP	1.93	8.99E-06	01765 Colicin I receptor precursor	4.29	1E-09
00083 NADH-quinone oxidoreductase subunit H	1.92	4.9E-24	00117 DNA polymerase III PolC-type	-6.42	5.04E-03	00727 hypothetical protein	-2.44	5E-10
00084 NADH-quinone oxidoreductase subunit 3	1.77	8.6E-21	00146 hypothetical protein	-7.20	1.18E-03	00262 hypothetical protein	4.15	1E-05
00089 NADH-quinone oxidoreductase subunit 6	1.54	3.9E-14	00147 Quinohemoprotein amine dehydrogenase subunit gamma	-6.76	3.07E-03	00260 hypothetical protein	2.97	4E-06
00109 Regulatory protein PchR	1.62	7.7E-09	00156 Blue-light-activated protein	6.23	1.14E-04	00977 Nitric oxide reductase subunit B	-1.77	3E-05
00121 Cytochrome c biogenesis protein CcsA	-1.68	6.3E-05	00157 Response regulator MprA	8.16	1.48E-05	01758 Fumarate hydratase class II	2.20	5E-08
00123 Cytochrome c-type protein NrfH	-1.67	4.3E-15	00172 Macrolide export ATP-binding/permease protein MacB	8.37	1.06E-46	00535 Cytochrome c-type protein TorY	-1.51	1E-04
00236 Cobalt-zinc-cadmium resistance protein CzcB	-2.09	9.2E-17	00173 Lipoprotein-releasing system ATP-binding protein LolD	8.08	4.65E-34	01289 hypothetical protein	2.01	1E-07
00237 Outer membrane efflux protein	-2.06	5.8E-	00174 hypothetical protein	6.81	1.69E-23	02041 Outer membrane efflux protein	2.17	4E-03

00238 Fatty acid metabolism regulator protein	-1.70	19.8E-17	00175 hypothetical protein	5.98	1.61E-13	00163 Regulatory protein PchR	2.52	9E-05
00260 Ribose-phosphate pyrophosphokinase	1.93	2.0E-27	00176 hypothetical protein	5.99	3.27E-25	01338 hypothetical protein	-1.65	8E-07
00287 tRNA-Val(tac)	1.53	1.8E-13	00177 hypothetical protein	4.79	1.61E-04	02289 hypothetical protein	5.12	1E-15
00290 tRNA-Asp(gtc)	1.54	6.4E-17	00189 Ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE	-2.27	1.45E-03	02045 FhuE receptor precursor	3.23	2E-08
00291 tRNA-Val(tac)	1.65	7.8E-12	00224 Small-conductance mechanosensitive channel	-1.74	5.09E-03	00998 hypothetical protein	-2.04	1E-04
00309 hypothetical protein	2.00	1.2E-11	00265 Cytochrome c-type protein NrfH	5.64	1.60E-25	01444 hypothetical protein	3.24	6E-06
00357 fec operon regulator FecR	1.78	1.4E-05	00266 Cytochrome c-552 precursor	4.41	1.54E-16	01331 YceI-like domain protein	2.81	1E-08
00381 ATP synthase epsilon chain	1.97	9.0E-22	00347 Fluoroacetyl-CoA thioesterase	1.54	2.19E-04	02048 hypothetical protein	2.25	2E-07
00382 ATP synthase subunit beta	1.69	1.4E-15	00352 Thiol-disulfide oxidoreductase ResA	-1.76	3.47E-03	00857 Dihydrolipoyl dehydrogenase	1.65	7E-07
00497 hypothetical protein	1.58	1.5E-	00371 hypothetical protein	2.93	3.05E-11	02042 hypothetical protein	2.70	5E-05

00504 tRNA-His(gtg)	1.80	12 1.7E-17	00376	hypothetical protein	-1.60	4.88E-03	00263	Lipoprotein-releasing system ATP-binding protein LolD	4.04	4E-06
00505 tRNA-Pro(tgg)	1.63	9.8E-10	00404	hypothetical protein	5.65	1.76E-10	01648	Succinate-semialdehyde dehydrogenase [NADP(+)]	1.53	1E-06
00521 50S ribosomal protein L31	1.96	2.7E-24	00405	hypothetical protein	-6.53	4.99E-03	02165	hypothetical protein	2.05	6E-08
00525 tRNA-Arg(tcg)	1.96	1.3E-17	00409	Lumazine-binding domain protein	-10.44	9.22E-10	01762	Catecholate siderophore receptor Fiu precursor	4.45	5E-18
00538 Spermidine synthase	1.58	1.8E-16	00411	hypothetical protein	1.88	1.59E-05	01766	Ferri-bacillibactin esterase BesA	2.12	2E-04
00544 hypothetical protein	-1.60	8.7E-14	00413	hypothetical protein	-9.56	2.03E-06	01761	Transcriptional regulatory protein WalR	1.74	9E-04
00564 Putative oxidoreductase CatD	1.83	3.8E-24	00458	Putative F0F1-ATPase subunit (ATPase_gene1)	-1.89	4.48E-05	02310	hypothetical protein	2.85	2E-04
00589 Peptide chain release factor 1	1.59	3.2E-17	00474	Cupin domain protein	2.21	1.29E-05	02049	hypothetical protein	2.27	6E-06
00590 30S ribosomal protein S20	1.53	1.6E-16	00522	hypothetical protein	2.09	1.15E-03	00778	Cytochrome c	2.67	1E-08
00598 50S ribosomal protein L20	2.28	1.3E-	00525	hypothetical protein	-1.93	5.38E-03	00994	hypothetical protein	-5.38	2E-03

		30								
00599 50S ribosomal protein L35	2.17	1.2E-26	00527 Signal transduction histidine-protein kinase/phosphatase MprB	-8.00	1.99E-04	00558 Outer membrane porin F precursor	1.83	1E-07		
00607 Cation efflux system protein CusA	1.99	4.2E-04	00542 Hemin transport system permease protein HmuU	-7.40	7.12E-04	01341 ECF RNA polymerase sigma factor SigE	1.99	5E-03		
00608 Cation efflux system protein CusB precursor	1.69	8.0E-03	00547 Sensor protein RstB	-2.23	3.19E-03	02287 Formate dehydrogenase H	4.42	1E-12		
00609 Outer membrane efflux protein	1.84	4.4E-03	00550 Flagellin	1.79	8.85E-05	00922 hypothetical protein	-1.60	4E-04		
00623 Ribosomal large subunit pseudouridine synthase B	1.57	5.9E-14	00555 Outer membrane porin F precursor	1.91	1.93E-03	01597 Transcriptional regulatory protein BaeR	2.10	5E-08		
00652 Ribonucleoside-diphosphate reductase subunit beta	1.61	5.5E-19	00568 3,4-dihydroxy-2-butanone 4-phosphate synthase	7.35	2.08E-05	01757 hypothetical protein	2.15	4E-04		
00653 Ribonucleoside-diphosphate reductase 1 subunit alpha	1.79	6.0E-19	00575 Bacterial transcription activator, effector binding domain	-9.06	9.72E-06	01606 hypothetical protein	-7.17	5E-07		
00654 Adenylosuccinate lyase	1.90	9.5E-20	00596 Ribosome-associated factor Y	1.58	3.07E-04	01759 hypothetical protein	2.25	3E-07		
00667 hypothetical protein	1.60	2.6E-16	00602 LPS-assembly protein LptD precursor	-1.58	1.37E-03	01442 Mu-like prophage I protein	3.22	2E-04		
00695 hypothetical protein	1.99	1.8E-	00612 Tyrosine recombinase XerC	2.61	9.46E-07	00513 putative transcriptional regulatory protein pdtaR	-5.35	2E-03		

		07							
00706 EamA-like transporter family protein	-1.67	7.7E-09	00629 YGGT family protein	-1.66	4.01E-03	01747 Ankyrin repeats (3 copies)	1.85	8E-06	
00715 hypothetical protein	-2.42	5.6E-12	00663 tRNA-Arg(tcg)	-12.27	1.27E-35	00633 EamA-like transporter family protein	-1.89	2E-04	
00723 Modulator of drug activity B	-1.65	3.1E-06	00668 hypothetical protein	8.85	1.27E-04	00989 hypothetical protein	-5.69	6E-04	
00933 hypothetical protein	-1.59	3.0E-15	00735 hypothetical protein	-6.37	4.99E-03	01656 hypothetical protein	2.03	7E-06	
00936 Fumarate hydratase class II	1.88	9.6E-22	00769 hypothetical protein	-8.66	7.18E-05	01345 mRNA interferase MazF	-1.57	1E-04	
00938 Blue-light-activated histidine kinase 2	1.59	9.3E-06	00773 hypothetical protein	1.91	1.42E-06	00011 zinc-responsive transcriptional regulator	-1.90	1E-07	
00941 PKHD-type hydroxylase	1.50	4.6E-05	00808 Anaerobic glycerol-3-phosphate dehydrogenase subunit C	1.69	9.36E-04	01429 hypothetical protein	2.75	6E-04	
00946 Nickel uptake substrate-specific transmembrane region	1.65	2.4E-20	00816 hypothetical protein	-8.63	6.70E-05	00285 Outer membrane efflux protein	-1.85	1E-04	
00987 Methionine aminopeptidase 1	1.55	8.1E-11	00823 hypothetical protein	3.64	9.45E-12	00286 Fatty acid metabolism regulator protein	-2.05	1E-06	
00988 Translation initiation factor IF-1	1.85	3.1E-	00832 N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	-7.89	2.23E-04	01082 hypothetical protein	1.88	8E-04	

		14							
00989 Uracil DNA glycosylase superfamily protein	1.76	5.5E-19	00851 Dihaem cytochrome c	2.30	2.29E-03	00619 Putative monooxygenase YcnE	-3.94	2E-22	
00995 Adenylate kinase	1.72	7.4E-20	00853 Dihaem cytochrome c	2.25	3.04E-03	00620 Modulator of drug activity B	-2.93	2E-15	
01009 Nickel uptake substrate-specific transmembrane region	2.00	3.4E-13	00854 Cytochrome c-type protein SHP precursor	4.90	6.92E-13	00618 NADPH dehydrogenase	-2.96	1E-21	
01055 fec operon regulator FecR	1.51	5.7E-03	00856 Phosphoethanolamine transferase EptA	2.18	2.23E-03	00284 Cobalt-zinc-cadmium resistance protein CzcB	-1.68	1E-03	
01192 Uracil-DNA glycosylase	1.80	5.4E-11	00880 hypothetical protein	-1.52	4.99E-03	01734 hypothetical protein	2.41	3E-07	
01201 hypothetical protein	2.04	1.0E-10	00885 Methylated-DNA--protein-cysteine methyltransferase, inducible	-8.94	1.53E-05	00617 EamA-like transporter family protein	-2.74	4E-17	
01211 fec operon regulator FecR	1.79	4.2E-08	00949 Peptide methionine sulfoxide reductase MsrB	-8.38	1.04E-04	00425 Aerotaxis receptor	-1.96	5E-10	
01212 RNA polymerase sigma factor YlaC	1.75	2.2E-10	00959 hypothetical protein	-1.56	1.65E-03	00991 hypothetical protein	-5.19	4E-03	
01246 putative D,D-dipeptide transport ATP-binding protein DdpF	1.57	4.2E-14	01021 Paraquat-inducible protein A	-8.38	8.47E-05	00630 Sodium:sulfate symporter transmembrane region	-1.57	4E-04	
01375 hypothetical protein	1.51	1.4E-	01045 tRNA-Phe(gaa)	-1.87	1.65E-05	00259 hypothetical protein	3.15	4E-03	

