Faculty of Veterinary Medicine
University of Helsinki

Prospects of whole-genome analysis in understanding the ecology and evolution of foodborne pathogens

Kaisa Jaakkola

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Veterinary Medicine of the University of Helsinki, in Sali 2, Metsätalo, on the 26th of August, 2022 at 12 o'clock.

Helsinki 2022

ISBN (pbk.) 978-951-51-8381-1

ISBN (PDF) 978-951-51-8382-8

http://ethesis.helsinki.fi

Unigrafia

Helsinki 2022

Supervising professor	Professor Miia Lindström, DVM, Ph.D. Department of Food Hygiene and Environmental Health Faculty of Veterinary Medicine University of Helsinki, Helsinki, Finland
Supervisors	Professor emeritus Hannu Korkeala, DVM, Ph.D., M.Soc.Sc. Department of Food Hygiene and Environmental Health Faculty of Veterinary Medicine University of Helsinki, Helsinki, Finland
	Professor Miia Lindström, DVM, Ph.D. Department of Food Hygiene and Environmental Health Faculty of Veterinary Medicine University of Helsinki, Helsinki, Finland
	Docent Riikka Keto-Timonen, DVM, Ph.D. Department of Food Hygiene and Environmental Health Faculty of Veterinary Medicine University of Helsinki, Helsinki, Finland
Reviewed by	Professor Thomas Alter, DVM, Ph.D. Institute of Food Safety and Food Hygiene Freie Universität Berlin, Berlin, Germany
	Professor Claudia Guldimann, DVM, Ph.D. Institute for Food Safety and Hygiene Ludwig-Maximilians Universität, München, Germany
Opponent	Professor Pentti Huovinen, MD, Ph.D. Institute of Biomedicine University of Turku, Turku, Finland

LIST OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following articles, which are referred to in the text by their roman numerals:

I Söderholm, H*., **Jaakkola, K.,*** Somervuo, P., Laine, P., Auvinen, P., Paulin, L., Lindström M. & Korkeala, H. Comparison of *Clostridium botulinum* genomes shows the absence of cold shock protein coding genes in type E neurotoxin producing strains. Botulinum Journal, 2013, 2. https://doi.org/10.1504/TBJ.2013.055662

II **Jaakkola, K.,** Virtanen K., Lahti P., Keto-Timonen R., Lindström M. & Korkeala, H. Comparative genome analysis and spore heat resistance assay reveal a new component to population structure and genome epidemiology within *Clostridium perfringens* enterotoxin-carrying isolates, Frontiers in Microbiology, 2021, 12. https://doi.org/10.3389/fmicb.2021.717176

III **Jaakkola, K.,** Somervuo, P., & Korkeala, H. Hybridization analysis of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* identifies genetic traits to elucidate their different ecologies. BioMed Research International, 2015. https://doi.org/10.1155/2015/760494

IV Virtanen J-P, Keto-Timonen R., **Jaakkola K.,** Salin N., Korkeala H. Changes in transcriptome of *Yersinia pseudotuberculo*sis IP32953 grown at 3 and 28°C detected by RNA sequencing shed light on cold adaptation, Frontiers in Cellular and Infection Microbiology, 2018, 8. https://www.frontiersin.org/article/10.3389/fcimb.2018.00416

*equal contribution

The original articles have been reprinted as appendices to this thesis with the kind permission of their copyright holders: Inderscience Publishers (I), Frontiers Media SA (II, IV), and Hindawi (III).

ABSTRACT

Globally, an estimated 1.9 billion people acquire a foodborne infection, and 715,000 die each year.

The burden of foodborne disease is significant, yet the reservoirs, attributable sources of infection, lifestyles, and ecological niches of many sporadic, common and preventable foodborne pathogens remain poorly understood.

The abundance and availability of genomic data make it an appealing new starting point for the study of ecological niches and evolution within bacteria. This thesis aimed to assess the suitability and prospects of comparative genome analysis and pangenomic approach for a better understanding of the ecology, evolution, and through these, also the epidemiology of foodborne pathogens.

Comparative genome analyses conducted on *C. botulinum, C. perfringens*, *Y. enterocolitica*, and *Y. pseudotuberculosis* shed light on genetic differences that contributed to cold adaptation and spore heat resistance. Psychrotrophic *C. botulinum* strains Type E in Group II lacked cold shock protein genes conserved in other strains suggesting that cold shock protein homologs are not necessary for cold adaptation in *C. botulinum* Type E. The comparative genome analysis of *cpe*-carrying *C. perfringens* strains revealed a novel subgroup, and a corresponding gene profile, of food poisoning-associated lineage IV strains that produced heat-sensitive spores. A transcriptome study on cold shock response and cold adaptation of *Yersinia pseudotuberculosis* serotype O1 identified genes contributing to cold survival. These genes included cold shock proteins and RNA helicases CsdA, RhlE, and DbpA forming the backbone of cold response, in addition to the transcription factors IF-1, RbfA, and Rho supporting the protein synthesis at suboptimal temperatures.

Additionally, genetic changes contributing to niche adaptation and survival in different reservoirs were identified. Comparative genome analysis of *C. perfringens* identified a putative reservoir or origin for lineage IV strains within swine and poultry farms. Exploration of genetic diversity within lineage IV strains also suggested that strains with different gene profiles had adapted to different ecological niches and reservoirs. The comparison of enteropathogenic *Yersinia* isolates revealed that all *Y. pseudotuberculosis* isolates shared several genetic traits, useful for survival in various environments that were absent from *Y. enterocolitica*. Most notably, the *Y. pseudotuberculosis* strains harbored a selection of type VI secretion systems targeting the competitive cells of other microbes and eukaryotes. The genomes of *Y. enterocolitica* were more streamlined and the biotypes had undergone reductive evolution.

The pangenomic approach was applied to provide a scalable, high-resolution view of population structure. The pangenome of *C. botulinum* reflected the heterogeneous nature of this pathogen. The isolates studied shared only a small core genome comparable to that of the clostridial backbone and the pangenome was large. Within *C. perfringens* the pangenome analysis revealed a substantial core genome with a large pangenome. The comparison of enteropathogenic *Yersinia* isolates revealed a large core genome shared between enteropathogenic *Yersinia*. A pangenomic approach successfully elucidated adaptive evolution within virulence and stress response mechanisms and allowed inference of evolutionary relationships between foodborne pathogens.

The studies on this thesis shed light on the genes contributing to stress tolerance (I, II, IV), niche adaptation (II, III), and ecology (I-IV) in foodborne pathogens and the pangenomic approach also shed light on their distribution within the population. Knowledge of genes associated with stress tolerance, reservoirs, and lifestyles are beneficial in the development of new targeted strategies and measures to identify and control the food safety risks caused by these bacteria.

ABBREVIATIONS AND GLOSSARY

AAD	Antibiotic-associated diarrhea
ABC	ATP- binding cassette
ADI	Arginine deiminase
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
Asp	Aspartate
BBH	Bi-directional best hit
BHI	Brain-heart infusion, broth
BLAST	Basic Local Alignment Search Tool, an algorithm widely used for comparing nucleotide sequences
BLASTP	Basic Local Alignment Search Tool, an algorithm widely used for comparing protein sequences
BoNT	Botulinum neurotoxin
c-cpe	Chromosomally cpe-carrying *C. perfringens*
cgMLST	Core genome multilocus sequence typing
CDS	Coding sequence
CFU	Colony forming unit
CGH	Comparative genome hybridization
CPE	*Clostridium perfringens* enterotoxin
CRISPR	Clustered regularly interspaced palindromic repeats
DEAD box	Collection of conserved nine motifs, one of which contains the amino acid sequence D-E-A-D (asp-glu-ala-asp)
DS	Duncan-Strong broth
DPA	Dipicolinic acid
FTG	Fluid thioglycolate medium

GFF3	General feature format
GC%	Guanine and cytosine content (percentage)
GI	Gastro-intestinal
GWAS	Genome wide association analysis
HGT	Horizontal gene transfer
IS	Insertion sequence
LD	linkage disequilibrium
Log OR	Logarithm of the odds ratio
Log FC	Logarithm of fold change
MIAME	Minimum information about a microarray experiment
MINSEQE	Minimum information about a next-generation sequencing experiment
MLST	Multilocus sequence typing
NGS	Next-generation sequencing techniques
OD	Optical density
PCR	Polymerase chain reaction
p-cpe	Plasmid borne cpe-carrying *C. perfringens*
PFGE	Pulsed-field gel electrophoresis
PTS	Phospho-transferase system
RNASeq	RNA-sequencing
RTqPCR	Real-time quantitative polymerase chain reaction (PCR)
SNP	Single nucleotide polymorphism
SMRT	Single-molecule real -time technology
T3SS	Type III secretion systems
T6SS	Type VI secretion systems
WGS	Whole genome sequencing
WGH	Whole genome hybridization

1 INTRODUCTION

Globally, an estimated 1.9 billion people acquire a foodborne infection each year, and 715,000 die from it. In as many as two-thirds of foodborne outbreaks, no causative agent is identified and new, emerging foodborne pathogens are identified every few years. Also the reservoirs, attributable sources of infection, lifestyles, and ecological niches of many sporadic, common and preventable foodborne pathogens remain poorly understood (1,2). These gaps in epidemiological and ecological understanding hinder the risk management and prioritization of public health measures for food safety control.

The first bacterial genome of *Haemophilus influenzae* (3) was sequenced in 1995, and since then, we generate more data every few days than was produced from the dawn of history up to the year 2000. The availability and affordability of whole-genome sequencing (WGS) methods have revolutionized foodborne disease surveillance and replaced traditional microbial typing tools. WGS generates a wealth of data as a by-product of typing, and the comparative genome analysis of sequenced genomes has been further utilized to study the biochemical and functional capacities of pathogens, and reveal new virulence factors and vaccine or drug molecule candidates. Exertion of the full potential of the generated sequence data is one of the great challenges in current microbiology. In the end, data is valuable only to the extent that it can, in one way or another, be transformed into knowledge and wisdom (4).

In this thesis, the prospects and suitability of whole-genome analysis for furthering our understanding of the ecology, lifestyle, and adaptation of foodborne pathogens are investigated using the sporadic foodborne pathogens *Clostridium botulinum, Clostridium perfringens, Yersinia enterocolitica,* and *Yersinia pseudotuberculosis* as study examples. The objective is to evaluate how much we can determine about these foodborne pathogens, their reservoirs, population structure, and lifestyle based on their genomes.

An important concept for this thesis is the pangenome. The term pangenome was first coined in 2005 as a comparative analysis of a few bacterial genomes indicated that an infinite number of genomes would have to be sequenced to completely present all genes within an organism (5). This infinite genetic pool was termed open pangenome and challenged the notion that the genome of a single isolate of a given species was sufficient to represent the genomic content of that species. Since then pangenomics and comparative genome analysis have become important fields of research.

Comparative genome analysis can help infer the lifestyles, gene repertoires, and minimal genome sizes of pathogens. In studying pathogens, it is sometimes beneficial to look beyond the clinical isolates and diseases caused. The driving force for the evolution of foodborne bacterial pathogens is survival in their reservoirs, and many bacterial pathogens are considered to infect humans only incidentally. Many of the known virulence factors in foodborne bacteria are also active against non-mammalian adversaries such as insects, protozoa, other bacteria, and bacteriophages. Additionally, the study of the whole population rather than just pathogenic isolates can provide greater insight and resolution to understanding the evolution of pathogens (6).

Further understanding of the reservoirs, population structures, source attribution, and pathogenomics of foodborne pathogens enables more effective and precise control and prevention measures to be applied in all steps of the food production chain. In-depth understanding and knowledge are required to effectively control food safety as food trends, climate change, and intensified food production alter the food chain and new risks emerge.

2 REVIEW OF THE LITERATURE

2.1 Bacterial genome sequencing and characteristics

2.1.1 Sequencing

The first bacterial genome of *Haemophilus influenzae* was sequenced less than 30 years ago, but the efficiency and improvement of DNA sequencing technologies have completely revolutionized the availability of bacterial genomic data (3). The introduction of next-generation sequencing techniques (NGS) has made the whole-genome sequencing (WGS) of bacterial isolates fast, available and affordable (7,8).

The first sequencing technique developed by Fred Sanger was combined with shotgun cloning in 1995, kick-starting the race to sequence bacterial pathogens (3,9). Shotgun sequencing was surpassed by the introduction of high-throughput sequencing (454, Illumina) increasing the speed and efficiency of sequencing (10,11). These short-read technologies delivered high-accuracy reads, at the price of assembly quality. This changed the emphasis from whole-genome sequencing of new organisms to the sequencing of closely related genomes to enable *de novo* assembly against a reference genome. The third revolution for sequencing was the introduction of sequencing methods using single-molecule templates and producing long read lengths with compromises on read accuracy (PacBio Single-molecule real-time technology or SMRT, Nanopore). The new technologies have made the assembly of genomes easier, leading to an increase in published complete genomes and plasmids (12,13).

Initially, it was commonly accepted that a single clinical isolate could be sufficient to represent the genomic content of an entire organism or pathogen. However, twenty years ago, the first three sequenced *E. coli* genomes already revealed the unexpectedly large role played by horizontal gene transfer (HGT) in strain diversification, with and the resulting genetic variation between genomes increasing the importance of comparative genomics and sequencing multiple isolates (14–16). Nowadays the WGS based methods and typing tools have become the golden standard for epidemiological investigation, and the ideal course of action is to sequence, not just the clinical isolates but also genetically varied isolates across different reservoirs, host species and biomes to efficiently screen for the genetic variety within pathogenic species.

2.1.2 Genome assembly

Genome assembly entails the identification of overlapping regions in sequenced reads and joining these together into longer sequences, also known as contigs (7). Formed contigs are sorted and connected by filling the gaps to form a complete genome. The shorter the reads produced, and the wider the gaps between them, the more challenging the complete assembly is. Uncompleted genomes consisting of several contigs are called draft assemblies. The emergence of third-generation sequencing methods such as SMRT has made the sealing of gaps in genome assemblies more convenient and inexpensive (12). Currently, relatively few complete genomes are available compared to draft assemblies, but the proportion of complete genomes is increasing (8,17,18).

2.1.3 The size and structure of a bacterial genome

The bacterial genome typically consists of one circular chromosome in addition to extrachromosomal elements such as plasmids and bacteriophages (19). Also, linear chromosomes and bacteria with several chromosomes have been described (20,21). Especially compared to eukaryotic genomes the typical bacterial genome is small, gene-rich, and under 10 megabases in length (22). The genome size and gene count correlate in bacterial genomes due to the scarcity of gene-poor regions (23).

Bacterial genomes evolve through clonal divergence, through the modification of existing genes, and by the acquisition of new sequences through HGT (19,24). The phylogeny of each HGT element is distinct from the phylogenetic signal of the rest of the genome, and therefore the bacterial genome can be described as a mosaic, where different genetic regions have their separate phylogenetic histories. Due to the integrated HGT elements obscuring and confounding the phylogenetic signal within the genome, quite often only the so-called core genes that are present in all genomes, are used in the phylogenetic analysis (24).

2.1.4 Horizontal gene transfer

The different modes of HGT in bacteria include transformation, transduction, and conjugation. Transformation is the acquisition of exogenous DNA from the environment, transduction is the introduction of exogenous DNA via viral vectors such as phages, and conjugation is the direct transfer of DNA from cell to cell (24). The HGT modes are under genetic, physiological and environmental modulation, so that in specific conditions a bacterial cell can either favor HGT or restrict it (25). Susceptibility to HGT is currently seen as an evolutionary adaptation or as a change in the

evolutionary mode of an organism or certain lineages, as the presence of anti-phage systems such as CRISPR-Cas and restriction-modification systems can play important roles in limiting HGT (26,27).

Some bacteria such as *Vibrio cholerae* and *Campylobacter jejuni* are naturally competent, meaning they can transform exogenous DNA from the environment (27). This competence in DNA internalization is usually transient, and certain substances and environmental conditions induce the competence (28). Typically, the transformation of DNA leads to homologous recombination within chromosomes and the generation of new alleles or allelic replacement (29,30). However, larger chunks of DNA can also be transferred through transformation. In 1928 Griffith observed that avirulent *Streptococcus pneumoniae* could turn virulent when exposed to killed cells of virulent *S. pneumoniae*. Today we know that this obtained virulence was due to the transformation of encapsulation genes from dead cells (31,32).

Integration of temperate bacteriophages, and transfer of genes via transduction, are important for virulence and the emergence of pathogenic strains in many bacteria. Bacteriophages carry and transfer many toxin and enzyme genes.

Conjugation is often mediated by extrachromosomal plasmids, self-repeating units of genetic sequences. Plasmid diversity and size vary between species and the exchange of plasmids can introduce up to 100 new genes in a single step (24). Plasmids are known to transfer important virulence factors such as toxins and also whole operons and their regulatory systems (33–35).

Conjugation and transduction are important modes of HGT for the organisms studied in my thesis. Bacteriophages and plasmids, in particular carry important toxin genes in *C. botulinum* and *C. perfringens* (36–38), and a well-defined virulence plasmid is required for pathogenicity in enteropathogenic *Yersinia* (33).

2.1.5 Functional annotation of bacterial genomes

Genetic sequences are annotated to identify known genetic structures such as genes, to detect homology with related genetic structures and to predict the biological function of identified sequences. This annotation process is often automated *in silico* by the use of algorithms that look for similarities between the known sequences and the predicted genetic structure, and the predicted function is then inherited from the reference sequence. This automatic annotation can then be enriched by manual curation or proteomic analysis data. Automatic functional prediction is based on the assumption that homologous proteins with a similar gene structure share similar biological functions (39).

The bacterial genome is gene-rich and typically around 88% of the genome encodes genes (40). The genes and other genomic features are first predicted by identifying RNAs and coding sequences (CDS) for proteins. Identified non-coding DNA also includes also pseudogenes with frameshift mutations, insertion sequences, and intergenic spacers (40). The identified RNAs and CDSs identified are assigned genetic functions through BLAST search of homologs, as the annotation of the best significant match is transferred to the newly identified gene (41,42). Hierarchy of data trustworthiness is used to produce better quality annotation, as proteins from curated databases such as Uniprot (43) are given preference. Common annotation algorithms include Prokka and RAST (42,44).

Automatic annotation can be a significant source of error for genetic research. Results may differ depending on which annotation pipelines are used, and any errors (gene name, predicted CDS) in existing annotations are quickly propagated to other strains and species (45). The generated annotation for a genome can be likened to a screenshot of the best available knowledge at the time.

2.2 Bacterial population structure and genome evolution

2.2.1 Genome and population structure

Traditional taxonomy and classification of pathogenic bacteria has largely relied on their observed toxinotypes and tested phenotypes under laboratory conditions. The first subtyping methods were based on phenotypic characteristics testable under laboratory conditions, including biotyping, serotyping, and phage typing (46). Subsequently, molecular typing methods such as pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), and multilocus sequence typing (MLST) have been used to identify epidemiological isolates and relatedness of isolates (47–49) . Additionally, the sequence of the 16S rRNA gene has been successfully used to decipher population structure (50). The administration of molecular typing methods has shed light on marked genetic diversity within toxinotype-based species such as *C. botulinum* (51–53).

The most accurate method for strain identification and discrimination is the direct examination of DNA sequences. WGS has already been established as the current gold standard for typing, and sequenced strains may be typed for clinical or scientific purposes based on single nucleotide variants (SNPs) or gene-by-gene allelic profiling of core genome genes (cgMLST) (49,54,55). Typing based on the actual genotype is well suited to evolutionary reconstruction and can be used to study phylogeny and the genetic context within the population. This enables the assignment of unknown strains into pre-existing clusters, revealing related strains, and examination of the prevalence of a

specific gene variant (allele) in a population (54). The genetic differences between clusters, strains, and alleles can also be easily extracted for further investigation. The transition from single-gene (<.0.07% of a genome) to MLST approach (~0.2%) to WGS (100%) has brought us closer to reliably representing the phylogenetic history of bacterial populations (56). The WGS data and genomic studies have also revealed much greater diversity and population structures than those presumed by molecular typing methods (56).

In the light of the molecular typing and WGS data, the concept of bacterial species has been questioned (55,57,58). The standard criteria used to define a species and genetic lineages within eukaryotic organisms are not directly applicable to bacteria as such, but bacteria can also be identified and clustered into genetically and ecologically cohesive entities such as species and lineages (59). Bacterial population structure can be described using a combination of three phylogenetic models: bifurcating trees, networks, and bundles (60). The bifurcating tree is suitable for picturing clonal populations, but networks are required to represent HGT between given members of the population, or even between different species (61). Another characteristic of a bacterial population is the emergence of clones and the associated strong selective sweep. Smith referred to these rapidly spreading clones dominating the pre-existing population, or "star-phylogenies", as bundles (61).

Depending on the organism and the study, terms such as genetic lineage, phylogroup, subgroup, or clade are used to describe distinct clusters within the bacterial population. This separation and speciation within bacterial populations is initiated and maintained by natural selection, driven by ecological niche adaptation and prevention of gene flow between populations through introduced restriction-modification systems, geographic range, or vector specificity (55). The clusters within the population are not constant, and the genetic material transferred between different clusters, in a "plug and play" manner, can lead to diversification of new clusters or despeciation of pre-existing clusters. Due to such type of proceedings within their populations, some bacterial species have famously been described as "fuzzy" (57).

2.2.2 Patterns in genome evolution

Bacterial genomes are molded by evolutive forces similar to those experienced by other organisms. Natural selection and adaptation change the genome size, gene content, and gene density. The classic evolutionary model is that genomes evolve via small changes within individual genes, but this does not seem to be the main mode for genome adaptation and diversification in bacteria (23). The mutation rates are fairly similar between bacterial species (62), but the acquisition of HGT and

homologous recombination rate varies (63). Therefore, in bacteria, the majority of genome size variation and genetic variation is due to the gain and loss of genes (29,40). New genes introducing new beneficial traits such as antimicrobial resistance or the ability to use new substrates are gained through modes of HGT described earlier; meanwhile, the gene loss is the result of a combination of genetic drift with a mutational bias towards deletions (19,22,40). Bias for deletion mutations seems to be universal for bacteria (23,40).

Bacterial genome evolution differs from eukaryotes on exposure to genetic drift. It has been estimated that the genetic drift affected by the clonal population is the most important evolutive force within bacteria (23). Unlike eukaryotes, on exposure to drift, bacterial genomes increase in size as selective pressure drives the accumulation of adaptive genetic modules, and shrink when selective pressure is lifted or limited. In eukaryotes the opposite is true: non-functional DNA accumulates when the selective pressure is low and genetic drift dominates (19).