01420 CTP synthase	1.60	1.51E-17	01062 hypothetical protein	-6.73	2.84E-03	00631 General stress protein 14	-2.82	3E-11
01442 DNA topoisomerase 1	1.53	7.8E-19	01067 Ferric-pseudobactin 358 receptor precursor	-7.11	1.79E-03	01083 Ferric-pseudobactin 358 receptor precursor	2.53	2E-04
01698 Phosphoserine phosphatase	1.74	7.6E-21	01069 hypothetical protein	9.07	1.35E-32	00013 CDP-diacylglycerol--inositol 3-phosphatidyltransferase	-2.98	2E-20
01699 Transaldolase	1.61	5.5E-21	01070 Fic/DOC family protein	7.96	6.58E-06	00015 1-acyl-sn-glycerol-3-phosphate acyltransferase	-2.64	2E-12
01744 Acyl carrier protein	1.87	1.8E-23	01081 hypothetical protein	-7.57	5.39E-04	00012 hypothetical protein	-3.74	4E-14
01793 Spermidine/putrescine import ATP-binding protein PotA	2.17	1.1E-21	01083 LemA family protein	1.69	5.03E-04	01485 hypothetical protein	-6.54	3E-05
01794 Putative 2-aminoethylphosphonate transport system permease protein PhnV	1.98	1.6E-17	01098 Sensor protein KdpD	6.63	2.33E-03	00169 Cytochrome c-552 precursor	-2.60	4E-09
01913 hypothetical protein	-1.74	4.4E-07	01102 hypothetical protein	8.92	1.51E-03	00014 Phosphatidate cytidylyltransferase	-2.67	1E-14
01922 Ferrous iron transport protein B	2.71	1.1E-23	01105 Phage integrase family protein	7.50	9.80E-06	00884 Fructose-1-phosphate phosphatase YqaB	2.27	3E-11
01954 hypothetical protein	2.41	1.3E-	01205 bifunctional dihydroneopterin aldolase/dihydroneopterin triphosphate 2'-epimerase	-8.44	6.88E-05	00946 hypothetical protein	-2.50	5E-03

		21							
01955 Imelysin	1.76	4.6E-15	01216 2-aminoadipate transaminase	-2.87	2.54E-06	01443 hypothetical protein	2.61	6E-05	
01956 Fatty acid hydroxylase superfamily protein	2.09	1.8E-12	01250 ATP-dependent RNA helicase RhlE	4.19	6.38E-20	00914 hypothetical protein	2.37	6E-11	
01969 Sensor protein ZraS	1.51	1.9E-08	01263 ribonuclease H	-8.82	2.01E-05	00426 Methyl-accepting chemotaxis protein IV	-1.59	1E-05	
01976 fec operon regulator FecR	1.53	4.8E-06	01273 fec operon regulator FecR	-9.14	1.40E-05	01596 Sensor protein BasS	2.73	5E-11	
02001 Quinone-reactive Ni/Fe-hydrogenase small chain precursor	-2.02	6.1E-21	01293 putative Ni/Fe-hydrogenase B-type cytochrome subunit	-3.32	4.42E-03	01293 Putative beta-lactamase HcpC precursor	-1.76	1E-07	
02073 protoporphyrinogen oxidase	-1.63	1.9E-10	01313 Methyl-accepting chemotaxis protein II	-1.90	4.13E-05	00568 Sensor protein RstB	-1.62	2E-04	
02112 50S ribosomal protein L28	1.51	8.3E-14	01316 Hydrogenase isoenzymes formation protein HypC	-9.62	1.37E-06	02034 Tyrosine recombinase XerC	-1.80	2E-04	
02231 hypothetical protein	1.51	5.7E-06	01332 Flagellar biosynthetic protein FlhB	-9.90	1.48E-07	01175 hypothetical protein	-1.61	2E-07	
02242 MerT mercuric transport protein	-2.19	8.8E-04	01400 hypothetical protein	-6.53	5.50E-03	02288 Putative formate dehydrogenase	3.52	1E-14	
02260 Cyclic di-GMP phosphodiesterase response regulator RpfG	1.66	2.5E-	01401 hypothetical protein	-8.26	1.47E-04	01087 Phosphoethanolamine transferase EptA	3.68	1E-16	

		09								
02261 hypothetical protein	1.65	2.2E-08	01403 hypothetical protein	-7.22	1.11E-03	01484 hypothetical protein	-1.62	2E-04		
00276 Cation/acetate symporter ActP	2.41	7.4E-12	01407 hypothetical protein	-6.42	4.85E-03	01307 Sensor protein ZraS	-1.99	8E-05		
00278 Cation/acetate symporter ActP	2.54	5.8E-11	01409 Virulence sensor protein BvgS precursor	-2.90	3.19E-03	01088 PAP2 superfamily protein	4.22	1E-14		
00279 Inner membrane protein YjcH	1.56	4.0E-06	01412 hypothetical protein	-7.74	3.12E-04	01141 Lactate utilization protein C	1.64	5E-04		
00056 Isochorismatase family protein	2.33	1.7E-11	01418 hypothetical protein	7.97	1.86E-04	01508 Transcriptional activator NphR	1.59	5E-05		
00074 Fumarate reductase iron-sulfur subunit	1.51	5.2E-14	01441 IMPACT family member YigZ	-7.01	1.94E-03	01740 transport protein TonB	1.55	2E-03		
00077 NADH-quinone oxidoreductase subunit N	1.95	7.0E-21	01444 Methyl-accepting chemotaxis protein IV	2.83	8.01E-04	01008 Prophage CP4-57 integrase	-1.51	5E-05		
00078 NADH-quinone oxidoreductase subunit M	1.80	1.6E-19	01456 Lipoteichoic acid synthase 1	-3.13	1.99E-03	01149 Membrane transport protein	-2.07	2E-03		
00079 NADH-quinone oxidoreductase subunit L	1.96	1.3E-23	01467 Peptidoglycan O-acetyltransferase	6.78	1.02E-05	02315 Formate dehydrogenase iron-sulfur subunit	1.99	1E-05		
00086 NADH-quinone oxidoreductase chain 1	1.54	4.9E-	01497 von Willebrand factor type A domain protein	2.46	1.81E-03	01603 Prophage CP4-57 integrase	-2.48	8E-11		

			17								
00106 hypothetical protein	2.28	1.5E-04		01500 Bifunctional adenosylcobalamin biosynthesis protein CobP	-7.90	2.41E-04		01609 hypothetical protein	-1.61	6E-04	
00113 Putative multidrug export ATP-binding/permease protein	1.53	2.7E-02		01582 Flagellar filament 33 kDa core protein	2.83	2.14E-10		01817 CheW-like domain protein	-2.29	7E-04	
00116 Regulatory protein PchR	3.38	4.3E-25		01588 Chaperone protein YajL	1.85	4.92E-04		02111 phosphoglycerol transferase I	1.77	1E-05	
00118 Transcriptional repressor RcnR	-1.87	2.4E-13		01610 hypothetical protein	-8.06	1.56E-04		00290 Quinohemoprotein amine dehydrogenase subunit gamma	-6.10	1E-04	
00122 Cytochrome c-552 precursor	-1.74	3.7E-14		01723 Flagellar motor switch protein FliM	1.72	2.79E-03		01507 Spermidine/putrescine import ATP-binding protein PotA	1.71	1E-04	
00130 Periplasmic nitrate reductase, electron transfer subunit precursor	-1.93	4.8E-11		01724 Flagellar P-ring protein precursor	1.58	1.39E-04		01102 hypothetical protein	1.53	1E-05	
00133 Periplasmic nitrate reductase precursor	-1.72	5.8E-13		01776 hypothetical protein	-1.54	2.56E-03		01398 Chaperone protein ClpB	-2.00	4E-06	
00220 S-adenosylmethionine synthase	1.58	5.4E-17		01794 Methylisocitrate lyase	1.86	6.72E-05		02237 Regulatory protein PchR	-2.77	4E-12	
00221 Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	1.62	2.0E-16		01795 2-methylcitrate synthase	2.06	1.25E-05		00170 Cytochrome c-type protein NrfH	-2.56	5E-12	
00254 Fibrobacter succinogenes major domain (Fib_succ_major)	2.58	9.2E-		01819 transport protein TonB	-1.95	1.91E-05		01613 hypothetical protein	-2.26	6E-04	

		18							
00275 Putative nucleotidyltransferase substrate binding domain protein	2.35	7.1E-06	01820 Biopolymer transport protein ExbD	-1.84	5.30E-05	01607 hypothetical protein	-5.63	2E-03	
00308 hypothetical protein	1.61	5.7E-16	01869 50S ribosomal protein L4	-1.53	6.04E-05	01608 hypothetical protein	-1.88	2E-03	
00348 hypothetical protein	1.70	2.0E-15	01879 50S ribosomal protein L24	-1.87	1.09E-03	02148 AI-2 transport protein TqsA	-1.73	4E-03	
00350 YceI-like domain protein	2.58	1.1E-21	01908 Citrate lyase subunit beta-like protein	1.61	1.47E-03	01138 Glycolate permease GlcA	1.91	2E-06	
00358 ECF RNA polymerase sigma factor SigE	2.31	1.8E-12	01914 Methyl-accepting chemotaxis protein IV	-7.72	4.22E-04	00254 10 kDa chaperonin	-1.78	4E-09	
00487 hypothetical protein	-1.58	4.0E-03	01917 HTH-type transcriptional repressor AseR	-10.74	1.27E-11	02238 Ferrichrome receptor FcuA precursor	-2.68	2E-03	
00492 hypothetical protein	-1.81	1.4E-15	01919 putative permease	-1.71	4.84E-03	00918 Inner membrane protein YedI	1.63	1E-05	
00542 NADPH-dependent FMN reductase	-1.70	2.1E-13	01979 NnrS protein	-6.93	2.08E-03	00331 Methyl-accepting chemotaxis protein 4	-1.91	2E-04	
00687 Cytochrome c-type protein SHP precursor	-1.89	1.2E-14	02006 Putative acetyltransferase EpsM	7.63	3.33E-03				
00806 Cytochrome c-type protein TorY	-1.82	2.4E-	02007 Putative pyridoxal phosphate-dependent aminotransferase EpsN	9.02	4.83E-06				

		20			
00925 Ankyrin repeats (3 copies)	2.08	1.5E-10	02008 UDP-N-acetyl-alpha-D-glucosamine C6 dehydratase	7.85	4.28E-05
00939 Transcriptional regulatory protein WalR	2.60	1.5E-20	02012 Tyrosine-protein kinase etk	8.60	7.70E-08
00948 hypothetical protein	1.80	8.3E-12	02039 hypothetical protein	-8.78	2.10E-05
01056 Ferrichrome receptor FcuA precursor	1.75	2.3E-06	02045 hypothetical protein	2.40	1.96E-08
01193 Sulfoacetaldehyde dehydrogenase	2.07	1.6E-27	02055 Outer membrane lipoprotein Blc precursor	2.03	2.50E-07
01194 Acetolactate synthase	2.31	1.7E-32	02056 hypothetical protein	2.30	2.07E-06
01325 Isocitrate dehydrogenase [NADP]	1.57	3.1E-15	02061 Transcriptional repressor MprA	8.56	1.22E-05
01923 hypothetical protein	3.35	1.7E-16	02067 Urease accessory protein UreE	-8.09	1.42E-04
01933 hypothetical protein	1.96	1.2E-11	02068 Urease subunit alpha	-7.17	1.19E-03
01934 hypothetical protein	2.74	3.4E-	02070 Urease accessory protein UreD	-6.63	3.40E-03

		18			
01953 Imelysin	2.40	4.8E-30	02079 Bifunctional NMN adenylyltransferase/Nudix hydrolase	-8.80	2.36E-05
02050 Helix-hairpin-helix motif protein	2.86	1.4E-36	02081 Response regulator PleD	-7.55	5.32E-04
02230 hypothetical protein	1.98	1.7E-10	02082 Transcriptional activator protein CopR	-6.83	2.48E-03
			02083 Methyl-accepting chemotaxis protein IV	2.09	2.75E-04
			02129 Spore protein SP21	1.60	9.84E-05
			02141 Copper chaperone CopZ	-8.09	1.60E-04
			02153 Cell wall-associated hydrolase	2.50	3.02E-06
			02163 Methyl-accepting chemotaxis protein I	3.16	1.60E-10

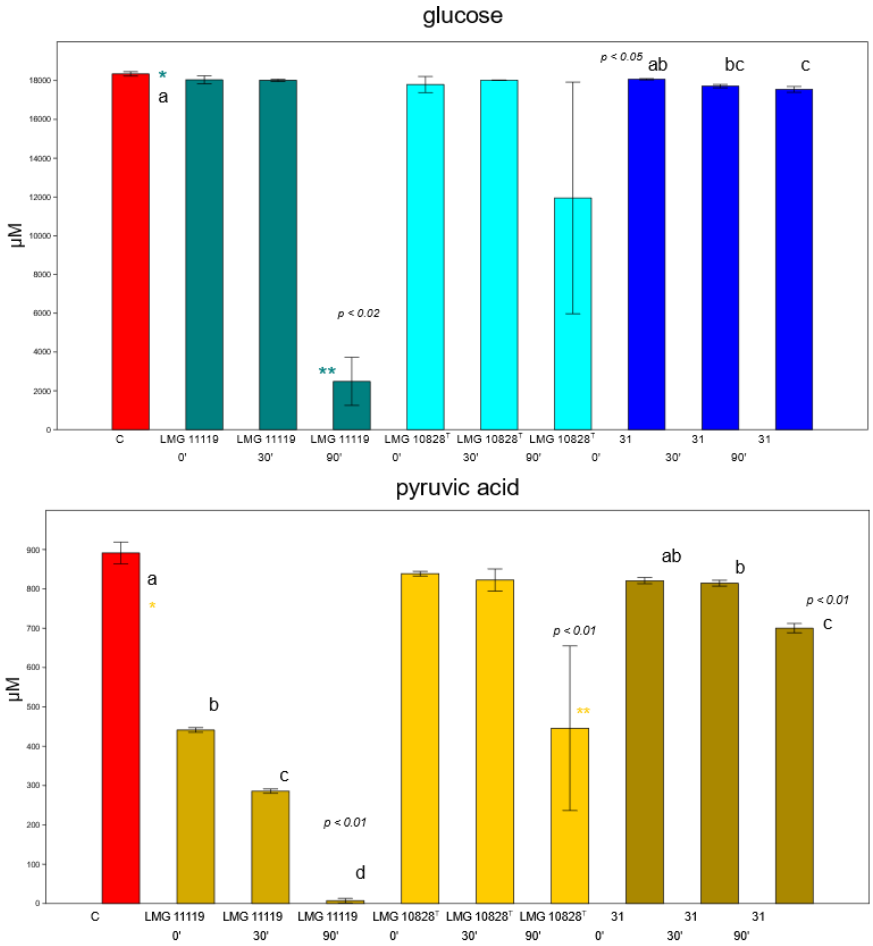
Supplementary Table 3.5. *A. butzleri* pyruvic acid and glucose related DEGs after 2 h of incubation in DMEM. The table shows logCPM, *p value* (< 0.05) and FDR (< 0.05) values of differentially expressed genes linked to pyruvate and glucose. The column "gene" shows the gene number and his locus tag. The logFC values indicated result higher than 1.5 or lower than 1.5 (except for LMG 11119 locus tag 01408 and 02015, logFC > 1.40).

gene	logFC	logCPM	PValue	FDR
Strain LMG 10828^T				
00081 Pyruvate kinase	-1.67	9.32	2.21E-12	7.53E-12
01697 UDP-glucose 6-dehydrogenase	1.64	7.11	4.45E-12	1.47E-11
02178 UTP--glucose-1-phosphate uridylyltransferase	1.97	9.28	3.70E-15	1.50E-14
02179 Glucose-6-phosphate isomerase	2.02	8.51	5.08E-16	2.18E-15
01698 dTDP-glucose 4,6-dehydratase	2.22	6.84	1.32E-17	6.40E-17
02286 Phosphoenolpyruvate carboxykinase [ATP]	2.31	7.97	3.98E-20	2.34E-19
01704 UDP-glucose 4-epimerase	2.31	6.68	9.20E-20	5.24E-19
02093 Glucose-1-phosphate thymidylyltransferase 2	2.48	5.89	6.84E-20	3.93E-19
02095 dTDP-4-amino-4,6-dideoxy-D-glucose transaminase	2.62	5.41	1.37E-20	8.38E-20
02094 dTDP-glucose 4,6-dehydratase	2.77	5.76	2.43E-22	1.65E-21
Strain LMG 11119				
01408 Phosphoenolpyruvate-carboxykinase ATP	1.44	7.29	3.91E-15	6.44E-15
02015 Pyruvate dehydrogenase E1 component	1.45	9.01	1.14E-13	1.78E-13
00007 Glucose-1-dehydrogenase-2	1.52	2.65	6.30E-04	7.37E-04
00897 Glucose-1-phosphate-cytidylyltransferase	2.59	4.36	1.17E-16	2.02E-16
00908 UDP-glucose-4-epimerase	2.70	4.98	1.44E-24	3.36E-24

02067 UTP glucose-1-phosphate uridylyltransferase	2.78	7.88	5.39E-36	2.08E-35
02068 Glucose-6-phosphate isomerase	3.53	7.06	8.78E-45	5.25E-44
02176 Glucose-1-phosphate thymidylyltransferase 1	1.66	7.52	7.86E-16	1.32E-15
02179 dTDP glucose 4,6 dehydratase	1.56	5.80	2.61E-14	4.17E-14
Strain 31				
01457 Glucose-1-phosphate thymidylyltransferase 2	2.75	5.44	3.76E-17	2.44E-16
01458 dTDP-glucose 4,6-dehydratase	2.82	4.96	2.57E-16	1.51E-15
01459 dTDP-4-amino-4,6-dideoxy-D-glucose transaminase	2.65	3.31	3.71E-09	9.29E-09
01703 Phosphoenolpyruvate carboxykinase [ATP]	2.29	8.35	9.15E-14	3.7E-13
01809 UDP-glucose 4-epimerase	2.54	3.49	7.26E-09	1.78E-08
02010 dTDP-glucose 4,6-dehydratase	2.36	3.39	1.74E-08	4.1E-08
02041 UTP--glucose-1-phosphate uridylyltransferase	2.18	9.57	2.31E-10	6.62E-10
02042 Glucose-6-phosphate isomerase	2.48	8.76	1.46E-15	7.49E-15

Supplementary Table 3.6. Annotation statistics of the 3 *A. butzleri* strains. In the table are indicated the genome size, coverage ((read count * read length)/genome size), number of total genes, number of total CDS, tRNA and hypothetical proteins number.

Strain code	LMG 11119	LMG 10828^T	31
Genomes size (Mbp)	2.3	2.31	2.13
coverage (X)	195.32	205.69	486.87
GC content (%)	26.88	26.88	26.98
Number of contings	51	26	25
total genes	2278	2317	2164
CDS	2236	2271	2120
tRNA	41	45	43
hypothetical proteins	535	553	491



Supplementary figure 3.1. Glucose and pyruvic acid concentration of DMEM inoculated with *A. butzleri* strains. The bar chart shows the concentrations in DMEM of glucose and pyruvic acid (µM). In the figure are indicated the strain codes and control (normal DMEM, C). The different sampling times are indicated as 0' (after acclimation), 30' and 90'. The error bars represent the standard errors (Past3). The figure shows statistical analysis *p*-value. The statistical differences between strains are indicated near bars.

Supplementary table 1. General information about *Arcobacteraceae* genomes. In the table are indicated different information about genomes general information (Quast), coverage (reads nr. * reads length)/genome size) and CRISPR/CAS sequences number. In the first line under the name of the strains the group to which they belong is indicated.