Bacteria with a restricted host range or specific ecological niche carry smaller genomes, with fewer genes compared to those present in a variety of niches and leading different lifestyles (19). The reduction in genome size is the combined result of reductive evolution due to genetic drift and gene deletions. Commonly, the genes deleted during niche dwelling are the genes that play roles in coping in different niches or lifestyles. In a typical scenario, the niche presents a constant environment rich in substrates, and as a result, some genes are rendered useless as the adoption of dependent lifestyles becomes a possibility (29). Gene deletions are also central in the emergence of bacterial pathogens and particularly for shifts in ecology (64–66). For example, in the emergence of *Y. pestis* from *Y. pseudotuberculosis* the reductive evolution and deletions of genes associated with virulence and metabolism were even more striking than the gene acquisitions (67). The shifts in ecology of *Salmonella enterica* serovars Typhi and Gallinarum (*Salmonella* Typhi, *Salmonella* Gallinarum), and *Mycobacterium leprae* are equally attributed to gene loss within their genomes (64,65,68). Chain et al. (2004) estimated that as much as 10% of the genes of the last common ancestor had been deleted from *Y. pestis.* The genome adaptation via mutational deletions and deletion of superfluous genes is irrevocable, making bacteria unable to independently revert to an ubiquitous lifestyle. Strictly host-restricted genomes also show less susceptibility to HGT (19,69).

Evolution from an unspecialized, free-living organism to a strictly host-restricted organism is a spectrum and the genomes at different stages of this adaptation spectrum (e.g., host-associated, but not yet host-restricted organisms) exhibit different characteristics (Table 1) (70). For example, most bacterial genomes maintain very low numbers (< 10) of insertion sequence (IS) elements (71) whereas some recent symbionts and pathogens (for example Shigella and *Serratia symbiotica*) possess

hundreds of copies (29,72,73). Also, larger numbers of pseudogenes have been reported from host-associated bacteria compared to their unspecialized relatives (74). It is assumed that the initial association with hosts and a recent population bottleneck leads to the accumulation of pseudogenes and mobile elements due to a decrease in effective selection. These changes lower the gene density within genomes and gradually these non-functional sequences are eroded and purged from the genomes (19,40).

Table 1 The characteristics and signature of bacterial lifestyle in their genome architecture (19,70)

Genome characteristic	Free-living	Host-associated	Host-restricted
GC% content	High	Intermediate	Low
Genome size	Large	Intermediate	Small
Genetic diversity	Large	Intermediate	Small
HGT	Open to HGT, diverse plasmidome	High number of IS	Not open to HGT Remnants of mobile elements
Count of pseudogenes	Intermediate	High	Low
Number of biosynthesis and metabolism genes	High	Intermediate	Reduced selection
Other	High number of virulence genes and genes related to stress resistance	High number of secretion systems	
Example species	*Listeria monocytogenes, Clostridium perfringens*	*Yersinia enterocolitica* *Salmonella enterica* serovar Typhimurium	*Salmonella* Typhi *Salmonella* Gallinarum *Yersinia pestis*

Finally, the bacteria with a large population size accommodate more genes and prophages than those with more limited population sizes and niches (75). The accessory genome contains toxins, virulence factors, and other genetic elements involved in bacterial warfare and survival (33,76,77). The large genetic repertoire is likely a competitive advantage in diverse ecological conditions or when competition is present. Genetic diversity and openness to HGT accumulation allow bacteria to adapt

to an ever-changing and competitive environment and enable them to maintain a large population size in a wide range of environments (19,78).

2.2.3 Genomic studies elucidating patterns of evolution and population structure

Although MLST and SNP genotyping techniques are ideal in establishing genetic distance and relatedness, they are less useful in providing information on the unique qualities of individual isolates, such as antibiotic resistance genes, virulence factors, mobile elements, or adapted niche. Whole-genome sequencing and analysis are needed to identify the acquisition and loss of genetic material that has occurred during the adaptation of bacteria in a new environment or host. Comparative genome studies have successfully deciphered the emergence of successful clones and adapted lifestyles of pathogens on the gene and operon levels (79,80).

The emergence of new genetic traits is often intuitively associated with the introduction of new genetic elements. For example, in bacteria, the acquisition of a plasmid enables the introduction of an entire functional operon at once. The acquisition of virulence plasmids encoding toxins or resistance genes is an important and well-known path for the emergence of new pathogenic strains (35,81–84). As an example, the acquisition of plasmid pCD1 marked the emergence of pathogenic *Yersinia* as the introduction of the Type III secretion system in this plasmid helps to impair the host immune system (33).

As established previously the niche adaptation in bacteria occurs through the means of gene decay and reductive evolution. For example, comparative genome analysis has revealed gene decay within *Salmonella* pathovars and helped to understand why some pathovars of *Salmonella* (e.g. host-restricted human-specific *Salmonella* Typhi, and bird-specific *Salmonella enterica* serovar Gallinarum and *Salmonella* Pullorum*)* are more likely to cause extraintestinal infections than others. Pathovars that cause extraintestinal (typhoidal) disease possess several inactive or degraded genes that play roles in chemotaxis, adhesion, and anaerobic metabolism. Deleted genes include the cobalamin synthesis operon and 1,2-propanediol utilization operon, important for ethanolamine utilization. The ethanolamine operon enables the utilization of gut inflammation-derived nutrients to outcompete other gut microbes (68,85,86). The hypothesis is that the loss of these pathways has contributed to host-restricted *Salmonella* pathovars transitioning from intestinal pathogens to invasive pathogens.

Similarly, the emergence of invasive *Y. pestis* from *Y. pseudotuberculosis* is associated with reductive evolution and deletions of genes associated with virulence and metabolism. The deletions include

three loss of function mutations that lower insect toxicity and increase biofilm formation in the flea foregut, which is the known reservoir for *Y. pestis* (67,87–89).

Generally, the genetic adaptations of pathogens enhance survival in their adapted ecological niche and do not primarily increase human pathogenicity or virulence within the mammal hosts. For example, extraintestinal virulence has been suggested to be a coincidental multigenic by-product of GI-tract commensalism (90), as determinants linked to colonization and commensalism (transcriptional regulation, iron metabolism, adhesion, lipopolysaccharide biosynthesis) within the mammalian host are also the same genes associated with extraintestinal virulence (90). Fittingly, the *E. coli* strains that have superior capacity to persist in the intestinal microbe population are also the ones most often involved in extraintestinal infections (91).

Often the evolution of virulence occurs stepwise along with the ecological adaptation. A classic example of this is the emergence of the O157:H7 (EHEC) serotype from the *E. coli* O55:H7-like ancestor and then the subsequent emergence of different clusters of EHEC isolates. The stepwise evolution of EHEC has occurred via HGT as well as gene decay (92,93) and is visualized in Figure 1. Although the importance of Shiga toxin *(Stx)* in disease caused by EHEC is well established, the relationships between nonfermenting sorbitol and the loss of β-glucuronidase activity are not known. The hypothesis is that these changes have occurred during the adaptation and have allowed EHEC O157 strains to occupy a new niche distinct from EPEC O55 (92).

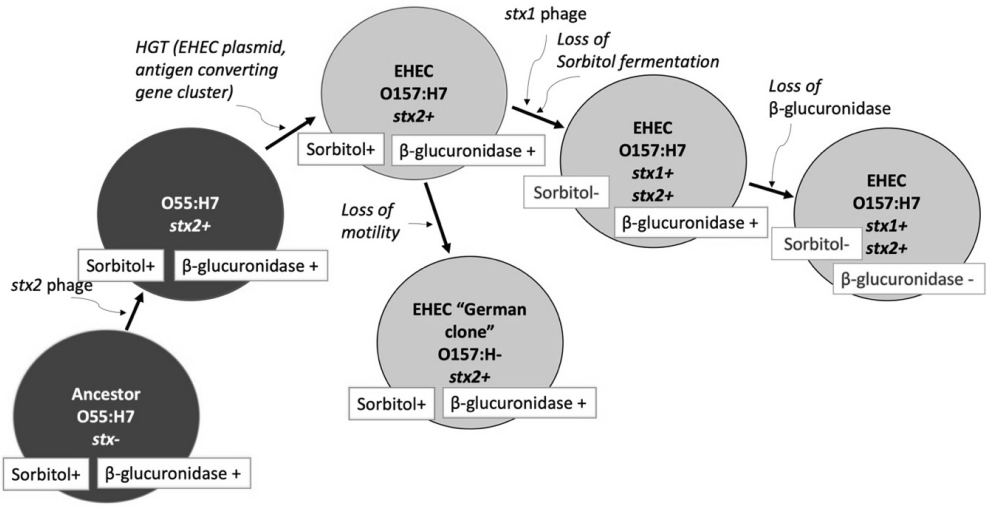

Figure 1. Phenotypic and genotypic changes of evolutionary emergence of EHEC O157:H7 from *E. coli* O55:H7. The figure has been modified from Feng et al. (92).

2.3 Bacterial pangenomics

2.3.1 The pangenome

One important notion of the genomic era is that the majority of genes are not strongly attached to any specific organism, and are rather "shared goods" for several organisms (94). The genetic repertoire available to the bacterial species or microbial community is called the pangenome, while the genome is the genetic repertoire of an individual organism. The pangenome is therefore always larger than the genome of a single organism and large pangenomes arise mainly in organisms with adaptive potential, susceptibility to HGT, and the ability to migrate to new niches (94). Comparative genome analysis and the pangenomic approach can help us infer the lifestyles, genetic repertoires, and minimal core genomes of pathogens.

In *Vibrio cholerae* one of the two chromosomes carries mainly genes related to essential cellular functions (core), and the other carries those concerned with virulence and adaptation (accessory) (20). Most bacteria carry just one chromosome, but also their genes can also be divided into "core genome" and "accessory genome". The pangenome is the sum of the core genome and the accessory genome.

The core genome is the conserved part of the pangenome, which is present in all genomes of all individuals comprising the stable backbone of an organism. Many of the core genes are "housekeeping genes" associated with the maintenance of the basic aspects of the biology of an organism.

The rest of the pangenome belongs to the accessory genome, which contains the mobile genes and strain- and group-specific genes, e.g. both stable and non-stable elements of genetic content. The accessory genome is the "flexible" gene pool, which encodes additional traits that are beneficial under certain circumstances. For example, some genes in accessory genes are ancestral genes that have been deleted from some strains during restrictive evolution, while some are HGT genes unique to a single strain (24,95). Together, the core genome and the accessory genome form the bacterial pangenome (95).

The term pangenome was first coined in 2005 as a comparative analysis of six *Streptococcus agalactiae* isolates indicated that an infinite number of genomes would have to be sequenced to completely present all the genes within *Streptococcus agalactiae* as a species (5). This infinite genetic pool was termed the open pangenome. The concept of pangenome quickly challenged quickly the notion that a genome of a single isolate of a given species was sufficient to represent the genomic content of that species. Since then pangenomics and comparative genome analysis have become important fields of research (96). The availability of sequenced genomes in public databases has opened new ways for scientific collaboration and comparative research to understand similarities and differences among the organisms.

2.3.2 Different types of homologous genes

Homologous genes are an important concept for comparative genomics and pangenome analysis. A homologous gene (or homolog) is a gene inherited in two species from a common ancestor. While homologous genes can be similar in sequence, similar sequences are not necessarily homologous. In pangenome analysis, the aim is to determine the core genome and identify the corresponding genes from each genome. This requires establishing which genes share ancestry and are conserved in studied genomes and which are not, and also discriminating between different types of homologous genes: orthologs from paralogs (Figure 2).

Homologous genes are orthologous if they were separated by a speciation event, but the gene and its main function are conserved (97).

When homologous genes have been separated by a gene duplication event they are called paralogs, and often have different roles or functions to justify their duplicated presence within the same genome (97). A typical example of paralogs is the ABC transporters: the duplicates have altered their substrate-specificity by introducing nucleotide substitutions in their binding periplasmic sites (98).

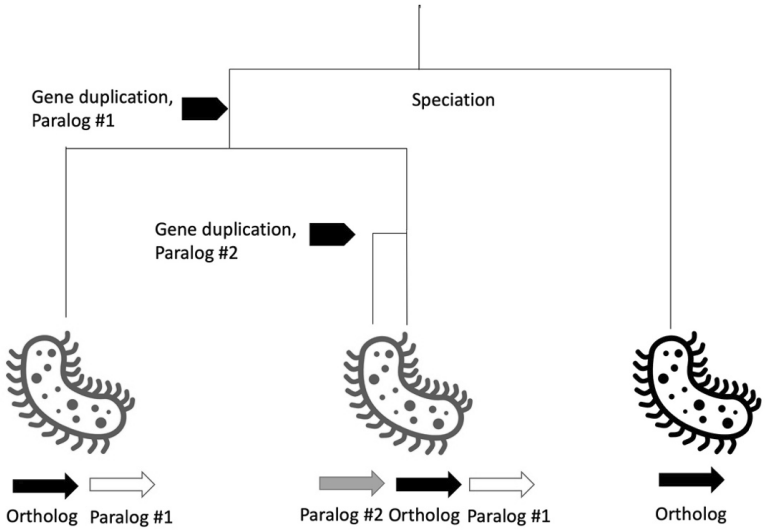

Figure 2. The evolutionary relationship between homologous genes and the difference between orthologs and paralogs.

Ideally, the orthologs are identified by a shared evolutionary history, conserved gene order or synteny within the chromosome, and shared biological function. However, the identification is based on sequence homology and estimation of shared ancestry. The bi-directional best hit (BBH) method has been widely used to discriminate between orthologs and paralogs (99). Sequences that are reciprocally the best BLAST alignment hits for each other between two different genomes are considered BBH (and orthologs). Normalization by BLAST score ratios has been used to filter out biologically irrelevant, distant hits (100). Current pangenome pipelines involve more efficient clustering algorithms to allow the formation of pangenomes for a large number of genomes quickly and efficiently. The actual sequence alignments are not performed all-against-all; instead the algorithms utilize short word filtering and user-set thresholds to determine whether the similarity

between two sequences falls above a certain value (101–103). Some pangenome methods also consider the surrounding genes and synteny to improve the recognition of orthologs (102).

Once genes have been clustered and orthologs and paralogs distinguished the pangenome can be formed. The core genome includes the gene clusters present in all genomes, and depending on the study, the core genome may be defined to include genes present in all (100%) or the great majority of genomes (90-99%). The rest of the gene clusters belong to the accessory genome.

2.3.3 Assessing pangenome

One important characteristic of the pangenome is its openness. The pangenome is considered to be "open" when new genes are added significantly to the total repertoire for each new additional genome and "closed" when the newly added genomes cannot be inferred to significantly increase the total repertoire of the genes. The openness of a pangenome can be calculated using Heaps' law (79). Heap's Law is an empirical law that describes the number of distinct genes in a pangenome as a function of its size and is represented by the formula $n=k*N^{-alpha}$. The k and alpha are empirically determined parameters, n is the expected number of genes for a given number of genomes, and N is the number of genomes (79,104). If the alpha < 1, the pangenome is open and the gene pool of species is not fully included in the genomes analyzed. If the alpha > 1 the pangenome is considered closed (79). Closed pangenomes have been observed in host-restricted organisms with limited genetic variety such as *Salmonella* Typhi (65). Also, intracellular pathogens show a tendency for reductive evolution and genome simplification via gene loss (19,70). Open pangenomes reflect adaptation potential and capability to thrive in different environments. Some important foodborne pathogens such as *Escherichia, Salmonella, Vibrio,* and *Yersinia* carry a wide variety of diverse plasmids that contribute to their large pangenomes (34,94,105). In *E. coli* the fitness of the organism seems to be highest when there is fine-tuning between the chromosomal backbone and the genes newly acquired by horizontal transfer (90,106).

Each pangenome has its distinct structure: the proportion of core genome and accessory genome, and the frequency and distribution of gene clusters. This pangenome structure, or matrix, reflects the lifestyle and adaptive potential within species and is molded by adaptive evolution. The pangenome reflects the lifestyle and adaptive capability within species. The obligate intracellular pathogen *Chlamydia trachomatis* has a core genome taking up to 84% of its pangenome, while perhaps the most common prokaryote on this planet, *Prochlorococcus marinus*, has a core genome that comprises only 18% of its vast pangenome (94).

The total number of core genes within the core genome (of an average genome) is also applicable to the assessment of the genetic diversity among the studied isolates. The core genome becomes smaller when the diversity increases among the input genomes; eventually, the core genome is reduced to the list of essential genes (104). The relative frequency of genes among the extant genomes is often relevant to pangenome assessment. Typically, the genes present in the majority of genomes (>90%) are considered persistent, while the genes present in very few genomes (<20%) might be considered volatile genes. When the pangenome is formed for a monophyletic sample, the majority of genes belong to the persistent or volatile genes, and only a small minority of genes is present in 20-90% of the genomes (29). The presence of volatile genes can indicate that the pangenome analysis contains subclusters or genomes belonging to different genetic lineages.

2.3.4 Challenges in pangenomic approach

In a typical pangenome analysis, several major considerations affect the result: i) the definition and methods used to determine the core genome and the accessory genome; ii) the method used to align and cluster genes to define similarity; iii) the sample of input genomes and the phylogenetic resolution they represent; iv) the type and quality of input genomes (56,96).

i) Selecting the biologically relevant threshold for the core genome depends on the study design and methodology. Therefore, the definition of core genome varies between studies, and comparison of results in different studies should be made with caution.

ii) The pangenome analysis and the resulting numbers for core genome and accessory genome are heavily reliant on chosen ortholog clustering. Suboptimal clustering can lead to core genome clusters being split into multiple groups (underestimating core genome, overestimating accessory genome) (107). To identify true homologs, good pangenome methodologies apply strategies to normalize similarity scores, context-dependent cutoffs, and acquisition of additional information such as protein domains and synteny (107). For identifying homologs, the alignment methods are suitable and useful for closely related genomes with a low count of unique genes.

iii) For good quality pangenome analysis, consideration of study design in terms of sample size and phylogenetic presentation is also needed. Pangenome analysis can be conducted between similar or vastly divergent genomes, and the analysis can either include or exclude distantly related isolates (56,108). The diversity of genomes included strongly affects the results, and the interpretation of results should be adjusted

accordingly to avoid confounding conclusions. In genome data sets with numerous low-abundance gene families, the mixed alignment-free methods with clustering provide the best results (107).

iv) While distantly related isolates introduce biologically relevant diversity to the data set, the diversity of sequencing and annotation techniques can introduce non-biological diversity to the data set. Pangenome analysis as a method is heavily dependent on annotation and discrepancies in gene prediction, and annotation within the genomes can cause significant bias like decreasing the size of the core genome and even affecting whether the pangenome will be predicted to be open or closed (56). Genome annotation is an error-prone process that at the best reflects a snapshot of methods and data available at the time of annotation, and annotation biases including over-prediction and under-prediction of genes, and inconsistently identified gene boundaries have confounded many downstream comparative analyses (56,109).

Caution is required as generalizations or comprehensive quantification (i.e., number of shared genes) from genome sequence comparisons are made. Recently inclusion criteria for input genomes for pangenome analysis have been proposed. As an extreme example, the genomes of genetically modified laboratory strains can confound the pangenome analysis (108).

The methodology used for pangenome analysis should be selected based on the data set, but consolidated guidelines and policies for this have yet to be established.

2.4 From genotype to phenotype

Identifying the function of gene products is one of the fundamental tasks in the post-genomic era and the basis for the interpretation of comparative genome analysis results. The gold standard for gene function prediction is still the wet-lab experiment. The existing computational gene function prediction solutions are an important supplementary technique to this wet-lab work. As more evidence of gene functions is accumulated from experiments, the gene function prediction solutions will become more competent. Computational gene function prediction is especially important for non-model species such as many foodborne pathogens (110–112).

Instead of directly identifying the genetic trait responsible for a certain phenotype, the genome-wide association study (GWAS) statistically identifies the genetic variation associated with the difference in phenotype. Traditionally this approach has been used to study human disease or measurable traits affected by multiple genes, such as height and weight. Association between genotype and phenotype is measured as linkage disequilibrium (LD), and loci are considered to be in LD when their different alleles are statistically differentially associated with the certain phenotype (110,113).

The recent availability of large collections of bacterial genomes has made the application of GWAS a possibility also for microbes as well. WGS data has been successfully used to determine variation associated with traits such as antibiotic resistance by looking for the presence of certain genes such as efflux pumps, insertion–deletions, and other polymorphisms associated with antibiotic resistance (114,115). However, the clonal nature of bacterial populations limits however the resolution of association studies and makes it challenging to apply this approach to bacteria. Linkage analysis is unable to produce reliable results when the number of variants is high and the number of samples is small. These genetic characteristics are often caused by a single gene and horizontally transferable elements making the associated LD smaller. In bacteria, any genetic trait (potentially thousands of SNPs or mutations) is likely in LD with the phenotype within the lineage of similar isolates (113). Therefore the identification of phenotype-associated lineages may be the best possible outcome (116). Saber et al. (117) have suggested that, in a high-recombining population, a sample size of ~ 1,000 is sufficient to detect variants of strong effects (log OR >= 2) such as antibiotic resistance. For more complex phenotypes dependent on several genes or lower heritability (log OR ~1), sample sizes > 3000 are likely needed.

With bacteria, different approaches to investigate the genetic characteristics of interesting phenotypes in less studied and less available organisms are therefore still warranted. The organisms studied in my thesis – clostridia and *Yersinia* – are examples of such organisms where, compared to widely studied organisms such as *Salmonella* and *Listeria*, the scarcity of data and difficulties in laboratory cultivation are still limiting factors (118,119).

2.5 Clostridium botulinum

Clostridium botulinum is an anaerobic, gram-positive, spore-forming bacteria known to cause a rare, but potentially fatal, neuroparalytic disease of humans and animals, botulism (120,121). Botulism is caused via ingestion of preformed botulinum neurotoxin (BoNT) in food (foodborne botulism), or by exposure to the toxin produced in vivo by germinating spores during growth in the intestines (intestinal toxemia botulism, infant botulism), or wounds (wound botulism) (122–124). The predisposing risk factors of affected patients are not important for the development of foodborne botulism, but for other forms of botulism certain risk factors (age, gastrointestinal disease, wounds, injected drug use) are critical (124–126).

The first *C. botulinum* genome was sequenced in 2007 and currently around 477 genome assemblies are available (www.ncbi.nlm.nih.gov/genome/, accession date 20.06.2022) (127). The assembly size ranges from 2.4 to 4.7 Mb with GC% contents ranging from 27.0 to 29.8% (www.ncbi.nlm.nih.gov/genome/), with the average genome including 3860 CDSs (https://www.patricbrc.org/).

2.5.1 Population structure

Early molecular studies revealed that the bacterial species known as *C. botulinum* is a heterogenous, polyphyletic group of bacteria (128,129). Botulism was first described in consumers of sausages (botulus in Latin) in Europe in the 18th century, and until the 1990s, any bacterium expressing neurotoxin BoNT was classified as *C. botulinum,* regardless of its genetic and physiological properties. This trend was discontinued only once certain strains of well-characterized clostridial species, *C. butyricum*, and *C. baratii* were found to express BoNT (130,131). Since then *C. sporogenes* isolates have also been found to carry BoNT encoding genes (132,133).