	A. bivalvium LMG26154^T	*A. butzleri* LMG10828^T	*A. cibarius* LMG 21996^T	*A. cryaerophilus* LMG24291^T	*A. cryaerophilus* LMG10829	*A. ellisii* LMG26155^T	*A. faecis* LMG28519^T	*A. halophilus* CCUG53805^T
group	4	1	1	1	1	2	1	3
nr. contigs (>= 0 bp)	23	27	70	29	27	42	65	49
nr. contigs (>= 1000 bp)	22	23	60	23	23	37	56	41
nr. contigs (>= 5000 bp)	17	20	40	21	22	27	33	29
nr. contigs (>= 10000 bp)	15	18	34	19	19	23	30	23
nr. contigs (>= 25000 bp)	13	16	23	16	15	20	23	23
nr. contigs (>= 50000 bp)	12	11	10	14	12	17	16	17
Total length (>= 0 Mbp)	2.66	2.30	2.15	2.08	2.01	2.75	2.39	2.76
Total length (>= 1000 Mbp)	2.66	2.30	2.14	2.08	2.01	2.75	2.38	2.75
Total length (>= 5000 Mbp)	2.65	2.29	2.09	2.07	2.01	2.72	2.34	2.72
Total length (>= 10000 Mbp)	2.64	2.28	2.05	2.06	1.99	2.70	2.32	2.68
Total length (>= 25000 Mbp)	2.61	2.25	1.86	2.01	1.93	2.65	2.22	2.68
Total length (>= 50000 Mbp)	2.58	2.04	1.39	1.92	1.83	2.54	1.97	2.46
Total length Mbp	2.66	2.30	2.15	2.08	2.01	2.75	2.39	2.76
coverage	353	217	197	441	471	277	348	333
N50	247420	234980	122275	194390	165914	177151	123980	177704
N75	178940	112969	35326	101744	104157	110030	75048	88435
L50	4	4	6	5	4	6	7	6
L75	7	8	16	9	8	11	14	12

GC (%)	28	26.87	26.78	27.19	27.39	26.76	26.96	27.43
CRISPR	0	0	2	2	1	5	3	5
CAS	2	0	4	0	1	1	3	4
CRISPR/CAS associated sequences	0	0	1	0	0	0	1	1

	A. lacus LMG29062T	*A. lanthieri* LMG28516T	*A. molluscorum* LMG25693T	*A. mytili* LMG24559T	*A. nitrofigilis* LMG 7704T	*A. porcinus* LMG2248 7T	*A. skirrowii* LMG6621T	*A. suis* LMG26152
group	1	1	3	3	5	1	1	2
nr. contigs (>= 0 bp)	23	30	68	59	1	29	19	57
nr. contigs (>= 1000 bp)	20	25	54	50	1	28	18	52
nr. contigs (>= 5000 bp)	18	18	36	42	1	24	14	36
nr. contigs (>= 10000 bp)	17	17	34	35	1	19	11	30
nr. contigs (>= 25000 bp)	15	14	31	29	1	15	9	22
nr. contigs (>= 50000 bp)	14	11	22	17	1	10	8	17
Total length (>= 0 Mbp)	2.22	2.24	2.73	2.97	3.19	1.79	1.96	2.58
Total length (>= 1000 Mbp)	2.22	2.23	2.72	2.97	3.19	1.79	1.96	2.58
Total length (>= 5000 Mbp)	2.21	2.22	2.67	2.95	3.19	1.79	1.95	2.54
Total length (>= 10000 Mbp)	2.21	2.21	2.66	2.90	3.19	1.75	1.93	2.51
Total length (>= 25000 Mbp)	2.18	2.17	2.61	2.81	3.19	1.68	1.89	2.35
Total length (>= 50000 Mbp)	2.13	2.04	2.28	2.41	3.19	1.49	1.84	2.19
Total length Mbp	2.22	2.24	2.73	2.97	3.19	1.79	1.96	2.58
coverage	75	365	464	382	Ref. genome	74	362	342
N50	207786	369058	88441	164261	3192235	140725	305996	142035
N75	107947	91794	60477	77475	3192235	75594	205942	75047
L50	5	2	9	7	1	4	2	14
L75	8	7	18	15	1	8	4	7

GC (%)	26.78	26.41	26.03	26.35	28.36	27.17	27.66	27.23
CRISPR	0	2	4	1	1	1	2	1
CAS	0	3	4	7	1	1	1	6
CRISPR/CAS associated sequences	0	1	2	1	1	0	1	1

	A. thereius LMG24486^T	*A. trophiarum* LMG25534^T	*A. vandammei* LMG31429^T	*A. venerupis* LMG26156^T	*A. vittoriensis* LMG30050^T	*C. jejuni* NCTC11168	*H. pylori* MT5135
group	1	1	1	2	1	-	-
nr. contigs (>= 0 bp)	9	35	56	89	40	1	1
nr. contigs (>= 1000 bp)	8	31	45	72	36	1	1
nr. contigs (>= 5000 bp)	7	27	34	57	32	1	1
nr. contigs (>= 10000 bp)	5	25	29	52	29	1	1
nr. contigs (>= 25000 bp)	5	19	21	38	24	1	1
nr. contigs (>= 50000 bp)	4	13	15	26	18	1	1
Total length (>= 0 Mbp)	1.90	1.87	2.21	3.16	2.42	1.64	1.62
Total length (>= 1000 Mbp)	1.90	1.87	2.20	3.15	2.41	1.64	1.62
Total length (>= 5000 Mbp)	1.90	1.86	2.17	3.12	2.41	1.64	1.62
Total length (>= 10000 Mbp)	1.89	1.85	2.13	3.08	2.39	1.64	1.62
Total length (>= 25000 Mbp)	1.89	1.74	1.99	2.84	2.31	1.64	1.62
Total length (>= 50000 Mbp)	1.85	1.54	1.76	2.42	2.11	1.64	1.62
Total length Mbp	1.90	1.87	2.21	3.16	2.42	1.64	1.62
coverage	59	253	215	149	339	Ref. genome	Ref. genome
N50	509145	122603	107602	82335	127074	1641481	1615199
N75	492236	63280	68457	53970	81644	1641481	1615199
L50	2	5	8	14	8	1	1
L75	3	11	14	26	14	1	1

GC (%)	26.92	28.04	27.6	27.17	27.02	30.55	39.28
CRISPR	5	0	2	0	1	2	1
CAS	2	0	3	5	3	1	0
CRISPR/CAS associated sequences	1	0	1	0	0	1	0

Supplementary table 4.2. Clusters of Orthologous Genes functions number. The table shows the number of annotated orthogroups COGs obtained from EggNOG mapper analysis. At the end of the table are indicated the coding of the letters relating to the different classes.

	A	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	S	T	U	V	Z
A. bivalvior um	1	193	25	2 0 4 1	85	7 8	13 5	5 5	1 5 8 1	12 2	99	14 5	80	1 0 1	14 9	36	6 8 3	4 8 1	5 9	3 2	0
A. butzleri	1	154	26	5 9 1	76	7 3	11 3	4 5	5 2 1	11 8	97	15 3	84	8 7	13 6	27	6 3 3	7 3 1	5 4	3 3	1
A. cibarius	0	137	25	3 4 1	73	6 0	11 4	4 5	5 5 1	80	15 5	14 0	73	8 8	11 0	20	0 9 3	5 8 1	6 2	2 6	0
A. cryaerop hilus (T)	0	133	26	3 4 1	73	6 5	11 9	5 5	5 2 1	78	10 9	13 8	76	8 3	10 6	30	1 8 3	2 8 1	5 7	2 7	0
A. cryaerop hilus	0	130	28	3 9 1	72	6 2	11 5	5 2	5 4 1	74	97	13 9	78	8 2 1	10 8	30	1 4 4	3 3 2	4 7	2 4	0
A. ellisii	1	208	29	6 8 1	81	7 0	12 4	5 7	5 9 1	12 8	14 0	15 6	84	1 3	14 2	40	2 4 3	3 1 1	5 7	3 4	1
A. faecis	0	155	26	4 5 1	75	6 3	12 0	5 3	5 3 1	11 3	12 4	15 7	90	8 3	14 0	29	5 7 3	8 8 2	5 1	2 8	0
A. halophilu s	1	225	25	8 9 1	83	7 8	14 1	5 7	5 7 1	11 5	10 7	13 2	88	8 8	17 1	33	8 1 3	5 0 1	5 7	3 3	0
A. lacus	1	138	22	6 6 1	75	7 2	10 2	6 9	5 1 1	11 8	10 4	13 1	71	8 9	13 0	43	3 2 3	5 0 1	4 6	3 4	1
A. lanthieri	1	137	24	5 4 1	71	6 8	11 6	5 0	5 0 1	10 3	11 8	14 0	87	8 5 1	16 8	31	2 0 4	5 1 2	5 6	2 5	0
A. molluscor um	1	235	19	9 9 1	79	7 5	14 3	5 2	5 8 1	12 6	11 4	14 5	79	0 0	14 1	31	1 3 3	1 7 2	6 1	3 4	0
A. mytili	1	223	24	9 9 2	80	7 7 1	14 3	6 6	5 4 1	12 1	11 1	16 0	84	9 4 1	16 7	46	7 2 4	4 7 2	4 9	3 5	0
A. nitrofigili s	0	265	23	5 6 1	92	2 8	15 3	7 2	6 3 1	16 4	12 5	18 6	88	1 2	20 8	60	9 6 2	5 1	7 5	3 8	0
A. porcinus	0	122	23	3 4 1	69	5 4	11 2	4 3	5 3 1	73	98	11 3	62	7 9	96	16	6 8 3	9 3 1	4 9	1 7	0
A. skirrowii	0	126	24	3 6 1	70	5 9	11 7	4 0	4 9 1	74	10 9	12 0	69	8 5	11 4	20	1 5 3	0 5 2	4 8	1 8	0
A. suis	1	210	23	6 4 1	78	7 2	13 0	6 4	6 1 1	12 2	10 2	14 8	81	9 7	13 3	40	8 8 2	0 2	5 5	3 2	1
A. thereius	0	120	25	3 3	70	5 5	11 5	3 8	5 3	82	11 4	11 0	65	8 0	10 7	18	7 1	9 5	4 4	2 4	0

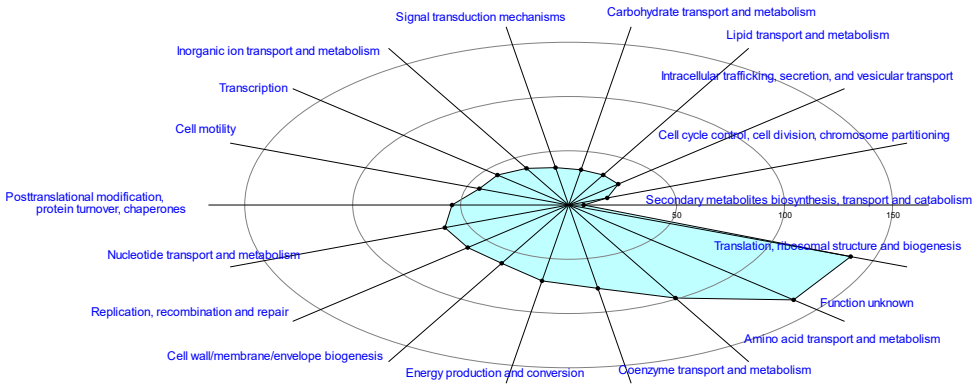
A. trophiarum	0	136	23	1 3 2	71	6 1	11 6	5 0	1 4 9	73	10 1	12 4	68	8 0	10 2	25	2 7 7	1 0 5	5 5	2 3	0
A. vandammei	0	135	27	1 3 2 2	71	6 2 1	11 7	5 2	1 5 5 1	90	13 8	13 3	62	7 6 1	10 8	24	3 3 4 4	1 2 4 2	5 4	4 3	1	0
A. venerupis	1	253	28	3 3	87	0 2	14 3	6 6	6 7	15 3	11 1	18 3	103	0 9	18 4	50	6 4	5 4	7 9	5 2	1
A. vittoriensis	1	140	27	1 6 4	74	7 4	11 9	5 3	1 5 1	13 0	10 0	15 3	75	8 6	19 5	32	3 4 0	1 5 2	5 5	3 2	0