In total botulinum neurotoxins are produced by at least seven different bacterial groups or species (134). Four of these bacterial groups are still identified as *C. botulinum* strains and form four phylogenetic groups (Groups I-IV). These Groups I-IV should be considered distinct bacterial species, joined by their ability to produce BoNT (51,133,135,136). *C. botulinum* strains are further typed by the serotype of produced BoNT (A–G) (Table 2) (121,137,138). Phylogenetically *C. botulinum* groups include also non-neurotoxigenic strains – in recent pangenomic studies 5% (22/452) and 16 % (33/208) of *C. botulinum* Group I and Group II genomes, respectively, did not carry the BoNT gene (132,133,139).

The Group I strains are monophyletic, and closely related with *C. sporogenes*, of which some strains are also able to produce BoNT (132). The genetic variation between Group I strains is limited (51,52,132,135,140–142) and the Latin name of *C. parabotulinum* has been proposed for Group I strains (134). The pangenome of 556 *C. botulinum* Group I and *C. sporogenes* isolates contained 18 731 genes, with a core genome of 2420 genes (present in >95% of genomes). The core genome comprised around 65.0% of the average genome (132).

The *C. botulinum* Group II, on the other hand, includes two distinct genomic lineages and is more diverse genetically than Group I (51,52,135,139,141–143). One lineage contains type E toxin carrying strains, and the other toxin types B and F, in addition to some E strains (139,142,144). The pangenome of 208 *C. botulinum* Group II isolates contained 16 571 genes, with a core genome of 1768 genes (present in >99% of genomes). The core genome presents around 47.6% of an average genome (139).

Table 2 Groups, toxin serotypes and selected characteristics within *C. botulinum* strains. Modified from Korkeala (136), Carter et al. (137) and Derman et al. (145).

Group, species	Toxin serotype	Proteo-lytic	Saccha-rolytic[a]	Growth temperatures (minimum;optimal; maximum)	Spore D-value (minutes, °C)
I, *C. parabotulinum*	A, B, AB, F, BF, neg	Yes	No	10 °C; 35-40 °C; 41-48 °C	0.1-1, 121 °C <25, 100°C
II, *C. botulinum*	B, E, F, neg	No	Yes	3-8.6 °C; 18-25 °C; 34.7-39.9 °C	1-98, 85 °C; <0.1, 100 °C
III, *C. novyi* sensu lato	C, D, C/D, D/C, neg	Varies	N/A	10-15 °C; 40 °C; N/A	N/A
IV, *C. argentinense*	G, neg	Yes	N/A	N/A; 37 °C; N/A	N/A

N/A: Unknown, information not available, neg: no neurotoxin gene, [a] sucrose, mannose

Fewer genomic and population studies are available on animal pathogenic Group III, or Group IV strains. Group III strains are closely related to *C. novyi* and are considered to belong to *C. novyi sensu lato* together with *C. novyi* and *C. haemolyticum* (37,146–148). Four main phylogenetic lineages within *C. novyi sensu lato* have been identified, all of them including *C. botulinum* Group III strains, and the distinction of *C. botulinum* Group III and *C. novyi* as separate species does not have a

phylogenomic basis (35,37,148). Group IV strains or *C. argentinense* form a separate species cluster, which is related to some *Clostridium subterminale* strains (134,149).

2.5.2 Pathogenic strains

The pathogenicity of *C. botulinum* is based on its capability to produce botulinum toxin BoNT. Group I and II strains, in addition to the BoNT-expressing *C. butyricum*, *C. baratii*, and *C. sporogenes* strains, have been identified as human pathogens (131,132). Group III strains are associated with animal botulism, while Group IV strains have not been associated with clinical disease in animals or humans (136).

The BoNT cluster contains a neurotoxin gene (*bot*) together with accessory proteins in one of the two conserved cluster types (*ha* cluster or *orfX* cluster) (137,138). The BoNT cluster itself can be situated on the chromosome (Group I, II), a plasmid (Group II, IV), or a bacteriophage (Group III) (136–138).

The most commonly applied approach to type *C. botulinum* strains is a multiplex PCR, with target sequences to identify and classify bacterial species and group with primers targeting toxin-production-related genes, such as the *ntnh* and *bot* genes, with the possible addition of flagellar genes (150–153). Pathogenic strains are further characterized by the serotype (A–G) of produced BoNT, and more recently the direct examination of *bot* sequences has led to the identification of more than 40 subtypes within the BoNT serotypes (138). While the alternative methods for active toxin detection in clinical cases have been proposed, in outbreak situations the mouse bioassay is still considered to be the gold standard method for laboratories to confirm botulism, and is thus by far still the method most commonly used by European laboratories (119,154).

The animal pathogenicity of strains within *C. botulinum* Group III is under some debate. Very often, the Group III *C. botulinum* colonies isolated from biological or environmental samples do not carry the *bontC* or *bontD* harboring phage (155). This has led to the hypothesis that the strains that do not produce BoNT might also be pathogenic for animals (35,37).

2.5.3 Epidemiology and reservoirs

Foodborne botulism outbreaks are not frequent but are steadily reported worldwide. According to an epidemiological review by Fleck-Derderian et al. (156), most of the 197 outbreaks reported between 1920 and 2014 occurred in North America (67%), followed by Europe (13%) and Asia (11%). The true prevalence of botulism globally is unknown, but in 2015 18 EU/EEA countries reported 201 cases of botulism and in Canada, approximately 4.3 outbreaks occur annually (157,158). Reviews of

foodborne botulisms in France and Canada have reported more than 400 outbreaks in 30 years (1987-2016), and 91 outbreaks in 20 years (1985-2005) respectively (155,157).

In Europe, animal botulism is contemporarily considered an emerging disease in poultry production (159). A high prevalence of botulism types C, D, and variants D/C and C/D has been reported in farmed and wild birds, and to a lower extent in cattle (154,159–163). However, human botulism cases caused by type C and D toxin-producing strains are exceptional, but not unheard of (155).

C. botulinum Groups I and II are the causative agents of foodborne botulism, and the natural occurrence of cross-contamination of food with *C. botulinum* spores is thought to be the main infection route (132,136,155). *C. botulinum* spores are common in soil and dust, aquatic environments, and in many foods including fish, meat, and honey (164–168). Foodborne botulism is often associated with the combination of insufficient cooking (home-canned and bottled foods) and time-temperature abuse (storing at room temperature) (156,157,169,170). Certain native foods (fermented seal), and fermented or smoked fish, in particular, are considered high-risk (157,170).

There are important epidemiological differences in the prevalence of Group I and Group II in foods. Group I strains producing A toxin are associated with canned and bottled vegetables, mushroom and meat products, and home-canning or bottled items, and are typical causative agents for outbreaks (136,157,170). Group I strains have been isolated (and sequenced) from all six inhabited continents and are considered widespread in the soil globally. Environmental isolates of *C. botulinum* Group I have been isolated, in particular from Argentina and the United States (132).

Group II strains forming type E neurotoxin are associated with healthy fish, fish roe, marine mammals, and the arctic/subarctic environment, but have also recently been isolated from the southern hemisphere (129,140,155,168–170). Meanwhile, Group II strains that produce type B4 neurotoxin have been isolated mainly from European soil, healthy pigs, and farms in addition to marine environments (155,171,172). All in all Group II genomes have been sequenced from 16 countries over four continents (North America, Europe, Japan, and Egypt) (139).

2.5.4 Other characteristics important for food safety

C. botulinum spores are ubiquitous in soil and aquatic environments, but cells grow and produce toxins only under specific conditions that include an anaerobic, low-salt, low-acid environment (164,166,168,173). Bacterial growth is inhibited by refrigeration below 4°C, heating above 121°C, high water activity, or acidity (pH <4.5) (136), but there is variation between different groups (Table

2). Most notably the strain of Group I produces extremely heat-resistant spores, and the Group II strains are psychrotrophic.

C. botulinum is weak for competition, and improperly executed canning and fermentation of foods are particularly conducive to creating the anaerobic conditions that allow *C. botulinum* Group I or II spores to germinate if products are stored at room temperature. Even the hardiest spores produced by Group I strains are inactivated by heating to 121°C for at least 20 minutes (136).

C. botulinum Group II strains pose a hazard in modern food processing as they can grow and produce BoNT at refrigeration temperatures. Ready-to-eat products with mild temperature treatments during processing, reduced use of salt and preservatives combined with extended shelf-life in vacuum packaging, offer a suitable growth environment for these strains (169).

2.6 Clostridium perfringens

Clostridium perfringens is an anaerobic, gram-positive spore-forming bacteria known to cause various human and animal diseases involving the gastrointestinal (GI) system (i.e., enteric diseases). Human enteric diseases include food poisoning; antibiotic-associated diarrhea (AAD); and necrotic enteritis, also known as darmbrand. Well-characterized veterinary enteric diseases include necrotic enteritis in chicken, dogs, and foals and abomasal disease in cattle; and rabbit and sheep enterotoxaemia. In addition to enteric diseases, *C. perfringens* causes gas gangrene and wound infections (174–178).

The first *C. perfringens* genome was sequenced in 2001 and currently around 405 genome assemblies are available (www.ncbi.nlm.nih.gov/genome/, accession date 20.06.2022) (178). The assembly size ranges from 2.7 to 4.2 Mb with the GC% content ranging from 24.3 to 31.1% (www.ncbi.nlm.nih.gov/genome/). The average genome includes 3146 CDSs (https://www.patricbrc.org/).

2.6.1 Population structure

The population structure of *C. perfringens* has been studied widely using WGS and MLST methods (76,105,179–182). Five lineages of *C. perfringens* have been described within data sets of 56-206 genomes, but unanimous nomenclature for these lineages is yet to be established (76,105,182).

Recently, Feng et al. (76) estimated that *C. perfringens* originated 40-80 000 years ago. Curiously, strains similar to the current isolates have been isolated from a 12 000-year-old mummified puppy (lineage I) and a 5000-year-old mummified human corpse (lineage IV) (76,183). The estimated core genome (present in >95% of the genomes) of *C. perfringens* comprises one-third of an average genome and the pangenome is notably large (76,105,184).

2.6.2 Pathogenic strains

The genetic lineages have not yet been established in use for the classification of *C. perfringens* strains. Instead, five toxins (alpha, beta, epsilon, iota, enterotoxin *cpe*) are used to toxinotype strains to types A-G (Table 3) (38). Most major toxins and virulence factors carried by *C. perfringens* belong to its large accessory genome. Chromosomal toxin alpha (plc) is present in all *C. perfringens* strains. Three of the major toxins (beta, epsilon, and iota) are located on a family of conjugative plasmids

(81,185), and the enterotoxin gene *cpe* can be located either on a plasmid or the chromosome. Since three out of five toxins used for toxinotyping are located on horizontally transferable plasmidial elements, they do not reflect the genetic relatedness of strains. To further distinguish different genetic lineages of strains, a virulence gene profile scheme (I-XV) has also been proposed more recently (186) (Table 3). This virulence gene profile has not yet been established in use for typing *C perfringens*.

Table 3 Toxinotyping A-G (38) and virulence gene profiling I-XV (186) of *C. perfringens*

Toxin type	Virulence gene profile	*plc*	*cpb*	*etx*	*iap, ibp*	*cpe*	*netB*	*cpb2*	*lam*	*pfoA*	*nagH*	*nanI*	*nanJ*
A	VIII	+	-	-	-	-	-	+	-	+	+	+	+
A	IX	+	-	-	-	-	-	+	-	+	-	+	-
A	X	+	-	-	-	-	-	-	-	+	+	+	+
A	XI	+	-	-	-	-	-	-	-	-	+	+	+
A	XII	+	-	-	-	-	-	-	-	+	-	+	-
A	XIII	+	-	-	-	-	-	-	-	-	+	-	-
A	XIV	+	-	-	-	-	-	+	+	+	+	+	+
A	XV	+	-	-	-	-	-	-	-	-	-	-	-
B	N/S	+	+	+	-	-	-	(+)	(+)	(+)	N/S	N/S	N/S
C	N/S	+	+	-	-	(+)	-	(+)	-	(+)	N/S	N/S	N/S
D	N/S	+	-	+	-	(+)	-	(+)	(+)	(+)	N/S	N/S	N/S
E	I	+	-	-	+	(+)	-	-	-	+	+	+	+
F	II	+	-	-	-	+	-	+	-	+	+	+	+
F	III	+	-	-	-	+	-	-	-	+	+	+	+
F	IV	+	-	-	-	+	-	-	-	-	+	-	+
F	V	+	-	-	-	+	-	-	-	-	+	-	-
F	VI	+	-	-	-	+	-	-	-	-	-	-	+
F	VII	+	-	-	-	+	-	-	-	-	-	-	-
G	N/A	+	-	-	-	-	+	-	-	(+)	N/S	N/S	N/S

+: present, –: not present, (+): variable, N/S: status not specified

Both food poisoning and AAD cases of *C. perfringens* diarrhea are primarily caused by pore-forming *C. perfringens* enterotoxin (CPE) producing type F strains (formerly enterotoxin carrying type A strains), in which the *cpe* gene can be located either chromosomally (c-cpe strains) or be carried by a family of conjugative toxin plasmids (36,187,188). There are three described cpe-carrying plasmids: pCPF5603 (IS1151), pCPF4969 (IS1470-like), and pCPBB-1 (36,81,189). These *cpe*-carrying strains are genetically heterogeneous, and the frequently cited estimation by McClane (190) is that only 5% of *C. perfringens* strains carry the *cpe* gene. However, all strains are presumed to be capable of carrying the *cpe* gene and producing enterotoxin. Within *cpe*-carrying strains, the chromosomally *cpe*-carrying (c-cpe) strains form a separate lineage IV and are monophyletic (76,179,180).

2.6.3 Epidemiology and reservoirs

C. perfringens is a relevant human enteric pathogen: approximately 15% of reported AAD cases are caused by *C. perfringens* and the number of *C. perfringens* food-poisoning outbreaks in the European Union has been estimated at around 850 000 to 5 million cases per year (191,192). While major outbreaks and small-scale outbreak cases are frequently reported, the source of infection or entry to the food system is often not identified. Generally, *C. perfringens* is considered to be an environmental bacterium associated with soil, water, sewage, and dust; but also, the GI tract of humans and animals (120). However, the transmission routes and reservoirs of enteropathogenic strains of *C. perfringens* are not fully known.

Both plasmid-borne *cpe*-carrying (p-cpe) and c-cpe strains cause foodborne outbreaks and the suggested reservoirs for *cpe*-positive strains include healthy animals and humans, sludge, soil, and retail meat (179,180,193–198). The reservoirs are likely different for p-cpe strains and c-cpe strains.

The p-cpe strains are often associated with non-food-borne human disease and animal enteric disease and the GI tract of animals has been suggested as their adapted niche (180,199). Humans have also been suggested as a reservoir for *cpe*-carrying strains and pCPF5603 plasmid-carrying strains in particular have been associated with care home outbreaks and AAD (36,200,201). A recent study on human enteric isolates suggested that *C. perfringens* strains, carrying *cpe*-carrying plasmid pCPF5603, might also persist and cause longitudinal outbreaks spanning several years, supporting the hypothesis of healthy humans serving as reservoirs (201).

Contrarily the c-cpe strains have mainly been isolated from samples associated with food items or food poisoning, and very few strains have been isolated from healthy humans or domestic animals

(200). Lahti et al. (180) have suggested that the c-cpe strains are specialized to a yet unknown environmental niche.

2.6.4 Other characteristics important for food safety

C. perfringens is a mesophilic bacterium and the ideal temperature for strains is around 43°C Psychrotrophic strains have not been reported and refrigeration is an efficient way to prevent the growth of *C. perfringens*. However, *C. perfringens* is unusually fast to reproduce and its generation time can be as short as 7.1 min. *C. perfringens* is metabolically very active: it is a strong gas producer and able to utilize a variety of carbohydrates and proteins (136,178). *C. perfringens* is anaerobic but able to tolerate low amounts of oxygen especially in food items with low redox potential (136)

For the epidemiology of foodborne illness, the spores produced carry the most significant role. The c-cpe strains in particular are considered more stress-resistant than other *C. perfringens* strains (202), which is one of the explanations for their association with foodborne illness. Hardiness and spore heat resistance are thought to protect c-cpe strains during food production and cooking, and enable proliferation during potential temperature abuse after cooking. Both vegetative cells and spores of representative c-cpe strains show higher overall resistance to heat or other environmental stresses such as pH and pressure than p-cpe strains (196,202–205). The heat resistance of *C. perfringens* vegetative cells and spores has been studied relatively widely, and the topic is further discussed in chapter 2.8.3.

2.7 Enteropathogenic *Yersinia*

The genus *Yersinia* belongs to the family *Enterobacteriaceae* and the *Yersinia* are gram-negative, facultative anaerobic rod-shaped, non-capsulated, and non-sporulating bacteria (206). Genus *Yersinia* includes 18 species of which three are human pathogens. *Yersinia pestis* is the causative agent of bubonic and pneumonic plague and is the most infamous member of this genus. However, enteropathogenic *Yersinia enterocolitica* and *Yersinia. pseudotuberculosis*, are also significant human pathogens (207).

Both species of enteropathogenic *Yersinia* cause a gastro-intestinal disease called yersiniosis, which is the fourth most frequently reported zoonosis in the European Union (208). Typical signs are fever, abdominal pain, and diarrhea; the illness is often clinically indistinguishable from acute appendicitis, which may result in unnecessary appendectomies (209). *Y. pseudotuberculosis* is also linked with a severe systemic disease known as Far East scarlet-like fever and has only been recognized as a foodborne pathogen since 2004 (210,211).

The first *Y. enterocolitica* genome was sequenced in 2006 and currently around 246 genome assemblies are available (212). The assembly size ranges from 3.8 to 6.1 Mb with GC% content ranging from 46.7 to 49.3% (www.ncbi.nlm.nih.gov/genome/, accession date 20.06.2022). The average genome includes 4466 CDSs (https://www.patricbrc.org/).

The first *Y. pseudotuberculosis* genome was sequenced in 2004 and currently around 96 genome assemblies are available (67). The assembly size ranges from 4.2 to 5.1 Mb with GC% content ranging from 47.1 to 47.7% (www.ncbi.nlm.nih.gov/genome/, accession date 20.06.2022). The average genome includes 4548 CDSs (https://www.patricbrc.org/).

2.7.1 Population structure

The enteropathogenic *Yersinia* diverged around 41–185 million years ago, and the third human pathogenic species of the *Yersinia* genus, the infamous *Y. pestis*, is a relatively recent clone of *Y. pseudotuberculosis* (213). Diversification of *Y. pestis* from *Y. pseudotuberculosis* is a typical example of adaptive evolution and niche restriction (67,214).

Y. enterocolitica is a heterogeneous species including two subspecies, based on the 16S rRNA gene sequence: *subsp. enterocolitica* and *subsp. palearctica*, in addition to six biotypes (BT1A, BT1B, and BT2-5) corresponding to six phylogenetic groups of varying pathogenicity (66,215,216). Biotypes

are further classified into over 30 serotypes based on the variation of the O antigen (217,218) (Table 4). Genetically and serotypically, the biotype 1A is the most heterogeneous and this group of environmental strains is thought to act as a reservoir for pathogenic strains, frequently acting as donors in recombination events (66,219).

Contrary to *Y. enterocolitica,* all *Y. pseudotuberculosis* strains are considered pathogenic and the species is homogenous. Serotypes are used to distinguish between different strains (206,220–222) (Table 4).

2.7.2 Pathogenic strains

The evolution of enteropathogenic *Yersinia* including *Y. pestis* is thought to have included multiple virulence gene acquisition events (pYV, *ail*) that have separated the pathogenic strains from the environmental, non-pathogenic lineages. This current hypothesis of parallel evolution(222) has rejected the previous one, which suggested that all pathogenic *Yersinia* species share a common pathogenic ancestor (33,222).

Identification of *Yersinia* spp. is mainly based on biochemical tests and serotyping, but molecular methods have increasingly also been used. Isolation of *Yersinia* from food and environment samples is challenging due to the long cultivation times required and some of the strains not thriving on commonly used agar plates (118). Thus, PCR methods with primers targeting virulence genes have been developed for both species (225,226). For serotyping, the multiplex PCR or WGS data-based methods targeting O-antigen are now preferred compared to the conventional serotyping as these approaches eliminate the issues with antiserum and antisera cross-activity and improve the proportion of typed strains (227,228).

The potential pathogenicity of the strains is mostly confirmed by PCR based on essential virulence genes. The virulence determinants are both chromosomally and plasmid-encoded. The pathogenic *Yersinia* carry a 70-kb virulence plasmid (pYV), which encodes type III secretion system Ysc and is essential for their survival and ability to multiply in different lymphoid tissues of the host (207,229). Customarily the strains that do not carry pYV are considered non-pathogenic (207,230). The pYV plasmid is considered the hallmark for pathogenicity of *Yersinia*, but even strains lacking this plasmid produce enterotoxin, invade epithelial cells, and survive within macrophages, suggesting some lever of virulence and pathogenicity (219).

Table 4: Typing of enteropathogenic *Yersinia* (219,223,224).

Species	Genetic group	Biotype	Serotype(s)	Biotype test results [a]
Yersinia enterocolitica	1	BT1A	O:1,2,3; O:3; O:4; O:4,33; O:5; O:6,30, O:6,31; O:7,8; O:7,13; O:8,19; O:10; O:11; O:11,13; O12,25; O:14; O:16; O:19,8; O:21;O:22, O:25, O:30; O:31; O:34; O:36; O:37, O:41,42, O:41,43; O:46; O:47, O:57, O:63; O:65, O:66; O:72, NAG[b]	+/+/+/+/+/+/+/+
	2	BT1B	O:4, O:8, O:13, O:20, O:21	+/+/+/+/+/+/+/+
	3	BT2, BT3	O:1, O:2, O:3, O:9, O:5,.27	BT2: -/-/+/+/+/+/+/+ BT3: -/-/-/+/+/+/+/+
	4	BT4	O:3	-/-/-/-/-/+/+/+
	5	BT5	O:2, O:3	-/-/-/-/-/-/-/-
Yersinia pseudotuberculosis	1	I or IV	O:1b, O:3, O:5a, O:5b, NAG	-/+/V
	2	I or IV	O:1a, O:1b, O:3, O:5b, O:13, O:14	-/+/V
	3	I or IV	O:1a, O:1b, O:2a, O:2b, O:2c, O:3, O:4a, O:4b, O:5a, O:5b, O:6, O:7, O:10, NAG	-/+/V
	4	II or III	O:1b, O:5a, O:5b, O:6, O:7, O:9, O:10, O:11, O:12	V/-/-
	5	II or III	O:3	V/-/-
	6	I or IV	O:1b, O:2a, O:2b, O:2c, O:3, O:4a, O:4b, O:5a, O:5b, O:6, O:7, O:10, O:11, O:13, NAG	-/+/V

[a] Results: +, positive for utilization/fermentation; -, negative; V, variable. Biotyping tests for *Y. enterocolitica*: salicin, esculin hydrolysis, indole production, xylose, lactose, nitrate reduction, trehalose, and sorbitol. Biotyping tests for *Y. pseudotuberculosis:* citrate, melibiose, and rhamnose

[b] NAG, non-agglutinable

Additionally, a high pathogenicity island HPI is found uniformly found in the highly pathogenic *Y. enterocolitica* strains (1B/O:8) and frequently in *Y. pseudotuberculosis* O:1 and O:3 strains (231,232). The evolution of non-pathogenic host-generalist to pathogenic host-restricted *Y. enterocolitica* is marked by a concomitant reduction in metabolic capacity through deletion and pseudogene formation and the expansion of the IS1667 element (222).

2.7.3 Epidemiology and reservoirs

While the gastrointestinal disease (yersiniosis) caused by enteropathogenic *Yersinia* is common, the true incidence of *Y. enterocolitica* and *Y. pseudotuberculosis* infections is unknown as the clinical diagnosis is often not confirmed and the disease is not notifiable (227,230).

Human and animal cases of yersiniosis are mainly sporadic and larger outbreaks are relatively rare. Often the sources of *Yersinia* infection are not identified or even investigated (227,230). Non-pathogenic *Yersinia* species and non-pathogenic strains of *Y. enterocolitica* have been abundantly isolated from food and environmental samples, but pathogenic strains of both species have mostly been associated with animal reservoirs (230,233–236).

Different *Y. enterocolitica* phylogroups or corresponding bioserotypes are associated with different animal reservoirs (230). Studies on recombination patterns and the accessory genome in *Y. enterocolitica* have led to the conclusion that phylogroups are ecologically separate and that certain genetic determinants have been fixed during lineation (66). Phylogroups also show evidence of gene decay and metabolic reduction (ABC transporters, dimethyl sulfoxide metabolism, etc.) which has been interpreted as evidence of niche adaptation (66). The pathogenicity of *Y. enterocolitica* strains varies from non-pathogenic to highly pathogenic, thus detection of virulence markers is also needed to determine the clinical significance of the isolated strains.

Based on current understanding, the most common strain types associated with human infections are *Y. enterocolitica* bioserotypes 2/O:9, 2/O5,27, 3/O:3 and 4/O:3, and *Y. pseudotuberculosis* serotypes O:1 and O:3 (118,206,230,237). The first reported *Y. enterocolitica* outbreaks were caused by bioserotype 1B/O:8 in North America and this bioserotype has also manifested commonly in Poland but has nevertheless fallen behind 2/O:9 and 4/O:3 in reported occurrence and epidemiological significance (209,238). The occurrence of bioserotypes varies between countries and the frequency of human and animal infections caused by certain *Yersinia* subgroups is likely to be related to the frequency of exposure to specific animal sources (230).

Y. enterocolitica infections are thought to be largely foodborne, even though pathogenic strains have not been frequently been isolated from foods. Different bioserotypes are associated with different animal reservoirs (230). *Y. enterocolitica* 4/O:3, which is the most common bioserotype for human infections in Africa, Europe, Japan, and Canada is associated with pigs in an almost host-restricted manner and *Y. enterocolitica* 2/O:9 is associated with domestic ruminants (230,239–244) . Outbreaks have been caused by pork products, pasteurized milk, tofu, and ready-to-eat salad mix (245–251).

Pigs and wild boar have also been identified as important reservoirs for *Y. pseudotuberculosis* (230,244,252). Foodborne outbreaks caused by *Y. pseudotuberculosis* have been reported in recent decades. *Y. pseudotuberculosis* has been isolated from several animal species (pigs, rabbits, and hare, rodents, birds) and the reservoir of *Y. pseudotuberculosis* is thought to be wild animals and also farms (220,230,233,244,253,254). During a *Y. pseudotuberculosis* O:1 raw milk outbreak in Finland, the same serotype was found in both milk and cattle feces from the same farm (255). In a large outbreak of *Y. pseudotuberculosis* caused by raw carrots the epidemic strain was also traced to soil and the production line at the farm, but while contamination by wildlife feces was suspected as the origin of this outbreak, it was not successfully confirmed (256). *Y. pseudotuberculosis* outbreaks have been linked to raw milk and vegetables (iceberg lettuce, carrots) and have been reported in Canada, Finland, New Zealand, Japan, and Russia (118,210,211,255–261). Also, outbreaks among captive animals have been frequently reported (118). Different serotypes of *Y. pseudotuberculosis* are not thought to have different reservoirs.

2.7.4 Other characteristics important for food safety

Enteropathogenic *Yersinia* are psychrotrophic and able to prosper in cold storage. The optimum growth temperature of *Yersinia* is 28-29°C and their growth range is 4-42°C, but both *Y. enterocolitica* and *Y. pseudotuberculosis* can grow at temperatures below 0°C (206,262–264). Testing by Bergann et al. (262) showed that 13% of tested strains were still able to grow at -5 °C. Both bacteria survive well in changing temperatures and endure repeated refreezing and thawing (206,265).

Both species are sensitive to heat treatment and do not produce spores. Their growth is easily inhibited by salt (> 7%) or nitrite (> 80 ppm) (206,266). Temperature-dependent motility is one of the hallmarks of enteropathogenic *Yersinia:* cells are peritrichously flagellated at 25 °C, but non-flagellated and nonmotile, at 37 °C (267).

2.8 Temperature stress resistance

2.8.1 Growth at low temperatures

Psychrotrophic foodborne pathogens, such as enteropathogenic *Yersinia* and *C. botulinum* Group II strains, which can survive and multiply at low temperatures, pose a risk in modern food systems, where cold chains are used to extend the shelf lives of food products. The current trend of minimal food processing and ready-to-eat products has further stressed the importance of the cold chain in maintaining food safety (169,268,269). Cold resistance and growth in cold temperatures are epidemiologically important, as the initial contamination level can increase to a clinically significant level in packaged food and cold stored products in as little as three days (255).

The ability to grow and survive at refrigeration temperatures, and to tolerate repeated freezing and thawing, varies among bacterial strains. In general, enteropathogenic *Yersinia* are able to grow at 0-4°C (206,269–271), *C. botulinum* Group I at 12.8-16.5°C (272), and Group II at 6.2 to 8.6°C (145). Cold resistance is thought to be a genomic adaptation dependent on many genes, but variation in observed cold resistance and cold growth has been reported for example in Group I *C. botulinum* strains (145) and *Listeria monocytogenes* (273), and to a lesser extent in *Y. pseudotuberculosis* (269).

2.8.2 Cold resistance and cold shock

An abrupt decrease in temperature causes a so-called cold shock response in bacteria, during which certain proteins are induced (274). An important group of cold resistance genes are the cold shock proteins (Csps), which are in many bacteria known to be central for cold shock survival and subsequent adaptation to low temperature. In addition to Csps, the fall in temperature induces changes in the transcriptome, RNA chaperones are induced, and cell membrane fluidity is adjusted by changing fatty acid composition and desaturations (274–277).

Csps are a highly conserved group of small nucleic acid-binding proteins that have been identified in a variety of bacteria, including psychrophiles, mesophiles, and thermophiles (274,278,279). Most studies of Csps have been conducted in mesophilic bacteria such as *E. coli* (280,281). Understanding how the Csps of psychrotrophic bacteria differ from those of mesophilic bacteria is important, for this may represent a common feature of psychrotrophic bacteria that distinguishes them from mesophiles (278). Furthermore, a deeper understanding of the Csps of psychrotrophic bacteria could contribute to preventing their growth in refrigerated food or in optimizing the production of their enzyme for

biotechnological purposes. Cold shock response and cold adaptation have been studied extensively in enteropathogenic *Yersinia* and *C. botulinum*, but still, the genetic traits that explain the differences in cold tolerance between the strains are still not fully understood (269,282–286).

Long periods of storage of fresh produce such as carrots at low temperatures have been associated with large *Y. pseudotuberculosis* outbreaks (256,259). These outbreaks and other foodborne outbreaks have been caused mainly by O1 strains (210,255–257,259,260), which are able to grow at a faster growth rate with a shorter lag period compared to O3 strains (118). This physiological diversity or adaptation is poorly understood, as previous studies have identified little genetic variation or lineages within *Y. pseudotuberculosis* isolates (221,287).

For the growth of *Y. enterocolitica* at low temperature, the regulators for signal transduction, chemotaxis, biodegradative metabolism, and defense against oxidative stress are known to be activated (288). Curiously also the expression of insecticidal toxins is also induced in cold, suggesting an ecological role in survival outside the mammalian host (289).

2.8.3 Heat resistance

Cooking is the oldest and most common method to destroy microbes in food, and survival during heat treatment and heat resistance are therefore important phenotypic traits for many foodborne pathogens. Most bacteria present in the food chain are destroyed by heating at 60 to 70 °C, making their destruction by cooking relatively easy. From the point of view of food safety, the problem is the bacterial spores that survive thermal processing at 70 to 130 °C.

As described in the earlier part of this dissertation, both *C. perfringens* and *C. botulinum* produce spores, and some of the strains such as chromosomally *cpe*-carrying *C. perfringens* strains, *C. botulinum* Group I strains produce extremely heat-resistant spores. For control of *C. botulinum* spores in canned food, a "botulinum cook" in a pressure cooker at 121 °C (250 °F) for 3 minutes has become the food industry standard. Genes associated with spore heat resistance in sporulating bacteria focus on three properties: (i) DNA damage prevention and repair (ii) dipicolinic acid (DPA) and cation concentrations in the spore core and (iii) the water content of the spore core (290–292).

The differences in heat resistance of *C. perfringens* vegetative cells and spores have been studied relatively extensively, and D-values for 25 strains have been published (203,293–295). The main finding is that c-cpe strains (10 strains) produce heat-resistant spores with high $D^{99°C}$-values averaging 53.2 (203). Contrarily, p-cpe strains and cpe negative strains produce spores, with low D99°C-values with an average of 1.0 (203). Vegetative cells of c-cpe strains are reported to have 1.5

to 2 folded $D^{55°C}$-values compared to p-cpe strains (203,294). A key mechanism behind this extreme spore resistance phenotype is the ability of strains to produce an unusual variant of small, acid-soluble protein 4, which binds particularly strongly to spore DNA to protect it from damage (295–297).

The extreme spore heat-resistance in *C. botulinum* Group I strains has also been studied extensively and a meta-analysis of studies included D-values for 11 strains in 38 studies. The mean D-values at the reference temperature of 121.1°C, in liquid media and pH neutral, was 0.19 min (298). The genetics behind the heat-resistance have been studied for *Clostridium sporogenes* PA3649, which produces extremely heat-resistant spores, is closely related to *C. botulinum* Group I strains and is often used in testing commercial thermal food processing procedures for their ability to prevent foodborne botulism. The most significant difference between *C. sporogenes* PA3649 and other *C. sporogenes* isolates was the acquisition of a second *spoVA* operon, *spoVA2*, which is responsible for the transport of DPA into the spore core during sporulation (299).

While certain genes have been found to explain heat-resistance phenotypes of produced spores the presence of outlier strains (203,298) has suggested that other factors including differential expression, altered function of sporulation proteins, and/or additional novel sporulation proteins could be involved (299).

3 AIMS OF THE STUDY

It is now acknowledged that environmental bacterial pathogens infect humans only incidentally. The driving force for evolution within their populations is not virulence against mammal hosts, but rather survival in their adapted reservoirs. However, the adapted reservoir and ecology of many environmental, foodborne pathogens remains poorly understood or undiscovered.

This study aimed to assess the suitability and prospects of an in silico comparative genome analysis and pangenomic approach for a better understanding of the ecology, evolution, and epidemiology of environmental foodborne pathogens. The concluding objective was to understand what we can determine about the pathogens, their reservoirs, and adapted lifestyle based on their genomes through interpretation of genetic differences and their functionalities. The opportunistic, environmental foodborne pathogens: *Clostridium botulinum, Clostridium perfringens*, *Yersinia enterocolitica,* and *Yersinia pseudotuberculosis* were selected as study organisms to approach this subject.

The specific aims of the sub-studies in this thesis were

1. To assess the population structure and pangenome of *C. botulinum* through comparative genome sequence analysis and to investigate how genomes of psychrotrophic *C. botulinum* strains differ from non-psychrotrophic strains.

2. To investigate the population structure and genomic epidemiology of *C. perfringens cpe*-carrying strains, and to study the genomics of spore heat resistance in *C. perfringens*.

3. To identify genetic differences between enteropathogenic *Yersinia* isolates from swine and other sources, and between environmental and non-environmental isolates, and to assess the suitability and limitations of whole-genome hybridization for comparative genome analysis.

4. To characterize the genes important for cold shock response and cold adaptation of *Y. pseudotuberculosis* strain IP32953 by transcriptome analysis.

4 MATERIALS AND METHODS

4.1 Strains and sequencing (I-IV)

For study I on the *C. botulinum* pangenome, 15 publicly available and one newly sequenced genome were used. The genomes were downloaded from the Patric database (300). The new genome included was *C. botulinum* II E CB11/1-1, which was isolated in a Finnish foodborne botulism outbreak (301). The strains studied represented Group I (n = 10), Group II (n = 4), and Group III (n = 2). Strains I A1 ATCC 3502 for Group I, and II E Alaska for Group II strains were used as reference genomes. The CB11/1-1 genome was sequenced using 454 Genome Sequencers GS20 and GS Flx (10). The reads obtained were assembled using GS *De novo* Assembler (version 1.1.03), resulting in a genome of 3.9 Mbps in 639 contigs.

For study II on *C. perfringens,* 30 *C. perfringens* strains were assayed for heat resistance and growth temperature, and their genomes were sequenced. Whole-genome sequencing was performed using PacBio RSII (Institute of Biotechnology, Helsinki, Finland). Sequenced genomes were assembled using HGAP3 and checked for circularity using GAP4 (302,303). To improve the draft assembly, Illumina MiSeq reads and Pilon tool were used for genome polishing (304). Complete genomes were 2.9 to 3.6 Mbps with one chromosome and 0 to 5 putative plasmid contigs. During phylogenetic and pangenome analysis publicly available genomes (n = 260), representing all described genetic lineages including 56 *cpe*-carrying isolates, were used. Strains str.13, ATCC 13124, and SM101 were used as reference genomes (178,305). The genomes were downloaded from the Patric database (300).

Sequenced genomes in I and II have been deposited to the GenBank.

For study III the seven gnomes publicly available at that time were used to design a DNA microarray. The genomes were downloaded from the Patric database (300). For comparative genome hybridization, 61 *Y. enterocolitica* and 38 *Y. pseudotuberculosis* strains isolated from a variety of sources were used. Strains *Y. enterocolitica* subsp *enterocolitica* 8081, *Y. enterocolitica* subsp *palearctica* Y11, and *Y. pseudotuberculosis* IP32953 were used as reference genomes and positive hybridization controls (67,212,306). Hybridized strains were selected from the *Yersinia* strain collection of the Department of Food Hygiene and Environmental Health to represent the various biotypes and serotypes of enteropathogenic *Yersinia*. In total, 41 strains represented the most common pathogenic *Y. enterocolitica* bioserotype 4/O:3. The majority (79/99) of the strains had been isolated

from swine or swine slaughterhouses. The rest of the strains (n = 20) had been isolated from human patients, wild birds, and other animals.

For transcriptome study IV, the strain *Y. pseudotuberculosis* IP32953 was used.

4.2 Annotation and ortholog identification (I, II, and III)

For annotation and proteome comparison, curated databases such as PATRIC, Uniprot and KEGG were utilized in studies I, II, and III. The annotation of protein-encoding genes was an important initial point for studies, and either the Prokka pipeline (42) or RAST (44), as implemented within the Patric annotation pipeline (307), were used to annotate newly sequenced genomes or to reannotate previously available genomes. Proteomes were compared using bi-directional best hits identified with BLAST as described by Ward (308) to distinguish positional homologs, orthologs, and paralogs. To explain this approach shortly, two genes were acknowledged as orthologs if a reciprocal best BLAST hit existed among them, and both hits scored over a set threshold (100). In studies, I and III the bi-directional best hits were identified using an in-house script, whereas in study II Patric proteome comparison tools and available pangenome pipelines were used (102,103,300,307). Studies I, II, and III required clustering of orthologous proteins and for this BLASTCLUST and CD-Hit were used (41,101).

BLASTCLUST does clustering by performing the exhaustive BLAST all-to-all pairwise alignments, which means that it is slow but accurate (308). CD-HIT was originally developed for clustering protein sequences to create a reference database with reduced redundancy (101). CD-HIT is widely applied and one of its applications lies within comparative genomics. This greedy incremental algorithm starts with the longest input sequence as the first cluster representative and then processes the remaining sequences from long to short to classify each sequence as a redundant or representative sequence based on its similarities to the existing representatives. Here CD-HIT was used to cluster proteins within enteropathogenic *Yersinia* to create a database for pan-microarray probe design (III).

CD-HIT is also used for clustering proteins within Roary and Panaroo pipelines (II). The Roary pipeline takes the annotated genomes as input and pre-clusters the proteins with CD-HIT, performs an all against all comparison using BLASTP, and then clusters those results with the Markov Clustering algorithm (103). Within the Panaroo pipeline, the neighboring nodes and contextual support of identified clusters are used to collapse diverse gene families and to merge and identify

fragmented or mistranslated genes (102). The idea behind this contextual curation of identified clusters is to produce robust clusters.

4.3 Comparative genome analysis (I, II, III)

To compare genomes and assess their encoded genetic toolsets, different approaches described in original articles I, II, and III were used.

For ease of comparison in this dissertation, the analyses for the datasets in studies I, and III were repeated with the same pangenome pipelines as in II. The Roary pangenome pipeline results are visualized with pangenome matrix and gene cluster frequency graphs in Figures 3, 4, and 5. The pangenome matrix shows the gene profile of each strain against phylogenetic tree and the distribution of core and accessory genome within the pangenome. The gene cluster frequency graph shows the relative frequency of genes among the extant genomes (core, persistent, and volatile clusters).

To consider the conserved gene order (synteny) of presumed orthologs, the genomes were aligned and visually compared with progressive Mauve (309) and SEED Viewer 2.0 (310). Also, alignment and comparison tools available at Patric were utilized (300,307).

4.4 Heat resistance assay (II)

For study IV, the spore heat-resistance phenotype of 30 *C. perfringens* strain was determined according to the established procedures (203,293). All strains were prepared as described by Duncan and Ando (293,311). Duncan-Strong (DS) medium cultures were prepared and grown for 24 hours at 37 °C, then heat-shocked at 75°C for 15 minutes to kill the remaining vegetative cells and facilitate spore germination. A sample was withdrawn, diluted, and plated to determine the initial spore count. The remainder of each heat-shocked DS medium culture was then heated at either 89 or 99 °C for periods ranging from 1 min to 4 h or until the spore count reached 0 (depending on the individual isolate and the temperature used). At each time point, the culture was mixed and the time point sample was withdrawn and diluted (dilution range, 10^{-2} to 10^{-7}). For vegetative cells, FTG (fluid thioglycolate) medium cultures grown for 24h at 37 °C were heated to 60 °C for a time ranging from 1 min to 2 h. The time point samples were withdrawn as explained previously. All dilutions were plated to determine the number of viable spores and cells present per milliliter (CFU/ml). The

logarithmic counts of viable spores and cells of every isolate were graphed against heating time to determine the slope of the survival curve. The estimation of the D value was determined from this curve. P values were calculated using ANOVA analysis of variance (IBM SPSS Statistics for Macintosh, Version 27.0).

4.5 Minimum and maximum growth temperatures (II)

The minimum and maximum growth temperatures for 30 *C. perfringens* strains were determined in study IV. The strains were examined using the Gradiplate W10 incubator (Biodata Oy, Helsinki, Finland) placed in an anaerobic workstation (MK III, Don Whitley Scientific, Ltd, Shipley, UK) to determine the minimum and maximum growth temperatures for the strains studied (145,272,312). Briefly, the strains were cultured in FTG at 37°C for 24 hours. Strains were refreshed and grown at 37 °C until turbidity OD_{600} reached 0.6-0.9. Cultures were diluted, and a sample was transferred to a Gradiplate cuvette. The temperature gradients tested for minimum and maximum growth temperatures were 8 – 18 °C and 47– 57 °C, respectively. Strains were incubated for 2 days for maximum temperatures and 21 days for minimum temperatures. Growth boundaries were determined using a stereomicroscope, and the growth temperature threshold was determined as the boundary where dense bacterial growth was discontinued. The formula $T = T_{low} + d * g$ was used to determine the minimum and maximum growth temperatures as described by Korkeala et al. (312). The d in the formula is the distance (in millimeters) of the growth boundary to the measurement point of T_{low}, and g is the temperature gradient.

4.5 Comparative genome hybridization and analysis (III)

4.5.1 Microarray design and genome hybridization (III)

In study III a multi-strain microarray was designed to perform a comparative whole-genome hybridization (WGH) of enteropathogenic *Yersinia* strains. The DNA microarray was designed based on the seven genomes and 14 plasmid sequences available at the time. Sequences (n = 29 786) were clustered (95% identity, 80% minimum alignment of the longer sequence) using Cd-hit-est (101) into 11,564 gene groups. Stringent clustering parameters were chosen to avoid problems with

uncomplimentary probes in the probe design. Probes were designed using Agilent Technologies Gene Expression Probe Design. A probe (45–60-mer) was designed to represent and separate each identified gene group (n = 11,564). Redundant gene groups were removed (n = 13, 14 CDSs) and, for long gene sequences (> 10 000 bases, n = 11), a tiling method was used to design extra probes (10 per sequence). The final custom arrays contained 11,661 probes on each of the subarrays on Agilent 8 ∗ 15 K chips (Agilent, Santa Clara, CA, USA).

The custom arrays thus created were hybridized with the extracted genomic DNA (313). Hybridization data analysis followed the routines set by Lindström et al. (52) and Lahti et al. (180).

The genomic DNA of *Yersinia* strain was fluorescently labeled with the BioPrime ArrayCGH labeling module (Invitrogen, Carlsbad, CA, USA) using either Cy3 or Cy5 (GE Healthcare, Buckinghamshire, UK). Labeled samples were hybridized at 65 °C for 16 hours and washed according to the manufacturer's instructions.

4.5.2 Data analysis (III)

The hybridized slides were scanned (Axon Genepix Autoloader 4200 AL, Molecular devices Inc., Sunnyvale, California, USA) with a resolution of 5 μm. Images were processed and manually checked using GenePix Pro 6.0/6.1 software. For data analysis, R software and the LIMMA package were used (314). For background correction, the normexp algorithm was applied (315).

The distribution of logarithmic signal intensities formed two clear peaks in all hybridizations, and a method conforming the positions of these density peaks was used to normalize the hybridization data. This approach assumed that all hybridizations exhibit high densities of both positive and negative hybridization signals from the pangenome array, and each sample was set with a threshold between the intensity peaks to classify probes as present, absent, or diverged in each strain. Probe signals ±0.3 from strain-specific threshold were classified as diverged (0 to 2 probes per hybridized strain) and these probes were considered absent in further data analysis. Visualization and clustering of data were conducted using MEV (316). The hybridization data were deposited in NCBI's Gene Expression Omnibus following the MIAME (Minimum Information About a Microarray Experiment) guidelines.