A RNA processing and modification	J Translation, ribosomal structure and biogenesis
C Energy production and conversion	K Transcription
D Cell cycle control, cell division, chromosome partitioning	L Replication, recombination and repair
E Amino acid transport and metabolism	M Cell wall/membrane/envelope biogenesis
F Nucleotide transport and metabolism	N Cell motility
G Carbohydrate transport and metabolism	O Posttranslational modification, protein turnover, chaperones
H Coenzyme transport and metabolism	P Inorganic ion transport and metabolism
I Lipid transport and metabolism	Q Secondary metabolites biosynthesis, transport and catabolism
S Function unknown	V Defense mechanisms
T Signal transduction mechanisms	Z Cytoskeleton
U Intracellular trafficking, secretion, and vesicular transport	

Supplementary table 4.3. Presence/absence of secondary metabolites related genes. The table shows presence (1) and absence (0) of secondary metabolites related sequences in the 21 genomes object of study.

	ectoine	RiPP-like	arylpolyene-resorcinol	ranthipeptide	thiopeptide	resorcinol	Acyl aminoacids	NRPS-T1PKS	redox cofactor	arylpolyene
A. bivalviorum LMG26154^T	1	1	1	0	0	0	0	0	0	0
A. ellisii LMG26155^T	1	0	0	1	0	0	0	0	0	0
A. halophilus CCUG53805^T	1	1	0	0	0	0	0	0	0	0
A. lathieri LMG28516^T	0	0	0	0	0	0	0	0	0	0
A. molluscorum LMG25693^T	1	0	0	1	0	0	0	0	0	0
A. mytili LMG24559^T	1	0	1	0	0	0	0	0	0	0
A. nitrofigilis LMG7704^T	1	0	1	0	0	0	0	1	1	0
A. suis LMG26152^T	0	0	0	1	0	0	0	0	0	0
A. trophiarum LMG25534^T	0	0	1	0	0	0	0	0	0	0
A. vandammei LMG31429^T	0	0	0	0	0	0	0	0	0	0
A. venerupis LMG26156^T	1	0	0	1	0	0	0	0	1	0
A. vittoriensis LMG30050^T	1	0	0	0	0	0	0	0	0	1
A. butzleri LMG10828^T	0	0	0	1	1	0	0	0	0	0
A. cibarius LMG21996^T	0	0	0	0	1	0	0	0	0	0
A. cryaerophilus LMG24291^T	1	0	1	0	1	1	1	0	0	0

A. cryaerophilus LMG10829	0	0	1	0	1	0	0	0	0	0
A. faecis LMG28519^{T}	0	0	0	0	1	0	1	0	0	0
A. lacus LMG29062^{T}	0	0	1	1	1	0	0	0	0	0
A. porcinus LMG24487^{T}	1	0	0	0	1	0	0	0	0	0
A. skirrowii LMG6621^{T}	1	0	0	0	1	0	0	0	0	0
A. thereius LMG24486^{T}	1	0	0	0	1	0	0	0	0	0



Supplementary figure 4.1. Orthogroups pathway present in all genomes. The figure shows the number of COGs codes relative to orthogroups present in all *Arcobacteraceae* species object of study.

7. Thesis Summary

7.1 PhD Thesis Summary

English version

The *Arcobacteraceae* family includes Gram-negative bacteria isolated from different environmental matrices and animals such as sewage, oil production environments, marine sediments, estuarine and river waters, oysters, snails, tube worms (abyssal annelid) and fish farms. The species belonging to *Arcobacteraceae* have been isolated from terrestrial animals, in particular from chickens, pigs, cattle. Some *Arcobacteraceae* species are considered food-borne pathogens as they have been isolated from different types of foods, though predominantly from food of animal origin and from human clinical cases. The species *Arcobacter butzleri* and *Arcobacter cryaerophilus* have been isolated from stool of humans with gastrointestinal diseases. The different ability of the species included in the *Arcobacteraceae* family to survive in different hosts and environments suggest an evolutionary pressure with consequent variation in genome content. Moreover, their different physiological and genomic characteristics have led to the recent proposal to subdivide the *Arcobacteraceae* family into different genera, which has been criticized due to the lack of biological and clinical significance.

The pathogenicity of some *Arcobacteraceae* species was already studied using *in vitro* assays. The *in vitro* cell models represent a valid instrument for the first host-pathogen interaction studies on a large strains number. These models can consist in a single cell line or several cell lines with different characteristics. The *in vitro* cell models are often used in the evaluation of bacterial adhesion and invasion and cytotoxic effects.

The cell models used for the study of intestinal pathogen are often produced with cells of tumor origin, as they are more easily cultured. The

Caco-2 and HT29-MTX-E12 cell lines have been used in the production of mixed models characterized by some significant intestinal characteristics. Caco-2 models were used for the evaluation of host colonization of bacterial pathogens such as *Campylobacter jejuni*, *Campylobacter coli*, *Salmonella enterica* and *Escherichia coli*. The HT29-MTX-E12 cell line (sub-homogeneous human intestinal mucus-producing cells) is derived from the parental line HT29. HT29-MTX-E12 cells. This cell line is able to produce mucus exhibiting a phenotype like globose cells with a predominant expression of mucins MUC5AC. The mucus represents an element produced by epithelial tissues which affects the behavior of pathogenic and non-pathogenic bacteria.

The phenotypic and physiological bacterial carachteristics can be linked to genome characteristics (e.g., presence or absence of specific genes). In the case of *Arcobacter* spp. DNA sequencing techniques have been used to characterize new species, to obtain information about specific sequences (e.g., 16s rRNA) and to design new molecular techniques. The sequencing techniques are used to study the whole genome bacterial gene content. Moreover, the RNA sequencing (RNA-seq) allows an evaluation of the entire transcriptome. The information obtained from the number of transcripts per gene are normalized and elaborated with statistical packages to evaluate the differentially expressed genes (DEGs) overexpressed and under-expressed. This method can be used in the gene expression evaluation of an organism in contact with another organism, allowing the transcriptome study of a single or multiple organisms.

The present Ph.D. thesis takes into consideration the study of *Arcobacteraceae* species and in particular of *A. butzleri* at different levels. The goal was to obtain genomic and transcriptomic information of

different *A. butzleri* strains. The choice to focus the analyzes on *A. butzleri* was based on higher isolation rate of this specie from different animals, foods and clinical cases compared to other *Arcobacteraceae*, as reported in literature. A comparative genomics analysis of 32 *A. butzleri* strains is reported in Chapter 2 jointly with the evaluation of colonization and invasion in contact with *in vitro* mucus producer and not mucus producer cell models. This experimental approach was followed to explore possible links between the colonization and invasion of *A. butzleri* and its genome content. Furthermore, the presence of putative virulence genes and their possible correlation with the strain isolation source was investigated. The genomes analysis was followed by transcriptomic analysis to evaluate the function of *A. butzleri* genes. More specifically, in chapter 3 is presented the global gene expression profile of three *A. butzleri* strains evaluated with RNA-seq. The strains were selected considering data about their adhesion and invasion ability. The gene expression was evaluated in bacterial cells in contact with a human mucus-producing gut model after 30 and 90 minutes. In chapter 4 the genome analysis of twenty *Arcobacteraceae* species recently reclassified at genus level is presented. The specific aim was to obtain genomic information about the whole bacterial family.

In Chapter 5 general conclusions are presented to show the advancements regarding *Arcobacter* spp. knowledge from the presented data.

The study shown in chapter 2 was performed to elucidate the genetic background of 32 *A. butzleri* strains of diverse origin. This study was performed to explore possible correlation between *A. butzleri* genomes and the ability to colonize and invade human intestinal cells *in vitro* through complementary use of comparative genomics and physiological tests on

human gut models. The simulated infection of human intestinal models showed a higher colonization rate in presence of mucus- producing cells (HT29-MTX-E12). For a part of strains, human mucus significantly improved the resistance to physical removal from the *in vitro* mucosa, while short time-frame growth was even observed. Pangenome analysis highlighted a hypervariable accessory genome. This hypervariable accessory genome was not strictly correlated to the *A. butzleri* isolation source. Likewise, the strain phylogeny was unrelated to their shared origin, while a certain degree of segregation was observed among strains isolated from different segments of the pig intestinal tract. The putative virulence genes detected were mostly encompassed in the accessory genome. The liposaccharides (LPS) biosynthesis and in particular the chain glycosylation of the O-antigen is harbored in a region of high plasticity of the pangenome. This aspect would indicate frequent horizontal gene transfer phenomena, as well as the involvement of this hypervariable structure in the adaptive behavior of *A. butzleri*. The results deepen the current knowledge on *A. butzleri* pangenome extending the pool of genes considered virulence markers and providing bases to develop new diagnostic approaches for the detection of those strains with a higher virulence potential.

The chapter 4 shows *A. butzleri* transcriptome RNA sequencing evaluation. Recent advancements in the genomes analysis of *A. butzleri* highlighted putative virulence genes and possible mechanisms of action. It is therefore now possible and relevant the transcriptome study to explore possible virulence mechanisms under conditions that mimic the infection process. The RNA-seq study focused on the transcriptome assessment of three *A. butzleri* strains isolated from human stool and displaying variable

in vitro virulence potential in the previous study (chapter 2). The study of the transcriptome was performed with the use of an *in vitro* human gut mucus producer cell model (Caco-2/HT29-MTX-E12). The simulated colonization and invasion conditions allowed the obtainment of physiological and expression data in parallel. The ability of all *A. butzleri* strains to colonize the *in vitro* human gut cells was confirmed. The transcriptome results identified the overexpression of genes currently considered putatively associated to *A. butzleri* virulence among which *irgA*, *oprF* and *iroE*. These three genes are currently used in the study of *A. butzleri*. A general gene overexpression was observed in *A. butzleri* incubated with DMEM especially in the strain LMG 11119 suggesting its greater adaptive capacity to environmental conditions Moreover, genes not currently considered *A. butzleri* virulence genes resulted differentially expressed during cell model colonization. The functions of these genes turned out to be several. Relevant differentially expressed genes (DEGs) were involved in organic acid metabolism (e.g. *actP* and *yjcH*). These genes were already linked to *Escherichia coli* virulence. Other relevant DEGs detected were related to iron transport. The strains LMG 11119 that showed a higher colonization led to an overexpression of the genes *tonB*, *exbB* and *exbD* linked to iron assimilation and proliferation in other Gram negative bacteria (*E. coli* and *S. enterica*).

The aim of the study presented in the chapter 4 was to assess the *Arcobacteraceae* pangenome and to characterize potential differences present in 20 validly described species. The analyzes have been conducted on the whole genomes of the corresponding type strains obtained by Illumina sequencing, applying different bioinformatic tools to compare different outputs. The results revealed the presence of pangenome

partitions with respective numbers of orthogroups and genes comparable to other Gram-negative bacteria genera such as *Campylobacter* spp., suggesting a taxonomic classification into a single genus, as originally proposed. Differences were present between genomes of species associated with animals and human clinical cases and those isolated from other sources, such as water. Furthermore, a smaller genome size has been observed in the animal and human associated species. Different gene class compositions were observed in animal and human-associated species with a higher percentage of virulence-related gene classes such as cell motility genes. The data of this study does not support the proposed division into different genera, though it does identify an adaptation to environmental and/or host conditions of some species.

The results suggest a division into pathogenic and non-pathogenic species to focus future studies to those species with a clinical relevance and impact on food safety and public health.