4.6 Transcriptome analysis of *Y. pseudotuberculosis* in cold temperature (IV)

4.6.1 Transcriptome preparation for RNASeq (IV)

Y. pseudotuberculosis IP32953 cells were grown at 3°C and 28°C, and samples for total RNA extraction were collected at six separate time points corresponding to different phases of growth across both temperatures sequenced. The extracted (GeneJET RNA Purification Kit, Thermo Fisher Scientific, Waltham, MA, USA) and cleaned (DNA-Free DNA Removal Kit, Ambion, Life Technologies, Carldbad, CA, USA; and Ribo-Zero rRNA Removal Kit for Bacteria, Epicenter, Madison, WI, USA) RNA was used to create cDNA libraries (Script-Seq v2 RNA-Seq Library Preparation Kit, Epicenter). PCR was used to amplify the libraries and a commercial kit was used to purify them. The libraries were sequenced (Illumina NextSeq500, Institute of Biotechnology, University of Helsinki) yielding four sets of biological replicates.

4.6.2 Data analysis (IV)

The reads were aligned and annotated with Bioconductor (317,318) against the reference genome (319). Bioconductor package BaySeq was used for transcript count analysis (320). For differential gene expression analysis, normalization and clustering of differentially expressed genes, packages such as ProOpDB and clusterProfiler were applied. Replicates were averaged, and averages were used instead of individual counts. The averaged transcript counts for each growth point at 3 °C were compared to the counts of the corresponding growth point at 28 °C. The $\log_2$ fold changes (logFC) between the two temperatures were calculated, and genes with logFC $\geq$ 2 were considered significantly expressed at 3 °C. RT-qPCR for selected genes (*betB, csdA, fabF, infA, mdtI, proX, rhlE*, and *rho*) was performed to validate results at both temperatures. The reads were deposited in the Sequence Read Archive following the MINSEQE (Minimum Information About a Next-generation Sequencing Experiment) guidelines.

5 RESULTS & DISCUSSION

5.1 Pangenomic approach to *C. botulinum* (I)

5.1.1 Probing the core genome and pangenome of *C. botulinum* (I)

Publicly available *C. botulinum* genomes (n = 16) were subjected to comparative genome analysis and their pangenome was constructed to investigate the population structure of *C. botulinum* and the genomic variety within it. The core genome contained 1076 gene clusters, while the core genomes of Group I (n = 10; 2758 clusters) and Group II (n = 4; 2456 clusters) contained more than twice the number of gene clusters compared to the core genome shared with all (Table 5). To numerically estimate the available genetic variety within *C. botulinum* the first estimate of its pangenome was constructed. Our analysis identified 18385 different genes suggesting a large, open pangenome (Table 5). Within this data set, each genome contained approximately 20% of the pangenome genes.

Table 5 Summary of pangenome studies for *Clostridium botulinum.*

Group	Genomes	Method		Size of the core genome		Size of the pangenome	Reference
				100%	99%		
C botulinum Groups I-III	16	BBH[a]		1076	N/A	18385	I
C. botulinum Group I	10	BBH[a]		2758	2758	N/A	I
C. botulinum Group II	4	BBH[a]		2456	N/A	N/A	I
C. botulinum Group I & *C. sporogenes*	556	Roary, identity	80%	N/A	2420[b]	18731	(132)
C. botulinum Group II	208	Roary, identity	80%	N/A	1768	16571	(139)

N/A, not analyzed. [a] BBH: Bi-directional best hit. [b] Present in 95% of genomes.

The Roary pipeline produced a pangenome profile with three separate gene cluster profiles, without a shared genetic backbone, corresponding to the *C. botulinum* Groups I, II, and III (Figure 3). In the gene cluster frequency graph for *C. botulinum,* the shared core genome is dwarfed by the group-specific genes, as is to be expected for a heterogeneous species (Figure 3).

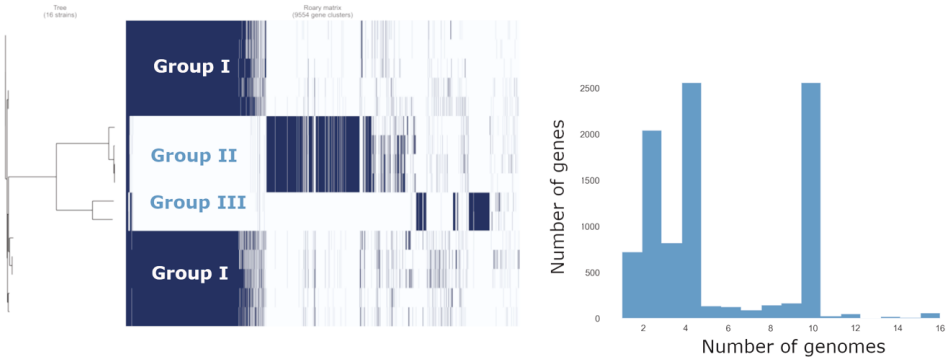

Figure 3. The pangenome profile (left) and gene cluster frequency graph (right) for *C. botulinum.*

The average similarity between compared *C. botulinum* genomes was under 70% and made this comparison markedly different from those of *C. perfringens* (II) and enteropathogenic *Yersinia* (III). The small core genome, distinct profiles of different groups of strains, and the large open pangenome are compliant with the notion that *C. botulinum* Groups I to III belong to different clostridial species and share only a distant genetic context between them, and just happen to produce the same toxin (51,121,142). Specifically, the estimated core genome was small and resembled the estimated core genome of clostridia in general (321).

For our analysis of the total core genome, the 16 genomes available at the time were used. Two of the genomes represented Group III, and Group IV genomes were not included. The true core genome of *C. botulinum* isolates is therefore likely considerably smaller than 1076 genes and likely very close to the size of the core genome of clostridia. Similarly, the pangenome of just 16 *C. botulinum* strains gives a scale for the size of the pangenome but was not sufficient to represent the entire available accessory genome to any level.

The core genome within clostridia has not been further explored, but since the study, I, Group I, and II pangenomes and core genomes have been estimated based on more than 200 strains each, and the results are summarized in Table 5.

5.1.2 Comparison of genomic differences (I)

To assess the differences between *C. botulinum* genomes an *in silico* proteome comparison was conducted. When the core genome within the Group I genomes was studied, the least conserved gene classes were mobile and extrachromosomal elements, toxin production, and resistance. Within the Group II genomes, the least conserved were genes related to mobile and extrachromosomal elements, pathogenesis, toxin production, and resistance. Notably, only 8 of the 18 genes related to pathogenesis in Group II E3 Alaska were conserved within the three other Group II genomes.

The analysis of protein-coding genes revealed that, while variation existed between isolates within Group I and Group II, no strain differed greatly from the others within each group. This suggests that phylogenetic groups are consistent and share a common ancestry.

5.1.3 Psychrotrophic Group II E strains lack csp homologs (I)

Group II strains are known to be psychrotrophic, but the genomic comparison revealed that while Group I and III strains contained 2-3 homologs for *csp* genes, the Group II genomes contained 0-1 homologs. All type E *C. botulinum* genomes (n = 3) carried no homologs for *csp* genes. Out of four Group II genomes analyzed, only II B Eklund contained a single *csp* homolog (CLL_A1515). This CDS shared moderate homology with Group I and Group III *csp* alleles (identity 49 to 51% with *cspA*, 41 to 49% with *cspB,* and 59 to 61% with *cspC*), but was even more closely related with *cspL* genes in *C. perfringens* (identity 84%).

Currently, at least 87 classic type E toxin genomes have been published, and none of them harbor *csp* homologs (Patric database, 24.1.2022). Contrarily, an untypical environmental *C. botulinum* strain (CDC66177) producing a variant of type E toxin (322), and some phylogenetically related strains, contained one *csp* homolog similar to the single cold shock protein-encoding CDS in II B Eklund.

The surrounding genetic context in Group II E strains was compared with that of II B Eklund, and the results suggested that the *csp* gene had been deleted from type E strains. The absence of *csp* homologs in the psychrotrophic Group II strains suggests that cold resistance and adaptation in these strains are not dependent on the role of *csp* genes.

The cold growth characteristics of various Group II *C. botulinum* strains have been studied and some strains seem to be more mesophilic than others (145). Derman et al. (145) tested minimum growth temperatures for Group II *C. botulinum* strains, and out of those 24 strains five have been sequenced (B Eklund, CB1171-1, Beluga, 202/ ATCC 23387, E strain K3). Two of these strains, B Eklund and

202/ ATCC 23387, carry a single *csp* homolog and three (Beluga, E strain K3 and CB1171-1) none. The presence or absence of *csp* homologs does not correlate with observed differences in minimum growth temperature.

Csps are known to some extent to compensate for each other's functions (323), meaning the impact of one lost homolog is likely less than the loss of all homologs (279). It is also known that not all *csp* genes are essential for cold tolerance but play other roles such as general stress resistance, host pathogenicity, and cell motility (278,279,323,324).

A similar loss of *csp* genes was also observed in c-cpe strains of *C. perfringens* (II). Also, these c-cpe *C. perfringens* strains are known to survive refrigeration and freezing significantly better than the cells and spores of other *cpe*-carrying isolates (204). Based on this, it is possible to hypothesize that loss of *csp*s is not detrimental to the cold resistance of clostridia. Further studies would be required to understand the putative mechanisms behind this.

5.2 Pangenomic approach to enteropathogenic *C. perfringens* (II)

5.2.1 Population structure

The population structure and genomics of enteropathogenic *C. perfringens* were studied to understand the population structure and elucidate the dispersal of the *cpe* gene. The *C. perfringens* strains analyzed belonged to five distinct lineages (I-V) (II, Figure 1). The core genome of 290 strains contained 1034, or 696 gene clusters depending on the pangenome pipeline used (Table 6).

Three gene cluster profiles corresponding to lineages III, IV, and V can be distinguished from the pangenome matrix (Figure 4). The most abundant gene cluster within *C. perfringens* isolates was the unique genes present in one or two strains, but within the *cpe*-carrying isolates, the shared core genome was the most abundant gene cluster highlighting the shared genetic backbone even between isolates representing different genetic lineages.

Notably, the reductive evolution within lineage IV is not visible on the gene cluster frequency graph. The gene frequency graph shows how many gene clusters of a certain size are shared by the isolates and does not reflect the genes lost. Reductive evolution is, however, visible in the strain profiles in pangenome matrix (Figure 4).

Table 6 Summary of pangenome results for *C. perfringens*.

Genomes	Method	Size of the core genome		Size of the pangenome
		100%	99%	
290	Roary, 95% identity	231	590	23148
290	Panaroo, strict	1034	1660	14306
283[a]	Roary, 95% identity	696	1170	16875
82, *cpe*-carrying isolates	Roary, 95% identity	1577	N/A	7835

N/A, not analyzed. [a] Confounding strains BER-NE33, K473, T3381, W1319, PC5, PBD1, and PBS5 were excluded from the analysis of 283 strains. [b] Present in 95% of genomes.

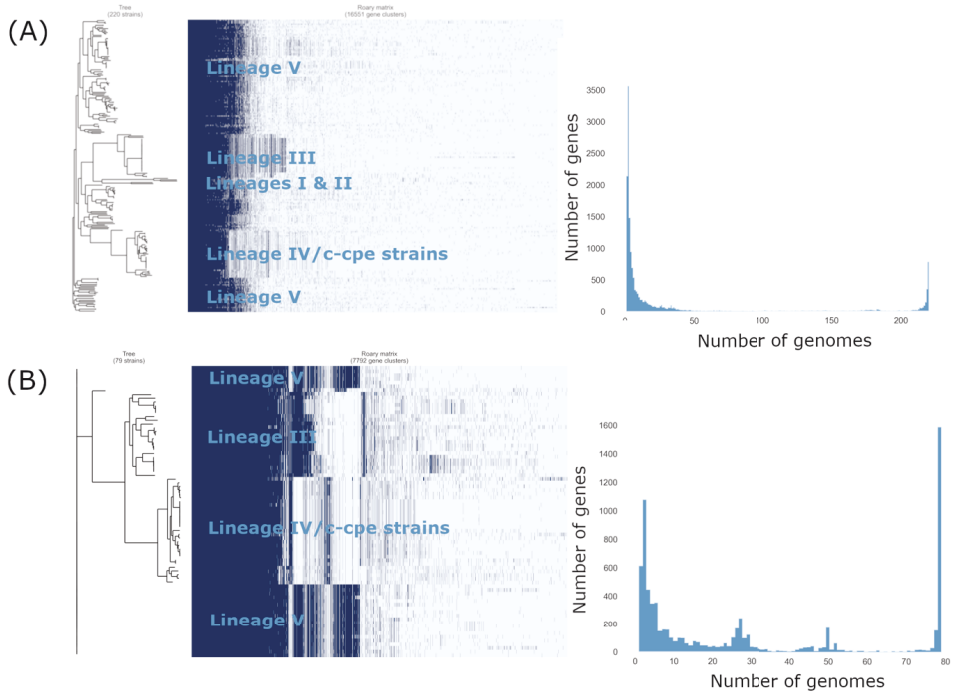

Figure 4 The pangenome matrix (left) and gene cluster frequency graph (right) for all *C. perfringens* strains (A) and *cpe*-carrying isolates of *C. perfringens* (B).

5.2.2 Dispersal of *cpe* gene within *C. perfringens* population (II)

Enterotoxin encoding *cpe*-gene in *C. perfringens* was carried out on a family of conjugative plasmids or in an integrated transposon within the chromosome. The understanding is that all *C. perfringens* types can carry the *cpe* gene and produce CPE, but that only approximately 5% of the strains do so (36). Three of the five lineages included *cpe*-carrying strains. All chromosomally *cpe*-carrying strains belonged to lineage IV, and plasmidially *cpe*-carrying strains were present in lineages III and V.

Comparative genome analysis revealed that two of the *cpe*-carrying toxin plasmids (pCPF4969, pCPBB-1) seemed constrained in their dispersal within the population (II, Figure 1) and that the core genome of *cpe*-carrying isolates from three lineages was, depending on the pangenome pipeline and methodology used, either significantly or moderately larger than that of all strains (Table 6). This suggests that conjugative *cpe*-carrying plasmids are transmitted within genetic lineages or reservoirs with yet uncharacterized limitations. This observation could be explained by a biased transfer rate of conjugative plasmids favoring the plasmid transfer to the kin similar to the observed pattern within the natural population of *E. coli* (325), or by certain genetic context (including for example the drug resistance genes and other genes fixed within the core genome of cpe-carrying isolates) that favor the *cpe*-carriage or allowing strains to flourish in the reservoirs where toxin plasmids are shared. Further studies on *cpe*-carrying plasmids and their incompatibilities in *C. perfringens* are warranted to understand this observation and its impact on the population structure of enteropathogenic *C. perfringens*.

5.2.3 The origin and reservoir of c-cpe strains (II)

Chromosomal *cpe*-carrying (c-cpe) strains are associated with food poisoning and have been, with some exceptions, generally isolated from food poisoning cases or food items (199,200,326). The strains have not been isolated from other environmental sources, and their reservoir from which they enter the food chain remains unknown. Our results revealed that the genetic lineage of c-cpe strains included i) 32 *cpe*-negative isolates (51%) from a variety of sources and geographic locations, and ii) 26 *cpe*-negative strains isolated from healthy swine and chicken (China, Finland). This is not directly indicative of the reservoir of c-cpe strains but suggests that the adapted niche for this lineage of strains or their ancestor might be in swine and poultry. Comparative genome analysis also indicated that the lineage IV might share ancestry with lineage V strains carrying IS1470-like plasmids. Additionally, our analysis revealed three groups of isolates within lineage IV: i) c-cpe Group 1, ii) c-cpe Group 2, and iii) *cpe*-negative isolates from swine and chicken. In conclusion, the results provided insights on

the population structure within lineage IV and revealed the putative origin of these epidemiologically relevant strains to be in domestic animals, suggesting that their entry to the food chain might occur at the farm level.

5.2.4 Niche adaptation within c-cpe strains (II)

Analysis of lineage IV strains in *C. perfringens* revealed distinct gene profiles within the c-cpe strains that corresponded to the phylogenetic clusters of c-cpe Group 1 and Group 2. Comparative genome analysis was conducted to characterize these gene profiles in detail, and evidence of niche adaptation and reductive evolution into two distinct directions was observed.

The c-cpe Group 1 strains had a reduced virulence gene profile, while the c-cpe Group 2 strains retained more virulence genes encoding sialidases and hyaluronidases and carried a fucose operon. The c-cpe Group 2 strains on the other hand had lost operons related to survival and resistance (arginine deiminase pathway operon, arsenic resistance operon, iron acquisition systems) in addition to an operon responsible for citrate metabolism. These results suggest that these groups of c-cpe strains have taken their reductive evolution in different directions and adapted to different ecological niches. We hypothesize that the c-cpe Group 1 strain has adapted to an acidic niche where exposure to stressful conditions in terms of pH and temperature may be common. The c-cpe Group 2 strains are likely more adapted to constant environmental conditions and their retainment of colonization-related virulence factors and fucose operon suggest they might prosper in some type of GI-tract niche.

5.2.5 Genetics of heat resistance in enteropathogenic *C. perfringens* (II)

In addition to the genetic differences between c-cpe Groups 1 and 2, we also discovered that c-cpe strains in Group 2 produced heat-sensitive spores. Previously all c-cpe strains have been presumed to produce heat-resistant spores (203,295) and to carry the heat-resistance allele of Ssp4 associated with this phenotype. Our results showed that the known heat-resistance allele of Ssp4 was confined to the c-cpe Group 1 and likely originated within this phylogroup.

One outlier c-cpe strain (310/85) produced heat-resistant spores despite carrying the allele of Ssp4 associated with the production of heat-sensitive spores, and despite belonging to c-cpe Group 2 strains producing heat-sensitive spores. Our results also suggest that spore heat resistance developed independently in a strain (310/85) without the known heat-resistance allele of Ssp4.

The genetic context behind the noted heat-resistance patterns was compared, and one allele of heat shock protein GrpE and Co-A-reductase were identified as putative genes that play a role in the heat-resistant phenotype of spores. Further studies with a wider selection of strains are required to confirm whether these alleles play a role in spore heat resistance in *C. perfringens* or not.

In conclusion, our results revealed that the spore heat resistance has developed mainly within one group of c-cpe strains and is likely affected by additional genetic characteristics in addition to the already established Ssp4 heat-resistance allele.

5.3 Pangenomic approach to enteropathogenic *Yersinia* (III)

5.3.1 Population structure of *Y. enterocolitica* and *Y. pseudotuberculosis* (III)

Hybridization results clustered *Y. enterocolitica* (n = 61) and *Y. pseudotuberculosis* (n = 38) strains into two distinct groups according to their species. Notably, one strain, ÅYV 7.1K2, was a clear outlier within *Y. pseudotuberculosis* strains and was reclassified as *Y. pekkanenii* (327). The distance between *Y. enterocolitica* and *Y. pseudotuberculosis,* based on Pearson's correlation on a scale from 0 to 2, was 1.36.

Y. enterocolitica strains belonging to different biotypes (1A, 1B, 2, and 4) formed distinct subclusters but *Y. enterocolitica* biotypes 2 or 3 (n = 10) clustered together. Within the *Y. enterocolitica* group, the genetic distance was the longest (0.25) between *Y. enterocolitica* biotypes 2-4 and biotypes 1A and 1B.

Two subclusters were observed within hybridized *Y. pseudotuberculosis* strains. The majority of hybridized *Y. pseudotuberculosis* strains had been obtained from swine samples in Finland, Estonia, Russia, England, and Belgium (n = 23), and they clustered together forming a "swine group". In contrast, 13 strains clustered separately from the "swine group" and formed a separate "diverse group". The genetic distance between these two groups was 0.15.

While the majority of *Y. pseudotuberculosis* strains isolated from swine were very homogenous and belonged to the "swine group", 5 of the 11 swine strains from English fattening pigs belonged to the "diverse group", together with strains isolated from wildlife and humans. This diverse group also included the type strain IP32953 of *Y. pseudotuberculosis*.

5.3.2 Probing pangenome and core genome in enteropathogenic *Yersinia* (III)

For the seven genomes used in array design, the core genome based on bi-directional best hits (BBH) contained 2772 sequences, implying that 68–76% of the genes of each genome were shared. Equally the analysis with the Roary pipeline produced a pangenome matrix showing the two gene cluster profiles with a shared genetic backbone (Figure 5). The enteropathogenic *Yersinia* share a substantial amount of core genome (39.7-67.7%) between them, yet the "volatile genes" specific to each species were abundant (Figure 5, Table 7).

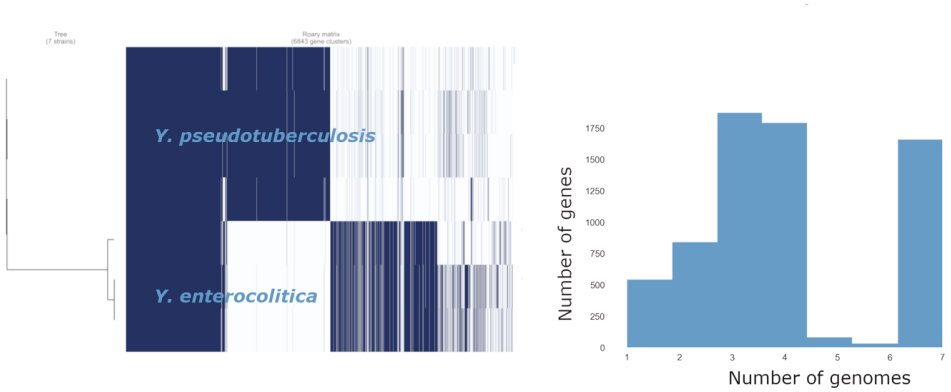

Figure 5. The pangenome matrix (left) and gene cluster frequency graph (right) for enteropathogenic *Yersinia.*

All hybridized *Yersinia* strains produced positive signals on 459 probes, which is the equivalent of 320–360 genes depending on the reference genome used (Table 7). This means that around 8% of the genome is fully conserved across the two species. The difference between the core genome estimation and the number of shared genes in WGH is explained by the differences in the stringency of applied methods: the BBH method can identify evolutionarily distant orthologs while the WGH discriminates between alleles on the SNP level.

5.3.3 Differences between species (III)

Y. enterocolitica strains shared notably fewer specific and conserved genes (n = 448) between them than *Y. pseudotuberculosis* strains (n = 906). This reflects the greater heterogeneity within *Y. enterocolitica* biotypes and also the ecological adaptation of biotypes 2-5 through reductive evolution and gene loss (222).

Table 7 Summary of pangenome studies and their results for enteropathogenic *Yersinia.*

Group	Genomes	Method	Size of the core genome		Size of the pangenome	Reference
			100%	99%		
Enteropathogenic *Yersinia*	7	BBH[a]	2772	N/A	N/A	III
Enteropathogenic *Yersinia*	99	WGH[b]	459	N/A	N/A	III
Y. enterocolitica	117	Bio-Pangenome		3181	11460	(66)

N/A, not analyzed. [a] BBH: Bi-directional best hit. [b] Whole genome hybridization.

The genetic differences between *Y. enterocolitica* and *Y. pseudotuberculosis* were also explored. In both species, many of the species-specific core genes were involved in the transportation of substances.

Y. pseudotuberculosis strains shared a variety of gene clusters involved in the use and/or uptake of various substrates, including phenolic compounds, rhamnose, xylose, myo-inositol, opines/polyamines, and aliphatic sulfonates. In total *Y. pseudotuberculosis* strains shared 18 ABC-transporters and 2 PTS transporters that were absent from *Y. enterocolitica* strains. *Y. enterocolitica* strains shared six ABC transporters and seven PTS that were all absent from *Y. pseudotuberculosis* strains. Putative substrates for these transportation systems included iron, glucoside, metallic ions, lactose/cellobiose, sorbitol, and sucrose.

These results highlight the differences between these two species in substrate utilization and ecological niches adapted to. Transportation variety within *Y. enterocolitica* seemed to be more limited than within *Y. pseudotuberculosis,* which also carried operons linked with the uptake and utilization of substances not found in living animal tissues but present in the soil, plants, and rotting flesh.

Also, several operons related with metabolism and biosynthesis differed between species. *Y. enterocolitica* had the genetic capability to utilize citrate and propanediol (including the related biosynthesis of cobalamin), and synthesize cellulose. These genetic traits are likely beneficial in gut colonization and survival during extracellular exposure. The *Y. enterocolitica* strains, excluding the

highly virulent 1B strains, also carried an operon involved in the utilization of N-acetylgalactosamine. This is a metabolite of fucose and is readily available in the gut environment.

One remarkable characteristic for *Y. pseudotuberculosis* was the presence of three type VI secretion systems (T6SS). These were not found in *Y. enterocolitica*. One operon of T6SS identified in *Y. pseudotuberculosis* genomes (CAH22876.1 in IP32953) shared strong similarity with the T6SSs of *Vibrio cholerae* and *Burkholderia thailandensis*, which are both considered to have cytotoxic effects against unicellular organisms, but also against mammal macrophages, and are therefore, putative virulence factors. We hypothesize that these defense and interaction systems help *Y. pseudotuberculosis* to survive and multiply in ecological niches in the environment from which it could easily end up as a contaminant of the food chain.

All in all, the results suggested that *Y. pseudotuberculosis* possesses many tools that are beneficial for survival in varied environments, while the *Y. enterocolitica* genome is more streamlined, tending towards host-associated lifestyles within their preferred animal reservoir.

5.3.4 Swine specificity (III)

The main animal reservoir for *Y. enterocolitica* 4/O:3 strains are pigs, which function as asymptomatic carriers. Genomic adaptations to this niche were also studied. Bioserotype 4/O:3 strains (n = 42) shared 51 gene clusters (1% of sequenced *Y. enterocolitica* 4/O:3 genome Y11) that were only present in strains of this bioserotype. These included mainly serotype O:3 antigen-related genes.

Biotype 2-4, but not biotype IB were observed to carry the N-acetylgalactosamine utilization operon associated with adaptation to the GI tract.

Subsequently and simultaneously, extensive research has been carried out to uncover the virulence factors of *Y. enterocolitica* and its different serotypes and the virulence factors explaining the swine specificity of *Y. enterocolitica* serotype O:3 (328–330). It has been proposed that the N- acetylgalactosamine utilization operon plays a role in host adaptation since it enables the use of intestinal mucin as a carbon source, and the gut of pigs is rich in N-acetylgalactosamine-containing mucin (331). However, the main adaptation mechanism has been attributed to changes in gene expression. Small but significant variations in cell adhesion and invasion properties via changed expression responses to temperature have been noted to result in high expression of the primary invasion factor at 37 ◦C, providing attunement to the higher body temperature of the natural host reservoir, the pig (330,332). The actual changes in transcription are attributed to two changes within

genes: i) insertion of an IS1667 to primary invasion factor promoter, and ii) a base-pair substitution in *rovA*, resulting in a more stable variant of the virulence regulator RovA (333).

We also compared the genetic content of the swine group and diverse group of *Y. pseudotuberculosis* but did not distinguish gene clusters to discriminate between these two. The subclusters observed were differentiated by genetic distance rather than gene or operon level differences. Niskanen et al. (220) have previously reported based on PFGE analysis that *Y. pseudotuberculosis* strains isolated from swine samples are homogenous. The genetic diversity within the same English fattening pigs was first reported by Ortiz Martinez et al. (243,244,334). To this date, the *Y. pseudotuberculosis* is considered a homogenous group, and clear, distinct subclusters within this pathogen have not been identified (207,221,222,261). One notable limitation of our study was the dependency of WGH on the representation of reference strains used in microarray design. The array contained only probes designed based on "diverse group" *Y. pseudotuberculosis* genomes, and repeating the comparative genome analysis with fully sequenced "swine group" genomes could perhaps elucidate the genetic differences between different *Y. pseudotuberculosis* strains.

5.4 Cold growth of *Yersinia pseudotuberculosis* (IV)

5.4.1 Differentially expressed genes during cold growth form clusters with different expression patterns (IV)

The RNA expression profile of *Y. pseudotuberculosis* IP32953 grown at 3°C was compared to that of the cells grown at 28°C to determine which genes were important for growth at cold temperatures. In total, 570 genes were expressed significantly more at 3°C than 28°C at least at one of the growth points measured (IV, Figure 1). Since genes that are expressed both in great numbers and differentially at 3 °C are likely to play important roles in cold growth, this result suggested that *Y. pseudotuberculosis* IP32953 has an extensive toolbox of hundreds of genes to facilitate cold growth.

The differentially expressed genes were clustered based on their expression profiles at different time points of the growth curve (IV, Figure 2). *Y. pseudotuberculosis* is known to be non-motile at 28°C, and unsurprisingly 88 of the 570 differentially expressed genes were related to motility and chemotaxis. The 482 genes not related to chemotaxis or motility were analyzed further, and five subclusters (A-E) of expression patterns were identified. The most interesting clusters were clusters

containing genes expressed highly at the beginning of growth (E, D) and logarithmic phase of growth (B, C) (IV, Figure 2). Cluster A held mainly genes that were expressed highly at the stationary phase.

5.4.2 Differentially expressed gene and their functions (IV)

The genes that were highly expressed at the very beginning of growth (E, growth point I) played roles in acquiring compatible solutes and various nutrients. Significant increases were observed in genes encoding a glycine betaine transporter, phosphotransferase systems (PTS) to import fructose, N- acetylglucosamine, and L-ascorbate, and ATP-binding cassette (ABC) transporters of maltose and aldopentoses. Some of these operons have also been identified as differing between *Y. pseudotuberculosis* and *Y. enterocolitica* (III). Also, genes encoding chaperone molecules such as DEAD-Box RNA helicase RhlE and Csps (yptb2950, yptb2414, yptb3585, and yptb3586), which destabilize nucleic acid secondary structures, were differentially and highly expressed at 3°C at the beginning of cold growth.

Subcluster D included genes with roles in substrate transport (glutamate, inositol, and arginine) in addition to betaine biosynthesis. This is likely due to cells trying to secure the availability of sufficient nutrients to ensure successful growth at low temperatures. PTSs in particular have been linked to regulatory functions during cold stress, and enable quick acquisition of carbohydrates from the environment (335,336). Additionally, genes related to spermidine efflux, synthesizing desaturated membrane lipids, and securing translation (biosynthesis of ribosomes, posttranscriptional modification of RNA, translation factors IF-1 and Rho, DEAD-Box chaperone *dbpA,* and *csp* yptb1423) were highly expressed at this point.

During logarithmic growth (growth points III-V) (IV, Figure 1) the nutrient strategy seemed to shift, as urease operon and PTSs to import fructose (IV, Figure 5), N-acetylglucosamine, and mannose first showed significantly more transcripts. Further on, at growth point IV, the sulfur metabolism operon was highly expressed. In addition to IF-1 and Rho, the translation factor *rbfA* was also highly expressed.

Entering the stationary phase, at growth points V and VI, the nutrient strategy continued shifting as operons involved in the metabolism of amino acids and other nitrogen compounds such as histidine, cystine, and methionine showed more transcripts at 3◦C (IV, Figure 5), and the *csp* gene expression levels tapered off.

There are five DEAD-box RNA helicases in *Y. pseudotuberculosis* and they all were expressed more at 3◦C throughout the growth (IV, Figure 8). The difference in expression was statistically significant in growth point III for *csdA*, growth points I-IV for *rhlE*, and growth points II and IV for *dbpA*.

In conclusion, we characterized the extensive toolbox that *Y. pseudotuberculosis* used to keep its protein synthesis running at low temperatures. Cold shock proteins encoded by yptb1423, yptb2414, yptb2950, yptb3585–86, and RNA helicases CsdA, RhlE, and DbpA, seemed to form the backbone of cold survival of *Y. pseudotuberculosis.* In addition to these, the network protecting *Y. pseudotuberculosis* from cold damage included transcription factors IF-1, RbfA, and also Rho, which seemed to support protein synthesis at cold temperatures.

6 CONCLUSIONS

The abundance and availability of genomic data make it an appealing new starting point for the study of ecology and the evolution of bacteria. This thesis aimed to assess the suitability and prospects of comparative genome analysis and a pangenomic approach for a better understanding of the ecology, evolution, and through these also the epidemiology of foodborne pathogens.

The pangenome of *C. botulinum* reflected the heterogeneous nature of this pathogen. The isolates shared only a small core genome comparable to that of the clostridial backbone and the pangenome was large. Psychrotrophic *C. botulinum* strains in Group II lacked cold shock protein genes conserved in Group I and III strains, suggesting that cold shock protein homologs are not necessary for cold adaptation in *C. botulinum* Group II.

The comparative genome analysis of *cpe*-carrying *C. perfringens* strains revealed a novel group of c-cpe strains associated within food poisoning. Comparative genome analysis also identified a putative reservoir or origin for c-cpe strains within swine and poultry. Exploration of genetic diversity within c-cpe strains suggested that strains with different gene profiles had adapted to different ecological niches and reservoirs. The gene profiles also corresponded with phenotypic differences in spore heat-resistance; unlike the previously described c-cpe strains, the novel gene profile of c-cpe strains produced heat-sensitive spores.

The comparison of enteropathogenic *Yersinia* isolates revealed a large core genome shared between enteropathogenic *Yersinia*. Results revealed that all *Y. pseudotuberculosis* isolates shared several genetic traits useful for survival in various environments that were absent from *Y. enterocolitica*. Most notably the *Y. pseudotuberculosis* strains harbored a selection of type VI secretion systems targeting competitive cells of other microbes and eukaryotes. The genomes of *Y. enterocolitica* were more streamlined and the biotypes had undergone reductive evolution during adaptation to their niches. The whole-genome hybridization was also able to distinguish two clusters ("swine group" and "diverse group") within *Y. pseudotuberculosis,* but the genetic differences between these clusters warrant further study.

The cold shock response and cold adaptation of *Yersinia pseudotuberculosis* serotype O1 were studied on the transcriptome level and a genome-wide transcription adaptation to cold growth was observed. Strain IP32953 engaged a large number of genes at different stages of the growth curve during the cold response. The backbone of cold survival for *Y. pseudotuberculosis* seemed to be formed by cold shock proteins and RNA helicases CsdA, RhlE, and DbpA. In addition to these the

transcription factors IF-1, RbfA, and Rho were highly expressed during cold growth and seemed to support protein synthesis at suboptimal temperatures.

A pangenomic approach successfully elucidated adaptive evolution within virulence and stress response mechanisms and allowed inference of evolutionary relationships between foodborne pathogens. The pangenome profiles can be used to identify main genetic lineages within less-studied organisms or to identify clear outliers (biological or technical) within the analyzed genomes. Compared to simple phylogenetic analysis, the pangenome profile enriched with gene cluster frequencies identifies clusters of isolates with accumulated and reduced genetic content. However, good quality pangenome analysis with prospects of elucidating reservoir and ecology requires a wealth of data, and is ideally combined with relevant phenotypic data. The comparative genome analyses focus on the direct examination of DNA sequence, the predicted genes, and their functions – these are also the limitations for the prospects of this method. Gene expression studies as transcriptome analysis are important to confirm when and where the genes of interest are turned on or off. Importantly, when studying and comparing environmental bacteria and their genomes, closely related environmental strains and strains not carrying toxin genes should also be included. Sequencing efforts of environmental and commensal isolates will probably aid in understanding the emergence and adapted niches of pathogenic strains.

These studies shed light on the genes that contribute to stress tolerance (I, II, IV), niche adaptation (II, III), and ecology (I-IV) in foodborne pathogens. The pangenomic studies also elucidated genomic evolution within foodborne pathogen populations and were efficient in providing scalable, high-resolution views on population structure. Knowledge of genes associated with stress tolerance, reservoirs, and lifestyles is beneficial in the development of new targeted strategies and measures to identify and control the food safety risks caused by these bacteria.

ACKNOWLEDGEMENTS

This thesis study was conducted at the Department of Food Hygiene and Environmental Health of the Faculty of Veterinary Medicine, University of Helsinki. The Finnish Foundation of Veterinary Research is gratefully acknowledged for their financial support.

Professor emeritus Hannu Korkeala, Professor Miia Lindström, and Docent Riikka Keto-Timonen are warmly acknowledged for their considerate and insightful supervision of this thesis. Each of you contributed in different ways to this work, and I feel obliged for the support and encouragement I've received during the years. My thesis committee members Docent Leena Maunula and Postdoctoral Researcher Ravi Kant are warmly thanked for their support and guidance.

Professor Claudia Guldimann and Professor Thomas Alter are highly appreciated for reviewing this thesis in such an expert and supportive manner.

I wish to equally acknowledge each person in the department for many advice and answers readily given. I am indebted to all of my co-authors of which Panu Somervuo deserves an honorary mention. I still vividly remember how you showed me the way to create a simple script to compare proteins within genomes and to identify orthologs. Little did I know how many times I'd return to those scripts in coming years! I also thank Katja Selby, Yagmur Derman, and Kirsi Ristkari for their indispensable help during this thesis.

Loput kiitoksista haluan antaa suomeksi. Katja, olen melkoisen varma, että tämä väitöskirja ei olisi valmistunut ilman sinua. Kiitos neuvoista ja terveestä vertaispaineesta! Kiitän sekä Katjaa, että siskoani Tarua myös siitä, että tietämättänne (vai tietoisesti) olette toistuvasti raahanneet minut pois mukavuusalueeltani. Olisin varmasti rapakunnossa ilman teitä. Siskoani Lauraa ja ystävääni Annaa kiitän hyvistä keskusteluista ja neuvoista tähänkin väitöskirjaan liittyen. Kiitos myös kaikille mainituille ja nimeltä mainitsemattomille sukulaisille ja ystäville tuesta ja ystävyydestä.

Suuri kiitos kuuluu vanhemmilleni hyvistä elämän eväistä, jotka kotoa Vihteljärveltä sain. Omistan tämän työn etenkin äidilleni Lealle. On täysin esimerkkisi ansiota, että alkujaan lähdin tälle uralle ja olen todella kiitollinen siitä vankkumattomasta tuesta ja uskosta, jota olet kaikille hankkeilleni aina antanut.

Tämä väitöskirja valmistui koronavuosina ja iso osa työstä tehtiin lockdownin ja erilaisten rajoitusten aikana. Lopuksi olenkin luvannut kiittää bcjodelia ja Heikin Statsikerhoa laadukkaasta prokrastinoinnista ja monista hyvistä nauruista kirjoitusprosessin aikana. It wasn't smart but it was fun.

REFERENCES

1. Tauxe R. Emerging foodborne pathogens. Int J Food Microbiol. 2002;78(1–2):31–41.

2. WHO. WHO Estimates of the Global Burden of Foodborne Diseases, Executive Summary. 2015.

3. Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 1995;269(5223):496–512.

4. Koponen J, Hilden J. Data Visualization Handbook. Aalto University; 2019.

5. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. 2005;102(39):13950–5.

6. McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, Sreecharan T, et al. Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. Hughes D, editor. PLOS Genet. 2016;12(9):e1006280.

7. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. Nature Reviews Microbiology. 2015; 13(12): 787–94.

8. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical Microbiology and Infection. 2018;24(4): 335–41.

9. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.

10. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nat 2005 4377057. 2005;437(7057):376–80.

11. Shendure J, Ji H. Next-generation DNA sequencing. Nature Biotechnology. 2008; 26 (10):1135–45.

12. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science (80-). 2009, 2;323(5910):133–8.

13. Eisenstein M. Oxford Nanopore announcement sets sequencing sector abuzz. Nature Biotechnology. 2012. 30:295–6.

14. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. Vol. 277, Science. American Association for the Advancement of Science; 1997. 1453–62.

15. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 2001;8(1):11–22.

16. Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A. 2002;99(26):17020–4.

17. Liao YC, Lin SH, Lin HH. Completing bacterial genome assemblies: Strategy and performance comparisons. Sci Rep. 2015;5(1):1–8.

18. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020 211. 2020;21(1):1–16.

19. Bobay LM, Ochman H. The evolution of bacterial genome architecture. Frontiers in genetics, 2017, 8: 72.

20. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nat 2000 4066795. 2000;406(6795):477–83.

21. Hashimoto Y, Kita I, Suzuki M, Hirakawa H, Ohtaki H, Tomita H. First Report of the Local Spread of Vancomycin-Resistant Enterococci Ascribed to the Interspecies Transmission of a *vanA* Gene Cluster-Carrying Linear Plasmid. mSphere. 2020;5(2).

22. Kuo CH, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. Genome Res. 2009;19(8):1450–4.

23. Ochman H, Moran NA. Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. Vol. 292, Science. Science; 2001. 1096–8.

24. Azarian T, Huang I-T, Hanage WP. Structure and Dynamics of Bacterial Populations: Pangenome ecology. In: Tettelin H, Medini D, editors. The Pangenome. Springer Nature Switzerland AG; 2020. 115–28.

25. Thomas CM, Nielsen KM. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. Nat Rev Microbiol 2005 39. 2005;3(9):711–21.

26. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of anti-phage defense systems in the microbial pan-genome. Science. 2018;359(6379).

27. Golz JC, Stingl K. Natural Competence and Horizontal Gene Transfer in Campylobacter. In: Current Topics in Microbiology and Immunology. Springer Science and Business Media Deutschland GmbH; 2021. 265–92.

28. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol 2014 123. 2014;12(3):181–96.

29. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. PLOS Genet. 2009;5(1):e1000344.

30. Hanage WP. Not So Simple After All: Bacteria, Their Population Genetics, and Recombination. Cold Spring Harb Perspect Biol. 2016;8(7):a018069.

31. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. 2006;2(3):e31.

32. Blokesch M. Natural competence for transformation. Current Biology. Cell Press; 2016. 26: 1126–30.

33. Wren BW. The *Yersiniae*--a model genus to study the rapid evolution of bacterial pathogens. Nat Rev Microbiol. 2003;1(1):55–64.

34. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol. 2015; 0:242.

35. Skarin H, Segerman B. Plasmidome interchange between *Clostridium botulinum*, *Clostridium novyi* and *Clostridium haemolyticum* converts strains of independent lineages into distinctly different pathogens. PLoS One. 2014;9(9).

36. Miyamoto K, Fisher DJ, Li J, Sayeed S, Akimoto S, McClane BA. Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type A non-food-borne human gastrointestinal disease isolates. J Bacteriol. 2006;188(4):1585–98.

37. Skarin H, Håfström T, Westerberg J, Segerman B. *Clostridium botulinum* group III: a group with dual identity shaped by plasmids, phages and mobile elements. BMC Genomics 2011 121. 2011;12(1):1–13.

38. Rood JI, Adams V, Lacey J, Lyras D, McClane BA, Melville SB, et al. Expansion of the *Clostridium perfringens* toxin-based typing scheme. Anaerobe. 2018; 53:5–10.

39. Mount D. Bioinformatics: sequence and genome analysis. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 2004.

40. Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. Trends Genet. 2001;17(10):589–96.

41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

42. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.

43. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004; 32:115–9.

44. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 2008, 8;9.

45. Salzberg SL. Next-generation genome annotation: we still struggle to get it right. Genome Biol 2019 201. 2019;20(1):1–3.

46. Hyytiä-Trees EK, Cooper K, Ribot EM, Gerner-Smidt P. Recent developments and future prospects in subtyping of foodborne bacterial pathogens. Future Microbiology. 2007. 2(2):175–85.

47. Vos P, Hogers R, Bleeker M, Reijans M, Lee T Van De, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995;23:4407–14.

48. Kaufmann ME. Pulsed-Field Gel Electrophoresis. Mol Bacteriol. 1998;33–50.

49. Maiden MCJ, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 2013;11(10):728–36.

50. Stackebrandt E, Goebel BM. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. Int J Syst Evol Microbiol. 1994;44(4):846–9.

51. Hill KK, Smith TJ, Helma CH, Ticknor LO, Foley BT, Svensson RT, et al. Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. J Bacteriol. 2007;189(3):818–32.

52. Lindström M, Hinderink K, Somervuo P, Kiviniemi K, Nevas M, Chen Y, et al. Comparative genomic hybridization analysis of two predominant Nordic group I (proteolytic) *Clostridium botulinum* type B clusters. Appl Environ Microbiol. 2009;75(9):2643–51.

53. Stackebrandt E, Kramer I, Swiderski J, Hippe H. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. FEMS Immunol Med Microbiol. 1999;24(3):253–8.

54. Cebula TA, Jackson SA, Brown EW, Goswami B, Leclerc JE. Chips and SNPs, Bugs and Thugs: A Molecular Sleuthing Perspective. J Food Prot. 2005;68(6):1271–84.

55. Bobay LM. The Prokaryotic Species Concept and Challenges. In: Tettelin H, Medini D, editors. The Pangenome. 2020.21–50.

56. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin Microbiol. 2015;23:148–54.

57. Hanage WP, Fraser C, Spratt BG. Fuzzy species among recombinogenic bacteria. BMC Biol 2005 31. 2005;3(1):1–7.

58. Hanage WP. Fuzzy species revisited. BMC Biol 2013 111. 2013;11(1):1–3.

59. Shapiro B, Polz M. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol. 2014;22(5):235–47.

60. Smith JM. Trees, bundles or nets? Trends Ecol Evol. 1989;4(10):302–4.

61. Spratt BG, Maiden MCJ. Bacterial population genetics, evolution and epidemiology. Philos Trans R Soc London Ser B Biol Sci. 1999;354(1384):701–10.

62. Sung W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, Cooper VS, et al. Evolution of the Insertion-Deletion Mutation Rate Across the Tree of Life. G3 Genes|Genomes|Genetics. 2016;6(8):2583–91.

63. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 2009;3(2):199–208.

64. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001;409(6823):1007–11.

65. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat Genet 2008 408. 2008;40(8):987–93.