In conclusion, *A. butzleri* showed the ability to overcome the human mucus barrier. The open pangenome of *A. butzleri* and the interchangeability of potential virulence genes has been observed in the 32 strains analyzed. These aspects have been proposed as key genomic features for the host adaptation of *A. butzleri*. The functional annotation of *A. butzleri* genomes showed the presence of putative virulence genes and antigen recognition markers. The *A. butzleri* strains from different hosts showed a similar colonization ability *in vitro*, without a marked invasiveness. However, part of the strains showed a different host colonization ability compared to other strains. These strains have been chosen for subsequent transcriptomics studies in contact with host cells.

The involvement of some genes currently considered virulence-associated by genomic analyzes has been confirmed by RNA-seq data. A fast response to environmental stimuli was observed in the strain LMG 11119 that showed a greater colonization. This aspect suggests an important role of environment-related gene expression response that could also play a role in adaptation under host colonization conditions. The importance of iron metabolism during *A. butzleri* infections has been suggested from the observation of overexpressed genes linked to iron transport in the strain characterized by higher colonization (LMG 11119). The differential gene expression of organic acid-related genes was confirmed by chemical analysis (High Pressure Liquid Chromatography). The transcriptome data allowed to understand the role of genes currently considered virulence-related from functional annotation studies.

The studies presented in this thesis were performed on simplified *in vitro* models. Although the gut characteristics are more complex compared to the cellular models, these data can be used for subsequent targeted *in vitro* and *in vivo* studies.

The *Arcobacteraceae* species are characterized by different sources of isolation and association with clinical relevance. This aspect may be linked to a different genome size and gene content among species. The *Arcobacteraceae* genome partitions (core, cloud, and soft genome) do not differ from other species belonging to other bacterial genera like *Campylobacter* spp., suggesting the existence of only one bacterial genus "*Arcobacter*". The smaller genome size of species considered animal and human clinical-related compared to environment related species was observed. This aspect led to detect groups-specific gene classes suggesting an evolutionary specialization of some species to animal hosts. The smaller

genomes size of clinical and animals related species leads to hypothesize a link between genome size pathogenicity of *Arcobacter* spp.. The maintaining and presence of specific sequences demonstrate the importance of pathways for specific bacterial metabolisms (e.g., cellulose inhibition in *A. nitrofigilis*). This aspect suggests the existence of sequences related to various bacterial lifestyle in different environmental niches. The study of *Arcobacteraceae* pangenome allowed the detection of some specific sequences linked to ecological niches. The increase in the coming years of genomic sequences and on *Arcobacteraceae* information will be useful to better understand their environmental and clinical role.

Dutch version (Samenvatting)

De familie *Arcobacteraceae* omvat Gram-negatieve, aerotolerante, staafvormige bacteriën die aanwezig zijn in zeer diverse matrices zoals riool, estuarium en rivier water, mariene sedimenten, zeevruchten, slakken en wormen. Sommige *Arcobacteraceae* species zijn algemeen aanwezig in het gastro-intestinaal stelsel van zoogdieren en vogels, in het bijzonder in landbouwhuisdieren als varkens, runderen en kippen. Sommige *Arcobacteraceae* species worden beschouwd als door voedsel overgedragen ziekteverwekkers omdat ze vaak worden geïsoleerd vanuit voedsel van dierlijke oorsprong. Twee species, *Arcobacter butzleri* en *Arcobacter cryaerophilus* worden frequent geïsoleerd uit stoelgang van mensen met gastro-intestinale infecties. Het verschillend vermogen van de species in de familie *Arcobacteraceae* om al dan niet te overleven in verschillende gastheren en diverse omgevingsomstandigheden suggereert een evolutionaire druk met een te verwachten variatie van het bacterieel genoom. De grote diversiteit aan fenotypische kenmerken tussen de species heeft recent ook geleid tot een voorstel om de species onder te verdelen in verschillende nieuwe genera, maar dit wordt vooralsnog sterk bekritiseerd door een gebrek aan biologische en klinische relevantie.

De pathogeniciteit van sommige *Arcobacteraceae* species werd al onderzocht met behulp van eenvoudige *in vitro* assays. Gelet dat *Arcobacteraceae* apathogeen zijn voor dieren vormen deze in vitro cel modellen een eerste belangrijk instrument voor interactie studies tussen gastheer en ziekteverwekker en om de diversiteit op stam niveau in te schatten. Deze modellen kunnen bestaan uit één enkele cellijn of meerdere cellijnen met verschillende kenmerken, en kunnen worden gebruikt bij de studie van bacteriële adhesie, invasie en cytotoxische effecten. De cel

modellen die worden gebruikt in het onderzoek naar darmpathogenen worden vaak uitgevoerd met cellijnen van tumorale oorsprong, omdat ze gemakkelijker te kweken zijn. De Caco-2 en HT29-MTX-E12 cellijnen zijn al gebruikt bij de constructie van gemengde modellen waardoor enkele significante darmkenmerken kunnen worden gesimuleerd. Caco-2 modellen werden gebruikt voor de evaluatie van gastheerkolonisatie door bacteriële pathogenen zoals *Campylobacter jejuni, C. coli, Salmonella enterica* en *E. coli.* De HT29-MTX-E12-cellijn (sub-homogene humane darmslijm producerende cellen) is afgeleid van de ouderlijn HT29, en zijn in staat mucus te produceren met een fenotype zoals *in vivo* mucus-producerende cellen. Mucus vertegenwoordigt een belangrijk element dat wordt geproduceerd door epitheliale weefsels en dat het gedrag van pathogene en niet-pathogene bacteriën sterk beïnvloedt.

Fenotypische bacteriële eigenschappen vinden hun oorsprong in genomische kenmerken, zoals bijvoorbeeld door de aan- of afwezigheid van specifieke genen. Ook voor verschillende *Arcobacter* species werd genoom DNA-sequencing toegepast om nieuwe soorten te karakteriseren, om informatie te verkrijgen over specifieke sequenties (bijv. 16s rRNA) en om de genoom inhoud te bestuderen. Daarnaast maakt RNA-sequencing (RNA-seq) een evaluatie van het volledige transcriptoom mogelijk. De informatie die wordt verkregen uit het aantal transcripten per gen wordt genormaliseerd en statistisch geëvalueerd om de differentieel tot expressie gebrachte genen (DEG's) te evalueren die tot over- en onderexpressie worden gebracht. Deze methode kan worden gebruikt om de genexpressie onder verschillende condities van een organisme te bestuderen, waardoor ondermeer de transcriptoom evaluatie van één enkele of meerdere organismen mogelijk is.

In het proefschrift worden de klinisch relevante *Arcobacter* species bestudeerd met een focus op *Arcobacter butzleri*. Het laatste met als doel om genomische en transcriptomische informatie te verkrijgen en te evalueren van verschillende *A. butzleri* stammen. Deze keuze voor het species werd ingegeven door de hogere prevalentie van het species in landbouwhuisdieren, voedingsmiddelen en klinisch humane gevallen in vergelijking met andere *Arcobacteraceae* species.

Na een literatuur overzicht in Hoofdstuk 1, omvat Hoofdstuk 2 een vergelijkende genomische analyse van 32 *A. butzleri*-stammen samen met de evaluatie van hun kolonisatie en invasie capaciteit in contact met mucus producerende en niet-mucus producerende in vitro cel modellen. Deze experimentele benadering werd gevolgd om mogelijke verbanden tussen kolonisatie en invasie capaciteit en genoominhoud te onderzoeken. Verder werd de aanwezigheid van vermeende virulentiegenen en hun mogelijke correlatie met de bron van isolatie onderzocht. De genoom analyse werd gevolgd door transcriptoom analyse (Hoofdstuk 3) om de functie van verschillende genen geannoteerd op de *A. butzleri* genomen te evalueren. Meer specifiek werd het globale genexpressieprofiel door RNAseq van drie *A. butzleri* stammen geanalyseerd. De stammen werden geselecteerd op basis van hun adhesie- en invasievermogen, en de genexpressie werd geëvalueerd na contact met een humaan mucus producerend darmmodel na 30 en 90 minuten. In hoofdstuk 4 werd de genoomanalyse van 20 *Arcobacteraceae* species gerapporteerd. Het doel van deze studie was om het pangenoom af te lijnen en om informatie te verkrijgen over de genomische kenmerken en diversiteit van deze bacteriële familie. In hoofdstuk 5 worden algemene conclusies geformuleerd en de bijdrage van

het doctoraal onderzoekswerk aan de kennis van *Arcobacteraceae* gepresenteerd.

De studie in **hoofdstuk 2** werd uitgevoerd om de genetische achtergrond van een set van 32 *A. butzleri* stammen van diverse oorsprong op te helderen en om de correlatie te onderzoeken met het vermogen om humane darmcellen in vitro te koloniseren en binnen te dringen door complementair gebruik van vergelijkende genomische analyse en fenotypische testen met behulp van meer complexe humane in vitro cel darmmodellen. De gesimuleerde infectie van humane darmmodellen toonde een hogere kolonisatiesnelheid in aanwezigheid van mucus producerende cellen (HT29-MTX-E12). Voor een deel van de stammen verbeterde humane mucus significant de weerstand tegen fysieke verwijdering van het *in vitro* slijmvlies, terwijl er zelfs sprake was van korte tijdsgroei. Pangenoom analyse bracht daarnaast een hypervariabel accessoire genoom aan het licht. Dit hypervariabele accessoire genoom is niet strikt gecorreleerd met de isolatiebron. Evenzo bleek de stamfylogenie niet gerelateerd aan hun gedeelde oorsprong, terwijl een zekere mate van segregatie werd waargenomen tussen stammen die waren geïsoleerd uit verschillende segmenten van het darmkanaal van varkens. De vermeende virulentiegenen die in alle stammen werden gedetecteerd, waren meestal aanwezig in het accessoire genoom. De LPS-biosynthese en in het bijzonder de ketenglycosylering van het O-antigeen bevindt zich in een gebied met hoge plasticiteit van het pangenoom. Dit aspect kan wijzen op frequente horizontale gen overdracht, als ook op de betrokkenheid van deze hypervariabele structuur in het adaptieve gedrag van *A. butzleri*. De resultaten van deze studie verdiepen de huidige kennis over het *A. butzleri* pangenoom door de pool van genen die als virulentiemarkers worden

beschouwd uit te breiden en bieden een basis om nieuwe diagnostische benaderingen te ontwikkelen voor de detectie van stammen met een hoger virulentiepotentieel.