66. Reuter S, Corander J, Been M de, Harris S, Cheng L, Hall M, et al. Directional gene flow and ecological separation in *Yersinia enterocolitica*. Microb Genomics. 2015;1(3).

67. Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A. 2004;101(38):13826–31.

68. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al. Patterns of genome evolution that have accompanied host adaptation in Salmonella. Proc Natl Acad Sci. 2015;112(3):863–8.

69. Hauck S, Maiden MCJ. Clonally Evolving Pathogenic Bacteria. Gd Challenges Biol Biotechnol. 2018;307–25.

70. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct. 2009;4.

71. Touchon M, Rocha EPC. Causes of Insertion Sequences Abundance in Prokaryotic Genomes. Mol Biol Evol. 2007;24(4):969–81.

72. McCutcheon J, Moran N. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 2012; 10:13–26.

73. Manzano-Marín A, Latorre A. Settling Down: The Genome of *Serratia symbiotica* from the Aphid *Cinara tujafilina* Zooms in on the Process of Accommodation to a Cooperative Intracellular Life. Genome Biol Evol. 2014;6(7):1683–98.

74. Lerat E, Ochman H. Recognizing the pseudogenes in bacterial genomes. Nucleic Acids Res. 2005;33(10):3125–32.

75. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 2009;33(2):376–93.

76. Feng Y, Fan X, Zhu L, Yang X, Liu Y, Gao S, et al. Phylogenetic and genomic analysis reveals high genomic openness and genetic diversity of *Clostridium perfringens*. Microb Genomics. 2020;6(10): e000441.

77. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol. 2008;190(20):6881–93.

78. Jackson RW, Vinatzer B, Arnold DL, Dorus S, Murillo J. The influence of the accessory genome on bacterial pathogen evolution. Mobile genetic elements, 2011, 1.1: 55-65.

79. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol. 2008;11(5):472–7.

80. Edwards DJ, Holt KE. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb Inform Exp. 2013;3(1):1–9.

81. Li J, Adams V, Bannam T, Miyamoto K, Garcia J, Uzal F, et al. Toxin Plasmids of Clostridium perfringens. Microbiol Mol Biol Rev. 2013;77(2):208–33.

82. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol. 2008;153(S1):347–57.

83. Pöntinen A, Aalto-Araneda M, Lindström M, Korkeala H. Heat Resistance Mediated by pLM58 Plasmid-Borne ClpL in *Listeria monocytogenes*. mSphere. 2017;2(6).

84. Hernandez BG, Vinithakumari AA, Sponseller B, Tangudu C, Mooyottu S. Prevalence, Colonization, Epidemiology, and Public Health Significance of *Clostridioides difficile* in Companion Animals. Front Vet Sci. 2020;0:663.

85. Nuccio SP, Bäumler AJ. Comparative analysis of *Salmonella* genomes identifies a metabolic network for escalating growth in the inflamed gut. MBio. 2014;5(2).

86. Gal-Mor O, Boyle EC, Grassl GA. Same species, different diseases: how and why typhoidal and non-typhoidal *Salmonella enterica* serovars differ. Front Microbiol. 2014; 0:391.

87. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature. 2001;413(6855):523–7.

88. Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, et al. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc Natl Acad Sci U S A. 2004;101(51):17837–42.

89. Sun YC, Jarrett CO, Bosio CF, Hinnebusch BJ. Retracing the Evolutionary Path that Led to Flea-Borne Transmission of *Yersinia pestis*. Cell Host Microbe. 2014;15(5):578–86.

90. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, et al. Extraintestinal Virulence Is a Coincidental By-Product of Commensalism in B2 Phylogenetic Group *Escherichia coli* Strains. Mol Biol Evol. 2007;24(11):2373–84.

91. Nowrouzian FL, Wold AE, Adlerberth I. *Escherichia coli* Strains Belonging to Phylogenetic Group B2 Have Superior Capacity to Persist in the Intestinal Microflora of Infants. J Infect Dis. 2005;191(7):1078–83.

92. Feng P, Lampel KA, Karch H, Whittam TS. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J Infect Dis. 1998;177(6):1750–3.

93. Eichhorn I, Heidemanns K, Semmler T, Kinnemann B, Mellmann A, Harmsen D, et al. Highly virulent non-O157 enterohemorrhagic *Escherichia coli* (EHEC) serotypes reflect similar phylogenetic lineages, providing new insights into the evolution of EHEC. Appl Environ Microbiol. 2015;81(20):7041–7.

94. McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. Nat Microbiol 2017 24. 2017;2(4):1–5.

95. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. Curr Opin Genet Dev. 2005;15(6):589–94.

96. Medini D, Donati C, Rappuoli R, Tettelin H. The Pangenome: A Data-Driven Discovery in Biology. In: Tettelin H, Medini D, editors. The Pangenome. Springer Nature Switzerland AG; 2020. 3–20.

97. Fitch WM. Distinguishing homologous from analogous proteins. Syst Zool. 1970;19(2):99–113.

98. Lapierre P, Gogarten JP. Estimating the size of the bacterial pan-genome. Trends Genet. 2009;25:107–10.

99. Tatusov RL, Koonin E V, Lipman DJ. A genomic perspective on protein families. Science (80-). 1997;278(5338):631–7.

100. Rasko DA, Myers GSA, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics. 2005; 6:2.

101. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.

102. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21(1):180.

103. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3.

104. Guimarães LC, Florczak-Wyspianska J, Jesus LB de, Viana MVC, Silva A, Ramos RTJ, et al. Inside the Pan-genome - Methods and Software Overview. Curr Genomics. 2015;16(4):245.

105. Kiu R, Caim S, Alexander S, Pachori P, Hall LJ. Probing genomic aspects of the multi-host pathogen *Clostridium perfringens* reveals significant pangenome diversity, and a diverse array of virulence factors. Front Microbiol. 2017;8.

106. Escobar-Páramo P, Clermont O, Blanc-Potard A-B, Bui H, Le Bouguénec C, Denamur E. A Specific Genetic Background Is Required for Acquisition and Expression of Virulence Factors in Escherichia coli. Mol Biol Evol. 2004;21(6):1085–94.

107. Bonnici V, Maresi E, Giugno R. Challenges in gene-oriented approaches for pangenome content discovery. Brief Bioinform. 2021;22(3).

108. Wu H, Wang D, Gao F. Toward a high-quality pan-genome landscape of *Bacillus subtilis* by removal of confounding strains. Brief Bioinform. 2020 (2).

109. Perna NT. Genomics of Escherichia and Shigella. In: Wiedmann M, Zhang W, editors. Genomics of Foodborne Bacterial Pathogens. Springer Science and Business Media LLC; 2011. 119–40.

110. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Studying Gene Expression and Function. 2002;

111. Makrodimitris S, Ham RCHJ van, Reinders MJT. Automatic Gene Function Prediction in the 2020's. Genes. 2020;11(11):1264.

112. Zhao Y, Wang J, Chen J, Zhang X, Guo M, Yu G. A Literature Review of Gene Function Prediction by Modeling Gene Ontology. Front Genet. 2020; 0:400.

113. Power RA, Parkhill J, de Oliveira T. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet 2016 181. 2016;18(1):41–50.

114. Liu F, Zhu Y, Yi Y, Lu N, Zhu B, Hu Y. Comparative genomic analysis of *Acinetobacter baumannii* clinical isolates reveals extensive genomic variation and diverse antibiotic resistance determinants. BMC Genomics 2014 151. 2014;15(1):1–14.

115. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, et al. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. Antimicrob Agents Chemother. 2016;60(9):5515–20.

116. Earle SG, Wu CH, Charlesworth J, Stoesser N, Gordon NC, Walker TM, et al. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. Nat Microbiol. 2016;1(5).

117. Saber MM, Jesse Shapiro B. Benchmarking bacterial genome-wide association study methods using simulated genomes and phenotypes. Microb Genomics. 2020;6(3).

118. Fredriksson-Ahomaa M, Joutsen S, Laukkanen-Ninios R. Identification of *Yersinia* at the Species and Subspecies Levels Is Challenging. Vol. 5, Current Clinical Microbiology Reports. Springer; 2018. p. 135–42.

119. Lindström M, Korkeala H. Laboratory diagnostics of botulism. Clin Microbiol Rev. 2006;19(2):298–314.

120. Hatheway CL. Toxigenic clostridia. Clin Microbiol Rev. 1990;3(1):66–98.

121. Smith T, Hill K, Raphael B. Historical and current perspectives on *Clostridium botulinum* diversity. Res Microbiol. 2015;166(4):290–302.

122. Midura T, Arnon S. Infant botulism: Identification of *Clostridium botulinum* and its toxins in faeces. Lancet. 1976;308(7992):934–6.

123. Davis J, Mattman L, Wiley M. *Clostridium botulinum* in a fatal wound infection. J Am Med Assoc. 1951;146(7):646–8.

124. Harris R, Anniballi F, Austin J. Adult Intestinal Toxemia Botulism. Toxins (Basel). 2020;12(2).

125. Passaro DJ, Werner SB, McGee J, Kenzie WR Mac, Vugia DJ. Wound Botulism Associated With Black Tar Heroin Among Injecting Drug Users. JAMA. 1998;279(11):859–63.

126. Koepke R, Sobel J, Arnon S. Global occurrence of infant botulism, 1976-2006. Pediatrics. 2008;122(1).

127. Sebaihia M, Peck MW, Minton NP, Thomson NR, Holden MTG, Mitchell WJ, et al. Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. Genome Res. 2007;17(7):1082–92.

128. Hutson RA, Thompson DE, Collins MD. Genetic interrelationships of saccharolytic *Clostridium botulinum* types B, E and F and related clostridia as revealed by small-subunit rRNA gene sequences. FEMS Microbiol Lett. 1993;108(1):103–10.

129. Hyytia E, Hielm S, Bjorkroth J, Korkeala H. Biodiversity of *Clostridium botulinum* type E strains isolated from fish and fishery products. Appl Environ Microbiol. 1999; 65:2057–64.

130. Aureli P, Fenicia L, Pasolini B, Gianfranceschi M, McCroskey L, Hatheway C. Two cases of type E infant botulism caused by neurotoxigenic *Clostridium butyricum* in Italy. J Infect Dis. 1986;154(2):207–11.

131. Suen J, Hatheway C, Steigerwalt A, Brenner D. Genetic confirmation of identities of neurotoxigenic *Clostridium baratii* and *Clostridium butyricum* implicated as agents of infant botulism. J Clin Microbiol. 1988;26(10):2191–2.

132. Brunt J, van Vliet AHM, Carter AT, Stringer SC, Amar C, Grant KA, et al. Diversity of the Genomes and Neurotoxins of Strains of *Clostridium botulinum* Group I and *Clostridium sporogenes* Associated with Foodborne, Infant and Wound Botulism. Toxins (Basel). 2020;12(9):586.

133. Williamson CHD, Sahl JW, Smith TJ, Xie G, Foley BT, Smith LA, et al. Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing clostridia. BMC Genomics 2016 171. 2016;17(1):1–20.

134. Smith T, Williamson CHD, Hill K, Sahl J, Keim P. Botulinum neurotoxin-producing bacteria. Isn't it time that we called a species a species? MBio. 2018;9(5).

135. Carter AT, Paul CJ, Mason DR, Twine SM, Alston MJ, Logan SM, et al. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. BMC Genomics. 2009; 10:115.

136. Korkeala H, Lindström M. Clostridium botulinum. In: Korkeala H, editor. Elintarvikehygienia - ympäristöhygienia, elintarvike- ja ympäristötoksikologia. 1st ed. Helsinki: WSOY Oppimateriaalit; 2007. 24–33.

137. Carter AT, Peck MW. Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. Res Microbiol. 2015;166(4):303–17.

138. Peck MW, Smith TJ, Anniballi F, Austin JW, Bano L, Bradshaw M, et al. Historical Perspectives and Guidelines for Botulinum Neurotoxin Subtype Nomenclature. Toxins 2017, Vol 9, Page 38. 2017, 18;9(1):38.

139. Brunt J, van Vliet AHM, Stringer SC, Carter AT, Lindström M, Peck MW. Pan-Genomic Analysis of *Clostridium botulinum* Group II (Non-Proteolytic *C. botulinum*) Associated with Foodborne Botulism and Isolated from the Environment. Toxins (Basel). 2020;12(5):306.

140. Hielm S, Björkroth J, Hyytiä E, Korkeala H. Prevalence of *Clostridium botulinum* in Finnish trout farms: pulsed-field gel electrophoresis typing reveals extensive genetic diversity among type E isolates. Appl Environ Microbiol. 1998;64(11):4161–7.

141. Hyytiä E, Björkroth J, Hielm S, Korkeala H. Characterisation of *Clostridium botulinum* groups I and II by randomly amplified polymorphic DNA analysis and repetitive element sequence-based PCR. Int J Food Microbiol. 1999;48(3):179–89.

142. Keto-Timonen R, Nevas M, Korkeala H. Efficient DNA fingerprinting of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis. Appl Environ Microbiol. 2005;71(3):1148–54.

143. Korkeala H, Stengel G, Hyytiä E, Vogelsang B, Bohl A, Wihlman H, et al. Type E botulism associated with vacuum-packaged hot-smoked whitefish. Int J Food Microbiol. 1998;43(1–2):1–5.

144. Weedmark KA, Mabon P, Hayden KL, Lambert v., Van Domselaar G, Austin JW, et al. *Clostridium botulinum* Group II isolate phylogenomic profiling using whole-genome sequence data. Appl Environ Microbiol. 2015;81(17):5938–48.

145. Derman Y, Lindström M, Selby K, Korkeala H. Growth of Group II *Clostridium botulinum* Strains at Extreme Temperatures. J Food Prot. 2011;74:1797–804.

146. Woudstra C, Le Maréchal C, Souillard R, Bayon-Auboyer M-H, Mermoud I, Desoutter D, et al. New Insights into the Genetic Diversity of *Clostridium botulinum* Group III through Extensive Genome Exploration. Front Microbiol. 2016; 0:757.

147. Sakaguchi Y, Suzuki T, Yamamoto Y, Nishikawa A, Oguma K. Genomics of *Clostridium botulinum* group III strains. Res Microbiol. 2015;166(4):318–25.

148. Fillo S, Giordani F, Tonon E, Drigo I, Anselmo A, Fortunato A, et al. Extensive Genome Exploration of *Clostridium botulinum* Group III Field Strains. Microorg 2021, Vol 9, Page 2347. 2021;9(11):2347.

149. Suen J, Hatheway C, Steigerwalt A, Brenner D. *Clostridium argentinense* sp. nov.: A Genetically Homogeneous Group Composed of All Strains of *Clostridium botulinum* Toxin Type G and Some Nontoxigenic Strains Previously Identified as *Clostridium subterminale* or *Clostridium hastiforme*. Int J Syst Evol Microbiol. 1988;38(4):375–81.

150. Lindström M, Keto R, Markkula A, Nevas M, Hielm S, Korkeala H. Multiplex PCR Assay for Detection and Identification of *Clostridium botulinum* Types A, B, E, and F in Food and Fecal Material. Appl Environ Microbiol. 2001;67(12):5694–9.

151. Lindström M, Nevas M, Korkeala H. Detection of *Clostridium botulinum* by Multiplex PCR in Foods and Feces. Food-Borne Pathog. 2006;37–45.

152. Williamson CHD, Vazquez AJ, Hill K, Smith TJ, Nottingham R, Stone NE, et al. Differentiating botulinum neurotoxin-producing clostridia with a simple, multiplex PCR assay. Appl Environ Microbiol. 2017;83(18).

153. Woudstra C, Le Maréchal C, Souillard R, Bayon-Auboyer MH, Anniballi F, Auricchio B, et al. Molecular gene profiling of *Clostridium botulinum* Group III and its detection in naturally contaminated samples originating from various European Countries. Appl Environ Microbiol. 2015;81(7):2495–505.

154. Skarin H, Åberg A, Woudstra C, Hansen T, Löfström C, Koene M, et al. The Workshop on Animal Botulism in Europe. Biosecurity and bioterrorism: biodefense strategy, practice, and science. 2013, 11: 183-190.

155. Rasetti-Escargueil C, Lemichez E, Popoff MR. Public Health Risk Associated with Botulism as Foodborne Zoonoses. Toxins 2020, Vol 12, Page 17. 2019;12(1):17.

156. Fleck-Derderian S, Shankar M, Rao AK, Chatham-Stephens K, Adjei S, Sobel J, et al. The Epidemiology of Foodborne Botulism Outbreaks: A Systematic Review. Clin Infect Dis. 2018;66:73–81.

157. Leclair D, Fung J, Isaac-Renton JL, Proulx JF, May-Hadford J, Ellis A, et al. Foodborne botulism in Canada, 1985-2005. Emerg Infect Dis. 2013;19:961–8.

158. ECDC. Botulism. In: ECDC. Annual epidemiological report for 2015. Stockholm; 2018.

159. Souillard R, Woudstra C, Maréchal C Le, Dia M, Bayon-Auboyer MH, Chemaly M, et al. Investigation of *Clostridium botulinum* in commercial poultry farms in France between 2011 and 2013. Avian Pathology. 2014;43(5):458–64.

160. Ono T, Azuma R, Kato T, Takeuchi S, Suto T. Outbreaks of type C botulism in waterfowl in Japan. National Institute of Animal Health Quarterly, 1982, 22.3: 102-114.

161. Takeda M, Tsukamoto K, Kohda T, Matsui M, Mukamoto M, Kozaki S. Characterization of the Neurotoxin Produced by Isolates Associated with Avian Botulism. Avian Dis. 2005;49(3):376–81.

162. Nakamura K, Kohda T, Umeda K, Yamamoto H, Mukamoto M, Kozaki S. Characterization of the D/C mosaic neurotoxin produced by *Clostridium botulinum* associated with bovine botulism in Japan. Vet Microbiol. 2010;140(1–2):147–54.

163. Ventujol A, Decors A, Maréchal C, Toux J, Allain V, Mazuet C, et al. Avian botulism in France: analysis of cases reported by two surveillance networks both in the wild and in poultry farms between 2000 and 2013. Epidémiologie Santé Anim. 2017; 72:85–102.

164. Smith L. The occurrence of *Clostridium botulinum* and *Clostridium tetani* in the soil of the United States. Health Lab Sci. 1978;15(2):74–80.

165. Midura TF, Snowden S, Wood RM, Arnon SS. Isolation of *Clostridium botulinum* from honey. J Clin Microbiol. 1979;9(2):282–3.

166. Huss HH. Distribution of *Clostridium botulinum*. Appl Environ Microbiol. 1980;39(4):764–9.

167. Hyytiä E, Hielm S, Korkeala H. Prevalence of *Clostridium botulinum* type E in Finnish fish and fishery products. Epidemiol Infect. 1998;120(3):245–50.

168. Leclair D, Farber JM, Doidge B, Blanchfield B, Suppa S, Pagotto F, et al. Distribution of *Clostridium botulinum* type E strains in Nunavik, Northern Quebec, Canada. Appl Environ Microbiol. 2013;79(2):646–54.

169. Lindström M, Kiviniemi K, Korkeala H. Hazard and control of group II (non-proteolytic) *Clostridium botulinum* in modern food processing. Int J Food Microbiol. 2006;108(1):92–104.

170. Sobel J, Tucker N, Sulka A, McLaughlin J, Maslanka S. Foodborne Botulism in the United States, 1990–2000. Emerg Infect Dis 2004. 10(9), 1606.

171. Myllykoski J, Nevas M, Lindström M, Korkeala H. The detection and prevalence of *Clostridium botulinum* in pig intestinal samples. Int J Food Microbiol. 2006;110(2):172–7.

172. Dahlenborg M, Borch E, Rådström P. Development of a Combined Selection and Enrichment PCR Procedure for *Clostridium botulinum* Types B, E, and F and Its Use to Determine Prevalence in Fecal Samples from Slaughtered Pigs. Appl Environ Microbiol. 2001;67(10):4781–8.

173. Kobayashi T, Watanabe K, Ueno K. Distribution of *Clostridium botulinum* and *Clostridium tetani* in Okinawa Prefecture. Kansenshogaku Zasshi. 1992;66(12):1639–44.

174. Van Kruiningen HJ, Nyaoke CA, Sidor IF, Fabis JJ, Hinckley LS, Lindell KA. Clostridial abomasal disease in Connecticut dairy calves. The Canadian Veterinary Journal, 2009, 50.8: 857.

175. Kiu R, Hall LJ. An update on the human and animal enteric pathogen *Clostridium perfringens*. Emerg Microbes Infect. 2018;7(1):1–15.

176. Mehdizadeh Gohari I, Unterer S, Whitehead AE, Prescott JF. NetF-producing *Clostridium perfringens* and its associated diseases in dogs and foals. J Vet Diagnostic Investig. 2020 ;32(2):230–8.

177. Songer JG. Clostridial enteric diseases of domestic animals. Vol. 9, Clinical Microbiology Reviews. American Society for Microbiology; 1996. 216–34.

178. Shimizu T, Ohtani K, Hirakawa H, Ohshima K, Yamashita A, Shiba T, et al. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. Proc Natl Acad Sci U S A. 2002;99(2):996–1001.

179. Deguchi A, Miyamoto K, Kuwahara T, Miki Y, Kaneko I, Li J, et al. Genetic characterization of type A enterotoxigenic *Clostridium perfringens* strains. PLoS One. 2009;4(5).

180. Lahti P, Lindström M, Somervuo P, Heikinheimo A, Korkeala H. Comparative Genomic Hybridization Analysis Shows Different Epidemiology of Chromosomal and Plasmid-Borne cpe-Carrying *Clostridium perfringens* Type A. Bruggemann H, editor. PLoS One. 2012;7(10): e46162.

181. Xiao Y, van Hijum SAFT, Abee T, Wells-Bennik MHJ. Genome-Wide Transcriptional Profiling of *Clostridium perfringens* SM101 during Sporulation Extends the Core of Putative Sporulation Genes and Genes Determining Spore Properties and Germination Characteristics. Zhou D, editor. PLoS One. 2015;10(5): e0127036.

182. Abdel-Glil M, Thomas P, Linde J, Busch A, Wieler L, Neubauer H, et al. Comparative in silico genome analysis of *Clostridium perfringens* unravels stable phylogroups with different genome characteristics and pathogenic potential. Sci Reports 2021 111. 2021;11(1):1–15.

183. Lugli GA, Milani C, Mancabelli L, Turroni F, Ferrario C, Duranti S, et al. Ancient bacteria of the ötzi's microbiome: A genomic tale from the Copper Age. Microbiome. 2017;5(1):5.

184. Lacey JA, Allnutt TR, Vezina B, Van TTH, Stent T, Han X, et al. Whole genome analysis reveals the diversity and evolutionary relationships between necrotic enteritis-causing strains of *Clostridium perfringens*. BMC Genomics. 2018;19(1).

185. Li J, Miyamoto K, Sayeed S, McClane BA. Organization of the cpe locus in CPE-positive *Clostridium perfringens* type C and D isolates. PLoS One. 2010;5(6).