Hoofdstuk 3 rapporteert een *Arcobacter butzleri* transcriptoom RNA sequencing evaluatie. Recente ontwikkelingen in de genoom analyse van *A. butzleri* identificeerden vermeende virulentiegenen en mogelijke werkingsmechanismen. Het is daarom nu mogelijk en relevant om het transcriptoom te bestuderen om mogelijke virulentiemechanismen te onderzoeken onder omstandigheden die het infectieproces nabootsen. De RNA-seq studie richtte zich op de transcriptoom beoordeling van drie *A. butzleri* stammen geïsoleerd uit menselijke stoelgang en die variabiliteit in vitro virulentiepotentieel vertoonden in de vorige studie (hoofdstuk 2). De studie van het transcriptoom werd uitgevoerd met een in vitro humaan darmmucus producerende celmodel (Caco-2/HT29-MTX-E12) en onder kolonisatie en invasie omstandigheden die het mogelijk maakten om parallel fysiologische gegevens te verkrijgen. Het vermogen van alle *A. butzleri* stammen om in vitro menselijke darmcellen te koloniseren werd bevestigd. De transcriptoom resultaten identificeerden de overexpressie van genen die momenteel verondersteld worden geassocieerd te zijn met de virulentie, waaronder *irgA*, *oprF* en *iroE*. Een algemene overexpressie van genen werd waargenomen in *A. butzleri* geïncubeerd met DMEM, vooral in de LMG 11119 stam, wat suggereert dat het een groter aanpassingsvermogen heeft aan omgevingsomstandigheden. Relevante differentieel tot expressie gebrachte genen (DEG's) die betrokken zijn bij het metabolisme van organische zuren, zoals actP en yjcH, die al zijn gekoppeld aan de virulentie van *Escherichia coli*, werden ook gedetecteerd. Andere relevante gedetecteerde DEG's waren gerelateerd

aan ijzertransport. De stam LMG 11119 die een hogere kolonisatie vertoonde, had een overexpressie van de genen tonB, exbB en exbD, genen geassocieerd met ijzerassimilatie en proliferatie in andere Gram-negatieve bacteriën zoals *E. coli* en *Salmonella enterica*.

Het doel van de studie gepresenteerd in **hoofdstuk 4** was om het *Arcobacteraceae* pangenoom te identificeren en de verschillen tussen de 20 beschreven species te karakteriseren. Hiervoor werd analyse uitgevoerd op de volledige genomen van de overeenkomstige typestammen die zijn verkregen door Illumina-sequencing, waarbij verschillende bioinformatica hulpmiddelen zijn toegepast om de verschillende outputs te analyseren en evalueren. De resultaten onthulden de aanwezigheid van pangenoompartities met aantallen orthogroepen en genen die vergelijkbaar zijn met andere Gram-negatieve bacteriesoorten zoals *Campylobacter*, wat een taxonomische classificatie in één enkel geslacht suggereert, zoals ook oorspronkelijk werd voorgesteld. Er werden verschillen gevonden tussen de genomen van de klinisch geassocieerde species die ook voorkomen in dieren en de species die tot nu toe enkel voorkomen in milieu matrices zoals water. Bovendien werd een kleinere genoomgrootte waargenomen bij de dier en mens gerelateerde species. Verschillende genklasses samenstellingen in dier en mens geassocieerde species zijn aanwezig met een hoger percentage virulentie gerelateerde genklasses zoals celmotiliteitsgenen. Gegevens vanuit de huidige studie ondersteunen de voorgestelde opdeling in verschillende genera niet, maar toont wel een aanpassing aan de omgevings- en/of gastheeromstandigheden van sommige species. Deze resultaten suggereren eerder een indeling in pathogene en niet-pathogene species, wat in de toekomst eventuele

oriëntatie naar klinische relevantie en impact op de voedselveiligheid en de volksgezondheid mogelijk maakt.

Samenvattend, *A. butzleri* toonde het vermogen om de humane slijmbarrière te overwinnen. Het open pangenoom van *A. butzleri* en de uitwisselbaarheid van potentiële virulentiegenen werd waargenomen in de 32 geanalyseerde stammen. Deze aspecten kunnen worden beschouwd als belangrijke genomische kenmerken in een gastheeraanpassing. De functionele annotatie van de *A. butzleri* genomen toonde de aanwezigheid van vermeende virulentiegenen en antigeenherkenningsmerkers. De *A. butzleri* stammen vanuit verschillende gastheren vertoonden een vergelijkbaar kolonisatievermogen in vitro, zonder een duidelijke cel invasiviteit. Een deel van de stammen vertoonde echter een ander gastheerkolonisatie vermogen. Deze stammen werden gekozen voor daaropvolgende transcriptoom onderzoeken in contact met gastheercellen. De betrokkenheid van sommige genen die momenteel door genomische analyses als virulentie-geassocieerd worden beschouwd, werd bevestigd door RNA-seq analyse. Een snelle reactie op omgevingsstimuli werd waargenomen in een stam die een grotere kolonisatie vertoonde. Dit aspect suggereert een belangrijke rol van omgeving gerelateerde genexpressierespons die ook een rol zou kunnen spelen bij aanpassing onder gastheerkolonisatie omstandigheden. Het belang van ijzermetabolisme tijdens *A. butzleri* infectie werd gesuggereerd door overexpressie van genen die zijn gekoppeld aan ijzertransport en aan organisch zuur gerelateerde genen. Dit aspect werd ook bevestigd door chemische analyse. De transcriptoom gegevens maakten het mogelijk om de rol te begrijpen van genen die momenteel als virulentie gerelateerd worden beschouwd uit functionele annotatiestudies.

De studies gepresenteerd in dit proefschrift zijn uitgevoerd op vereenvoudigde in vitro modellen. Hoewel darmkenmerken complexer zijn in vergelijking met de cellulaire modellen, kunnen deze gegevens worden gebruikt voor daaropvolgende gerichte in vitro en in vivo studies. De *Arcobacteraceae* species worden gekenmerkt door verschillende bronnen van isolatie en associatie met klinische relevantie. Dit aspect kan verband houden met een verschillende genoomgrootte en gen inhoud tussen de species. De *Arcobacteraceae* genoompartities verschillen niet van andere species die behoren tot andere bacteriële genera zoals *Campylobacter*, wat suggereert dat er slechts één genus "*Arcobacter*" bestaat. De kleinere genoomgrootte van species die als gerelateerd aan dieren en klinische relevant in mensen worden beschouwd in vergelijking met species die verband houden met het milieu, werd waargenomen. Dit aspect leidde tot het detecteren van groep specifieke genklasses die een evolutionaire specialisatie van sommige species tot dierlijke gastheren suggereren. De kleinere genomengrootte van klinische- en dier-gerelateerde species leidt tot de hypothese dat er een verband bestaat tussen de pathogeniteit van de genoomgrootte van *Arcobacter*. Het behoud en de aanwezigheid van specifieke sequenties demonstreert het belang van routes voor specifieke bacterieel metabolisme, zoals de celluloseremming in *A. nitrofigilis*. Dit aspect suggereert het bestaan van sequenties die verband houden met verschillende bacteriële levensstijlen in verschillende omgevingsniches. De studie van het *Arcobacteraceae* pangenoom maakte de detectie mogelijk van enkele specifieke sequenties die verband hielden met ecologische niches. De toename in de komende jaren van genomische sequenties en informatie over *Arcobacteraceae* zal nuttig zijn om hun ecologische en klinische rol beter te begrijpen.

Italian version (**Riassunto**)

La famiglia delle *Arcobacteraceae* include batteri Gram-negativi isolati da diversi ambienti e animali, tra cui acque reflue, ambienti di produzione petrolifera, sedimenti marini, estuari e fiumi, lumache marine, ostriche, anellidi abissali e allevamenti ittici. Le specie appartenenti alle *Arcobacteraceae* sono state isolate da animali terrestri, in particolare da polli, suini e bovini. Alcune specie sono considerate patogene alimentari essendo state isolate da diversi tipi di cibo, e prevalentemente da alimenti di origine animale e da casi clinici. Le specie *Arcobacter butzleri* e *Arcobacter cryaerophilus* sono state isolate da campioni fecali di pazienti con disordini gastrointestinali. L'abilità di specie incluse nelle *Arcobacteraceae* di sopravvivere in differenti ambienti e ospiti suggerisce una pressione evolutiva che ha portato a una conseguente variazione del loro contenuto del genoma. Inoltre, le differenti caratteristiche fisiologiche e genomiche hanno portato a proporre una suddivisione della famiglia *Arcobacteraceae* in diversi generi, aspetto criticato per l'assenza di evidenze biologiche e genomiche.

La patogenicità di alcune specie appartenenti alle *Arcobacteraceae* è già stata studiata usando test *in vitro*. I modelli cellulari *in vitro* rappresentano un valido strumento per i primi studi di interazione ospite-patogeno su un ampio numero di ceppi. Questi modelli possono consistere in una singola linea cellulare o più linee cellulari con differenti caratteristiche. I modelli cellulari *in vitro* sono spesso utilizzati nella valutazione dell'adesione, invasione e citotossicità batterica.

I modelli cellulari utilizzati per lo studio di patogeni intestinali vengono spesso prodotti utilizzando cellule di origine tumorale essendo più facilmente coltivabili. Le linee cellulari Caco-2 e HT29-MTX-E12 sono

state utilizzate nella produzione di modelli misti caratterizzati da alcune significanti caratteristiche intestinali. I modelli prodotti con le cellule Caco-2 sono stati utilizzati per la valutazione della colonizzazione di batteri patogeni come *Campylobacter jejuni*, *Campylobacter coli*, *Salmonella enterica* and *Escherichia coli*. La linea cellulare HT29-MTX-E12 (cellule intestinali sub-omogenee produttrici di muco) deriva dalla linea parentale HT29. Questa linea cellulare produce muco esibendo un fenotipo simile alle cellule globose con un'espressione predominante delle mucine MUC5AC. Il muco rappresenta un elemento prodotto dai tessuti epiteliali che influenza batteri patogeni e non patogeni.

Le caratteristiche batteriche fisiologiche e fenotipiche possono essere collegate a caratteristiche genomiche (es. presenza o assenza di specifici geni). Nel caso di *Arcobacter* spp. le tecniche del sequenziamento del DNA sono state utilizzate per caratterizzare nuove specie, per ottenere informazioni riguardanti specifiche sequenze (es. 16s rRNA) e per progettare nuove tecniche molecolari. Le tecniche di sequenziamento sono state utilizzate nello studio dell'intero genoma batterico e del contenuto genico. Inoltre, il sequenziamento dell'RNA (RNA-seq) permette la valutazione dell'intero trascrittoma. Le informazioni ottenute dal numero di trascritti per gene vengono normalizzate ed elaborate con pacchetti statistici per valutare i geni differenzialmente espressi (DEGs) sovra-espressi e sotto-espressi. Questo metodo può essere usato nella valutazione dell'espressione genica di un organismo in contatto con un altro, permettendo lo studio del trascrittoma di uno o più organismi.

Il presente studio di tesi di Dottorato prende in considerazione lo studio di specie relative alle *Arcobacteraceae* e in particolare di *A. butzleri* a diversi livelli. L'obiettivo è stato quello di ottenere dati genomici e trascrittomici

relativi a diversi ceppi di *A. butzleri*. La scelta di focalizzare le analisi su *A. butzleri* è dovuta alla sua maggiore frequenza di isolamento, rispetto alle altre specie appartenenti alle *Arcobacteraceae*, da differenti animali, alimenti e casi clinici come riportato in letteratura.

Nel capitolo 2 è riportata l'analisi genomica comparativa di 32 ceppi di *A. butzleri* con la valutazione della loro colonizzazione e invasione su modelli produttori e non produttori di muco *in vitro*. Questo approccio è stato seguito per esplorare possibili collegamenti tra la colonizzazione e invasione di *A. butzleri* e il suo contenuto genomico. Inoltre, è stata investigata la presenza di geni imputati alla virulenza e la loro possibile correlazione con la fonte di isolamento dei ceppi. L'analisi dei genomi è stata seguita dalle analisi del trascrittoma per valutare la funzione dei geni di *A. butzleri*. Più specificamente, nel capitolo 3 viene presentato il profilo dell'espressione globale di tre ceppi relativi a *A. butzleri* studiato tramite RNA-seq. I ceppi sono stati selezionati considerando i dati relativi alla loro abilità di colonizzazione e invasione. L'espressione genica è stata valutata nei batteri in contatto con un modello intestinale produttore di muco dopo 30 e 90 minuti.

Nel capitolo 4 viene presentata l'analisi dei genomi di 20 specie relative alle *Arcobacteraceae*, recentemente riclassificate a livello di genere. L'obiettivo specifico di questo studio era di ottenere informazioni genomiche relative all'intera famiglia batterica.

Le conclusioni generali della tesi sono presentate nel capitolo 5 per illustrare gli avanzamenti riguardanti la conoscenza di *Arcobacter* spp. dopo l'ottenimento dei dati presentati.

Lo studio mostrato nel capitolo 2 è stato eseguito per valutare il contenuto genetico di 32 ceppi relativi a *A. butzleri* con diversa origine di isolamento.

Questo studio è stato eseguito per esplorare possibili correlazioni tra i genomi di *A. butzleri* e l'abilità dei ceppi di colonizzare e invadere cellule intestinali umane *in vitro* tramite l'uso della genomica comparativa e test fisiologici eseguiti su modelli intestinali. L'infezione simulata di modelli intestinali umani mostrava una colonizzazione batterica più alta in presenza di cellule produttrici di muco (HT29-MTX-E12). Per una parte dei ceppi, il muco umano aumentava significativamente la resistenza alla rimozione fisica dalla mucosa *in vitro,* inoltre è stata osservata una crescita batterica per alcuni ceppi dopo un breve periodo. Le analisi del pangenoma ha mostrato un variabile genoma accessorio. Questo genoma accessorio ipervariabile non era strettamente correlato alla fonte di isolamento di *A. butzleri*. Similarmente, la filogenesi dei ceppi non era correlata alla loro origine, mentre un certo raggruppamento è stato osservato per i ceppi isolati da diversi tratti dell'intestino del maiale. Molti dei geni imputati alla virulenza sono stati rilevati nel genoma accessorio. La sintesi dei lipopolisaccaridi (LPS), e in particolare la glicossilazione della catena dell'O-antigene, è presente in una regione di alta plasticità del pangenoma. Questo aspetto indica un frequente fenomeno di trasferimento genico orizzontale e il coinvolgimento della sua struttura ipervariabile nel comportamento adattativo di *A. butzleri*. I risultati approfondiscono la conoscenza relativa al pangenoma di *A. butzleri* estendendo l'insieme di geni considerati indici di virulenza e provvedendo la base per lo sviluppo di nuovi approcci diagnostici per la rilevazione di ceppi con alto potenziale di virulenza.

Il capitolo 4 mostra la valutazione del trascrittoma di *A. butzleri* tramite RNA-seq. Recenti avanzamenti nell'analisi dei genomi di *A. butzleri* hanno sottolineato la presenza di possibili geni di virulenza e relativi

meccanismi di azione. Con l'ottenimento dei dati genomici risulta quindi possibile e importante la valutazione del trascrittoma per studiare possibili meccanismi di virulenza in condizioni che mimano il processo di infezione. Lo studio RNA-seq è stato focalizzato sulla valutazione del trascrittoma di tre ceppi di *A. butzleri* isolati da feci umane e che hanno mostrato un differente potenziale di virulenza *in vitro* (capitolo 2). Lo studio del trascrittoma è stato eseguito con l'uso di un modello intestinale umano produttore di muco (Caco-2/HT29-MTX-E12). La colonizzazione e invasione simulate hanno permesso l'ottenimento di dati fisiologici e di espressione genica in parallelo. L'abilità dei tre ceppi di *A. butzleri* di colonizzare le cellule intestinali *in vitro* è stata confermata. I risultati ottenuti dal trascrittoma hanno permesso di identificare la sovra-espressione di geni correntemente considerati associati alla virulenza. Tra questi geni *irgA*, *oprF* e *iroE*, frequentemente impiegati nello studio di *A. butzleri*. Una generale sovra-espressione è stata osservata in *A. butzleri* incubato in DMEM specialmente nel ceppo LMG 11119 suggerendo la sua maggiore adattabilità a condizioni ambientali. Inoltre, geni non considerati normalmente associati alla virulenza di *A. butzleri* sono risultati differenzialmente espressi durante la colonizzazione dei modelli cellulari. Le funzioni di questi geni sono risultate diverse. Alcuni di essi erano coinvolti nel metabolismo degli acidi organici (es. *actP* e *yjcH*). Questi geni sono considerati legati alla virulenza di *E. coli*. Altri rilevanti DEGs erano collegati al trasporto del ferro. Il ceppo LMG 11119 che mostrava una maggiore colonizzazione mostrava anche la sovra-espressione dei geni *tonB, exbB* e *exbD* legata all'assimilazione del ferro e alla proliferazione di altri batteri Gram negativi (*E. coli* e *S. enterica*).

Gli obiettivi degli studi presentati nel capitolo 4 erano di valutare il pangenoma della *Arcobacteraceae* e di caratterizzare potenziali differenze presenti in 20 specie descritte e validate. Le analisi sono state condotte su genomi interi di ceppi tipo ottenuti con sequenziamento Illumina applicando diversi strumenti bioinformatici per comparare diversi *output*. I risultati hanno rilevato la presenza di partizioni del pangenoma e rispettivi ortogruppi e geni comparabili a altri batteri Gram negativi come *Campylobacter* spp., suggerendo la classificazione in un singolo genere come originariamente proposto. Erano presenti differenze tra genomi associati ad animali e casi clinici in uomo e quelli isolati da altre fonti come l'acqua. Inoltre, è stato osservato un genoma più piccolo in specie isolate da animali e casi clinici in uomo rispetto alle altre specie. È stata osservata una composizione diversa delle classi geniche in specie isolate da animali e casi clinici in uomo con una più alta percentuale di geni legati alla virulenza come geni collegati alla motilità. I dati di questo studio non supportano la proposta di suddividere le *Arcobacteraceae* in diversi generi identificando al contempo un adattamento a condizioni legate all'ambiente e all'ospite di alcune specie.
I risultati suggeriscono una divisione in specie patogene e non patogene per focalizzare futuri studi su quelle specie di rilevanza clinica e con impatto sulla sicurezza alimentare e sanità pubblica.

In conclusione, *A. butzleri* ha mostrato l'abilità di superare la barriera di muco umano. Il pangenoma batterico di *A. butzleri* e la sua intercambiabilità di potenziali geni di virulenza sono stati osservati nei 32 ceppi analizzati. Questi aspetti sono stati proposti come caratteristiche genomiche legate all'adattamento all'ospite di *A. butzleri*. L'annotazione funzionale dei genomi di *A. butzleri* mostrava la presenza di geni imputati

alla virulenza e di antigeni marcatori di riconoscimento da parte dell'ospite. I ceppi di *A. butzleri* isolati da diversi ospiti mostravano una colonizzazione simile *in vitro*, senza una marcata invasività. Però, parte di questi ceppi mostravano un'abilità di colonizzazione diversa comparata a quella di altri ceppi. Questi ceppi sono stati scelti per i successivi studi di trascrittomica in contatto con le cellule ospite.

Il coinvolgimento di alcuni geni correntemente considerati di virulenza da analisi genomiche è stato confermato dai dati RNA-seq. Una veloce risposta a stimoli ambientali è stata osservata nel ceppo LMG 11119 che mostrava una maggiore colonizzazione. Questo aspetto suggerisce un importante ruolo di geni legati alla risposta all'ambiente che possono giocare un ruolo nell'adattamento a condizioni legate alla colonizzazione dell'ospite. L'importanza del metabolismo del ferro durante le infezioni di *A. butzleri* è stata suggerita dall'osservazione di geni sovra-espressi legati al trasporto di ferro nel ceppo LMG 11119 caratterizzato da una maggiore colonizzazione. L'espressione genica differenziale di geni legati ad acidi organici è stata confermata con analisi chimiche (*High Pressure Liquid Chromatography*). I dati trascrittomici hanno permesso di comprendere il ruolo di geni correntemente considerati legati alla virulenza da studi di annotazione funzionale.

Gli studi presentati in questa tesi sono stati eseguiti su modelli semplificati *in vitro*. Benché le caratteristiche dell'intestino siano più complesse rispetto a modelli cellulari, i dati ottenuti possono essere usati per futuri studi specifici *in vitro* e *in vivo*.

Le specie relative alle *Arcobacteraceae* sono caratterizzate da diverse fonti di isolamento e associazione di rilevanza clinica. Questo aspetto può essere collegato a differenti taglie genomiche e contenuto genico delle specie. Le

partizioni dei genomi (*core, cloud,* e *soft genome*) non differiscono da altri generi come *Campylobacter* spp., suggerendo l'esistenza di un solo genere batterico, "*Arcobacter",* nella famiglia *Arcobacteraceae*. Sono stati osservati genomi più piccoli in specie legate ad animali e casi clinici in comparazione con genomi di specie isolate dall'ambiente. Questo aspetto ha portato alla rilevazione di classi geniche gruppo-specifiche suggerendo una speciazione legata a ospiti animali. Questi genomi di taglia ridotta portano a ipotizzare un legame tra taglia genomica e la patogenicità di *Arcobacter* spp.. Il mantenimento e la presenza di specifiche sequenze dimostra l'importanza di *pathways* per specifici metabolismi batterici (es. inibizione della cellulosa in *Arcobacter nitrofigilis*). Questo aspetto suggerisce l'esistenza di sequenze batteriche legate a differenti nicchie ambientali e al relativo adattamento batterico a differenti condizioni ambientali. L'aumento nei prossimi anni di sequenze genomiche relative alle *Arcobacteraceae* e le relative informazioni sarà utile per comprendere meglio il ruolo nell'ambiente e in ambito clinico delle specie relative a questa famiglia batterica.

Acknowledgements

I would like to thank my Supervisors, Prof. Kalliopi Rantsiou and Prof. Kurt Houf, for their supervision during these years of PhD and for the supervision during the writing of the thesis. Moreover, I would like to offer my special thanks to Prof. Kurt Houf for the assistance in organizing my period abroad.

I wish to show my appreciation to Prof. Valentina Alessandria for her constant support during these years.

Infine, vorrei ringraziare la mia Famiglia, Papà, Erica, Patrizia e Andrea per il supporto in questi anni di Dottorato specialmente nei momenti più complicati.