186. Mahamat Abdelrahim A, Radomski N, Delannoy S, Djellal S, Le Négrate M, Hadjab K, et al. Large-Scale Genomic Analyses and Toxinotyping of *Clostridium perfringens* Implicated in Foodborne Outbreaks in France. Front Microbiol. 2019;10(03):777.

187. Cornillot E, Saint‐Joanis B, Daube G, Katayama S‐i, Granum PE, Canard B, et al. The enterotoxin gene (cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne. Mol Microbiol. 1995;15(4):639‐47.

188. Brynestad S, Synstad B, Granum PE. The *Clostridium perfringens* enterotoxin gene is on a transposable element in type A human food poisoning strains. Microbiology. 1997;143(7):2109–15.

189. Miyamoto K, Yumine N, Mimura K, Nagahama M, Li J, McClane BA, et al. Identification of novel *Clostridium perfringens* type E strains that carry an iota toxin plasmid with a functional enterotoxin gene. PLoS One. 2011;6(5).

190. McClane BA. *Clostridium perfringens*. In: Doyle MP, Beuchat LR, Montville TJ, editors. Food Microbiology: fundamentals and frontiers. Washington, D.C.: American Society for Microbiology (ASM); 1997. 305–26.

191. Asha NJ, Tompkins D, Wilcox MH. Comparative analysis of prevalence, risk factors, and molecular epidemiology of antibiotic-associated diarrhea due to *Clostridium difficile*, *Clostridium perfringens*, and *Staphylococcus aureus*. J Clin Microbiol. 2006;44(8):2785–91.

192. Da Silva Felício MT, Hald T, Liebana E, Allende A, Hugas M, Nguyen-The C, et al. Risk ranking of pathogens in ready-to-eat unprocessed foods of non-animal origin (FoNAO) in the EU: Initial evaluation using outbreak data (2007-2011). Int J Food Microbiol. 2015;195(02):9–19.

193. Keto-Timonen R, Heikinheimo A, Eerola E, Korkeala H. Identification of *Clostridium* species and DNA fingerprinting of *Clostridium perfringens* by amplified fragment length polymorphism analysis. J Clin Microbiol. 2006;44(11):4057–65.

194. Li J, Sayeed S, McClane BA. Prevalence of Enterotoxigenic *Clostridium perfringens* Isolates in Pittsburgh (Pennsylvania) Area Soils and Home Kitchens. Appl Environ Microbiol. 2007;73(22):7218 LP – 7224.

195. Lahti P, Heikinheimo A, Johansson T, Korkeala H. *Clostridium perfringens* type A strains carrying a plasmid-borne enterotoxin gene (genotype IS1151-cpe or IS1470-like-cpe) as a common cause of food poisoning. J Clin Microbiol. 2008;46(1):371–3.

196. Lindström M, Heikinheimo A, Lahti P, Korkeala H. Novel insights into the epidemiology of *Clostridium perfringens* type A food poisoning. Food Microbiology. Academic Press; 2011. 28:192–8.

197. Ma M, Li J, Mcclane BA. Genotypic and phenotypic characterization of *Clostridium perfringens* isolates from darmbrand cases in post-world war II Germany. Infect Immun. 2012;80(12):4354–63.

198. Hu WS, Kim H, Koo OK. Molecular genotyping, biofilm formation and antibiotic resistance of enterotoxigenic *Clostridium perfringens* isolated from meat supplied to school cafeterias in South Korea. Anaerobe. 2018; 52:115–21.

199. Sparks SG, Carman RJ, Sarker MR, McClane BA. Genotyping of enterotoxigenic *Clostridium perfringens* fecal isolates associated with antibiotic-associated diarrhea and food poisoning in North America. J Clin Microbiol. 2001;39(3):883–8.

200. Heikinheimo A, Lindström M, Granum PE, Korkeala H. Humans as reservoir for enterotoxin gene-carrying *Clostridium perfringens* type A. Emerg Infect Dis. 2006;12(11):1724–9.

201. Kiu R, Caim S, Painset A, Pickard D, Swift C, Dougan G, et al. Phylogenomic analysis of gastroenteritis-associated *Clostridium perfringens* in England and wales over a 7-year period indicates distribution of clonal toxigenic strains in multiple outbreaks and extensive involvement of enterotoxin-encoding (Cpe) plasmids. Microb Genomics. 2019;5(10).

202. Li J, McClane BA. Comparative Effects of Osmotic, Sodium Nitrite-Induced, and pH-Induced Stress on Growth and Survival of *Clostridium perfringens* Type A Isolates Carrying Chromosomal or Plasmid-Borne Enterotoxin Genes. Appl Environ Microbiol. 2006;72(12):7620–5.

203. Sarker MR, Shivers RP, Sparks SG, Juneja VK, McClane BA. Comparative experiments to examine the effects of heating on vegetative cells and spores of *Clostridium perfringens* isolates carrying plasmid genes versus chromosomal enterotoxin genes. Appl Environ Microbiol. 2000;66(8):3234–40.

204. Li J, McClane BA. Further comparison of temperature effects on growth and survival of *Clostridium perfringens* type A isolates carrying a chromosomal or plasmid-borne enterotoxin gene. Appl Environ Microbiol. 2006;72(7):4561–8.

205. Paredes-Sabja D, Gonzalez M, Sarker MR, Torres JA. Combined Effects of Hydrostatic Pressure, Temperature, and pH on the Inactivation of Spores of *Clostridium perfringens* Type A and *Clostridium sporogenes* in Buffer Solutions. J Food Sci. 2007;72(6):M202–6.

206. Fredriksson-Ahomaa M, Lindström M, Korkeala H. *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. Pathog Toxins Foods. 2014;164–80.

207. McNally A, Thomson NR, Reuter S, Wren BW. "Add, stir and reduce": Yersinia spp. as model bacteria for pathogen evolution. Nature Reviews Microbiology. 2016;14:177–90.

208. European Food Safety Authority. The European Union One Health 2019 Zoonoses Report. Approved: 19 January 2021. Available from: https://doi.org/10.2903/j.efsa.2021.6406

209. Bottone E. *Yersinia enterocolitica*: Revisitation of an enduring human pathogen. Clin Microbiol Newsl. 2015;37(1):1–8.

210. Nuorti JP, Niskanen T, Hallanvuo S, Mikkola J, Kela E, Hatakka M, et al. A widespread outbreak of *Yersinia pseudotuberculosis* O:3 infection from iceberg lettuce. J Infect Dis. 2004;189(5):766–74.

211. Amphlett A. Far east scarlet-like fever: A review of the epidemiology, symptomatology, and role of superantigenic toxin: *Yersinia pseudotuberculosis*-derived mitogen A. Vol. 3, Open Forum Infectious Diseases. Oxford University Press; 2016.

212. Thomson N, Howard S, Wren BW, Holden MTG, Crossman L, Challis GL, et al. The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. PLoS Genet. 2006;2(12):e206.

213. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci. 1999;96(24):14043–8.

214. Zhou D, Han Y, Song Y, Tong Z, Wang J, Guo Z, et al. DNA microarray analysis of genome dynamics in *Yersinia pestis*: Insights into bacterial genome microevolution and niche adaptation. J Bacteriol. 2004;186(15):5138–46.

215. Tauxe R V, Vandepitte J, Wauters G, Martin SM, Goossens V, De Mol P, et al. *Yersinia enterocolitica* infections and pork: the missing link. Lancet. 1987;1(8542):1129–32.

216. Hall M, Chattaway MA, Reuter S, Savin C, Strauch E, Carniel E, et al. Use of whole-genus genome sequence data to develop a multilocus sequence typing tool that accurately identifies *Yersinia* isolates to the species and subspecies levels. J Clin Microbiol. 2015;53(1):35–42.

217. Wauters G, Kandolo K, Janssens M. Revised biogrouping scheme of *Yersinia enterocolitica*. Contrib Microbiol Immunol. 1987; 9:14–21.

218. Neubauer H, Aleksic S, Hensel A, Finke EJ, Meyer H. *Yersinia enterocolitica* 16S rRNA gene types belong to the same genospecies but form three homology groups. Int J Med Microbiol. 2000;290(1):61–4.

219. Bhagat N, Virdi JS. The Enigma of *Yersinia enterocolitica* biovar 1A. Crit Rev Microbiol. 2011;37(1):25–39.

220. Niskanen T., Fredriksson-Ahomaa M., Korkeala H. *Yersinia pseudotuberculosis* with Limited Genetic Diversity Is a Common Finding in Tonsils of Fattening Pigs. J Food Prot. 2002;65(3):540–5.

221. Laukkanen-Ninios R, Didelot X, Jolley KA, Morelli G, Sangal V, Kristo P, et al. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environ Microbiol. 2011;13(12):3114–27.

222. Reuter S, Connor TR, Barquist L, Walker D, Feltwell T, Harris SR, et al. Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proc Natl Acad Sci U S A. 2014;111(18):6768–73.

223. Bottone EJ. *Yersinia enterocolitica*: overview and epidemiologic correlates. Microbes Infect. 1999;1(4):323–33.

224. Fredriksson-Ahomaa M. Enteropathogenic *Yersinia* spp. In: Zoonoses-Infections Affecting Humans and Animals: Focus on Public Health Aspects. Springer, Dordrecht; 2015. 213–34.

225. Fredriksson-Ahomaa M, Korkeala H. Low Occurrence of Pathogenic *Yersinia enterocolitica* in Clinical, Food, and Environmental Samples: a Methodological Problem. Clin Microbiol Rev. 2003;16(2):220–9.

226. Skurnik M, Peippo A, Ervela E. Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. Mol Microbiol. 2000;37(2):316–30.

227. Hunter E, Greig DR, Schaefer U, Wright MJ, Dallman TJ, McNally A, et al. Identification and typing of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* isolated from human clinical specimens in England between 2004 and 2018. J Med Microbiol. 2019;68(4):538–48.

228. Bogdanovich T, Carniel E, Fukushima H, Skurnik M. Use of O-Antigen Gene Cluster-Specific PCRs for the Identification and O-Genotyping of *Yersinia pseudotuberculosis* and *Yersinia pestis*. J Clin Microbiol. 2003;41(11):5103–12.

229. Rakin A, Garzetti D, Bouabe H, Sprague LD. *Yersinia enterocolitica*. Mol Med Microbiol Second Ed. 2014, 26;2–3:1319–44.

230. Le Guern AS, Martin L, Savin C, Carniel E. Yersiniosis in France: Overview and potential sources of infection. Int J Infect Dis. 2016, 1;46:1–7.

231. Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E. The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal asn tRNA genes. Mol Microbiol. 1998;30(5):965–78.

232. Schubert S, Rakin A, Heesemann J. The high-pathogenicity island (HPI): evolutionary and functional aspects. Int J Med Microbiol. 2004, 24;294(2–3):83–94.

233. Joutsen S, Laukkanen-Ninios R, Henttonen H, Niemimaa J, Voutilainen L, Kallio E, et al. *Yersinia* spp. in Wild Rodents and Shrews in Finland. Vector-Borne Zoonotic Dis. 2017, 1;17(5):303–11.

234. Platt-Samoraj A, Syczyło K, Szczerba-Turek A, Bancerz-Kisiel A, Jabłoński A, Łabuć S, et al. Presence of ail and ystB genes in *Yersinia enterocolitica* biotype 1A isolates from game animals in Poland. Vet J. 2017,1;221:11–3.

235. Fredriksson-Ahomaa M, Wacheck S, Koenig M, Stolle A, Stephan R. Prevalence of pathogenic *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* in wild boars in Switzerland. Int J Food Microbiol. 2009;135(3):199–202.

236. Ossiprandi M, Zerbini L. *Yersinia enterocolitica* survival in aquatic environment: epidemiological significance. J Adv Biol. 2022;6(3):1113–20.

237. Wang X, Li Y, Jing H, Ren Y, Zhou Z, Wang S, et al. Complete genome sequence of a *Yersinia enterocolitica* "Old World" (3/O:9) strain and comparison with the "New World" (1B/O:8) strain. J Clin Microbiol. 2011;49(4):1251–9.

238. Schubert S, Bockemühl J, Brendler U, Heesemann J. First isolation of virulent *Yersinia enterocolitica* O8, biotype 1B in Germany. Eur J Clin Microbiol Infect Dis. 2003;22(1):66–8.

239. Fredriksson-Ahomaa M, Hallanvuo S, Korte T, Siitonen A, Korkeala H. Correspondence of genotypes of sporadic *Yersinia enterocolitica* bioserotype 4/O:3 strains from human and porcine sources. Epidemiol Infect. 2001;127(1):37–47.

240. Gütler M, Alter T, Kasimir S. Prevalence of *Yersinia enterocolitica* in fattening pigs. J Food Prot. 2005;68(4):850–4.

241. Fredriksson-Ahomaa M. Sporadic human *Yersinia enterocolitica* infections caused by bioserotype 4/O : 3 originate mainly from pigs. J Med Microbiol. 2006,1;55(6):747–9.

242. Gierczyñski R, Szych J, Rastawicki W, Wardak S, Jagielski M. Molecular characterization of human clinical isolates of *Yersinia enterocolitica* bioserotype 1B/O8 in Poland: emergence and dissemination of three highly related clones. J Clin Microbiol. 2009;47(4):1225–8.

243. Ortiz Martinez P, Mylona S, Drake I, Fredriksson-Ahomaa M, Korkeala H, Corry JEL. Wide variety of bioserotypes of enteropathogenic *Yersinia* in tonsils of English pigs at slaughter. Int J Food Microbiol. 2010;139(1–2):64–9.

244. Ortiz Martínez P, Fredriksson-Ahomaa M, Pallotti A, Rosmini R, Houf K, Korkeala H. Variation in the prevalence of enteropathogenic *Yersinia* in slaughter pigs from Belgium, Italy, and Spain. Foodborne Pathog Dis. 2011;8(3):445–50.

245. Fredriksson-Ahomaa M, Stolle A, Korkeala H. Molecular epidemiology of *Yersinia enterocolitica* infections. Vol. 47, FEMS Immunology and Medical Microbiology. FEMS Immunol Med Microbiol; 2006. 315–29.

246. Ackers ML, Schoenfeld S, Markman J, Smith MG, Nicholson MA, DeWitt W, et al. An outbreak of *Yersinia enterocolitica* O:8 infections associated with pasteurized milk. J Infect Dis. 2000;181(5):1834–7.

247. MacDonald E, Heier BT, Nygård K, Stalheim T, Cudjoe KS, Skjerdal T, et al. *Yersinia enterocolitica* outbreak associated with ready-to-eat salad mix, Norway, 2011. Emerg Infect Dis. 2012;18(9):1496–9.

248. Grahek-Ogden D, Schimmer B, Cudjoe KS, Nygård K, Kapperud G. Outbreak of *Yersinia enterocolitica* serogroup O:9 infection and processed pork, Norway. Emerg Infect Dis. 2007;13(5):754–6.

249. Tacket C, Narain J, Sattin R, Lofgren J, Konigsberg C, Rendtorff R, et al. A Multistate Outbreak of Infections Caused by *Yersinia enterocolitica* Transmitted by Pasteurized Milk. JAMA J Am Med Assoc. 1984;251(4):483–6.

250. Tacket C, Harris N, Allard J, Nolna C, Nissinen A, Quan T, et al. An outbreak of *Yersinia enterocolitica* infections caused by contaminated tofu (soybean curd). Am J Epidemiol. 1985;121(5):705–11.

251. Sakai T, Nakayama A, Hashida M, Yamamoto Y, Takebe H, Imai S. Laboratory and Epidemiology Communications Outbreak of Food Poisoning by *Yersinia enterocolitica* Serotype O8 in Nara Prefecture: the First Case Report in Japan. Vol. 58, Jpn. J. Infect. Dis. 2005.

252. Ortiz Martinez P. Prevalence of enteropathogenic *Yersinia* in pigs from different European countries and contamination in the pork production chain. Faculty of Veterinary Medicine at the University of Helsinki; 2010.

253. Fukushima H, Gomyoda M. Intestinal carriage of *Yersinia pseudotuberculosis* by wild birds and mammals in Japan. Appl Environ Microbiol. 1991;57(4):1152–5.

254. Magistrali CF, Cucco L, Pezzotti G, Farneti S, Cambiotti V, Catania S, et al. Characterisation of *Yersinia pseudotuberculosis* isolated from animals with yersiniosis during 1996-2013 indicates the presence of pathogenic and Far Eastern strains in Italy. Vet Microbiol. 2015;180(1–2):161–6.

255. Pärn T, Hallanvuo S, Salmenlinna S, Pihlajasaari A, Heikkinen S, Telkki-Nykänen H, et al. Outbreak of *Yersinia pseudotuberculosis* O:1 infection associated with raw milk consumption, Finland, spring 2014. Eurosurveillance. 2015;20(40):30033.

256. Jalava K, Hakkinen M, Valkonen M, Nakari U-M, Palo T, Hallanvuo S, et al. An outbreak of gastrointestinal illness and erythema nodosum from grated carrots contaminated with *Yersinia pseudotuberculosis*. J Infect Dis. 2006;194(9):1209–16.

257. Jalava K, Hallanvuo S, Nakari U-M, Ruutu P, Kela E, Heinäsmäki T, et al. Multiple outbreaks of *Yersinia pseudotuberculosis* infections in Finland. J Clin Microbiol. 2004;42(6):2789–91.

258. Tertti R, Granfors K, Lehtonen O-P, Mertsola J, Makela A-L, Valimaki I, et al. An Outbreak of *Yersinia pseudotuberculosis* Infection. J Infect Dis. 1984;149(2):245–50.

259. Rimhanen-Finne R, Niskanen T, Hallanvuo S, Makary P, Haukka K, Pajunen S, et al. *Yersinia pseudotuberculosis* causing a large outbreak associated with carrots in Finland, 2006. Epidemiol Infect. 2009;137(3):342–7.

260. Kangas S, Takkinen J, Hakkinen M, Nakari U-M, Johansson T, Henttonen H, et al. *Yersinia pseudotuberculosis* O:1 traced to raw carrots, Finland. Emerg Infect Dis. 2008;14(12):1959–61.

261. Williamson DA, Baines SL, Carter GP, Da Silva AG, Ren X, Sherwood J, et al. Genomic insights into a sustained national outbreak of *Yersinia pseudotuberculosis*. Genome Biol Evol. 2016;8(12):3806–14.

262. Bergann T, Kleemann J, Sohr D. Model studies of psychrotrophia in *Yersinia enterocolitica*. Zentralbl Veterinarmed B. 1995;42(9):523–31.

263. Bottone E, Bercovier H, Mollaret H. Genus XLI. *Yersinia*. In: Garrity G, Brenner D, Krieg N, Staley J, editors. Bergey's Manual of Systematic Bacteriology. East Lansing: Springer-Verlag; 2005. 838–48.

264. Walker SJ, Archer P, Banks G. Growth of *Yersinia enterocolitica* at chill temperatures in milk and other media. Milchwissenschaft. 1990;45.8:503–6.

265. Toora S, Budu-Amoako E, Ablett R, Smith J. Effect of High-Temperature Short-Time Pasteurization, Freezing and Thawing and Constant Freezing, on the Survival of *Yersinia enterocolitica* in Milk. J Food Prot. 1992;55(10):803–5.

266. Asplund K, Nurmi E, Hirn J, Hirvi T, Hill P. Survival of *Yersinia enterocolitica* in Fermented Sausages Manufactured With Different Levels of Nitrite and Different Starter Cultures. J Food Prot. 1993;56(8):710–2.

267. Preston NW, Maitland HB. The Influence of Temperature on the Motility of *Pasteurella pseudotuberculosis*. Microbiology. 1952;7(1–2):117–28.

268. Laukkanen-Ninios R, Fredriksson-Ahomaa M, Korkeala H. Enteropathogenic *Yersinia* in the Pork Production Chain: Challenges for Control. Compr Rev Food Sci Food Saf. 2014;13(6):1165–91.

269. Keto-Timonen R, Pöntinen A, Aalto-Araneda M, Korkeala H. Growth of *Yersinia pseudotuberculosis* strains at different temperatures, ph values, and NaCl and ethanol concentrations. J Food Prot. 2018;81(1):142–9.

270. Harrison WA, Peters AC, Fielding LM. Growth of *Listeria monocytogenes* and *Yersinia enterocolitica* colonies under modified atmospheres at 4 and 8°C using a model food system. J Appl Microbiol. 2000;88(1):38–43.

271. Goverde RLJ, Kusters JG, Veld JHJH in't. Growth rate and physiology of *Yersinia enterocolitica*; influence of temperature and presence of the virulence plasmid. J Appl Bacteriol. 1994;77(1):96–104.

272. Hinderink K, Lindström M, Korkeala H. Group I *Clostridium botulinum* Strains Show Significant Variation in Growth at Low and High Temperatures. Vol. 72, Journal of Food Protection. 2009.

273. Ribeiro VB, Destro MT. *Listeria monocytogenes* Serotype 1/2b and 4b isolates from human clinical cases and foods show differences in tolerance to refrigeration and salt stress. J Food Prot. 2014;77(9):1519–26.

274. Barria C, Malecki M, Arraiano CM. Bacterial adaptation to cold. Vol. 159, Microbiology (United Kingdom). Microbiology (Reading); 2013. 2437–43.

275. Jones PG, Krah R, Tafuri SR, Wolffe AP. DNA gyrase, CS7.4, and the cold shock response in *Escherichia coli*. J Bacteriol. 1992;174(18):5798–802.

276. Wemekamp-Kamphuis HH, Karatzas AK, Wouters JA, Abee T. Enhanced levels of cold shock proteins in *Listeria monocytogenes* LO28 upon exposure to low temperature and high hydrostatic pressure. Appl Environ Microbiol. 2002;68(2):456–63.

277. Weber MHW, Marahiel MA. Coping with the cold: the cold shock response in the Gram-positive soil bacterium *Bacillus subtilis*. Philos Trans R Soc Lond B Biol Sci. 2002;357(1423):895–907.

278. Yu T, Keto-Timonen R, Jiang X, Virtanen J-P, Korkeala H. Insights into the Phylogeny and Evolution of Cold Shock Proteins: From Enteropathogenic *Yersinia* and *Escherichia coli* to Eubacteria. Int J Mol Sci. 2019;20(16):4059.

279. Eshwar AK, Guldimann C, Oevermann A, Tasara T. Cold-shock domain family proteins (Csps) are involved in regulation of virulence, cellular aggregation, and flagella-based motility in *Listeria monocytogenes*. Front Cell Infect Microbiol. 2017; 7:453.

280. Goldstein J, Pollitt NS, Inouye M. Major cold shock protein of *Escherichia coli*. Proc Natl Acad Sci U S A. 1990;87(1):283–7.

281. Yamanaka K, Fang L, Inouye M. The CspA family in *Escherichia coli*: Multiple gene duplication for stress adaptation. Vol. 27, Molecular Microbiology. Mol Microbiol; 1998:247–55.

282. Annamalai T, Venkitanarayanan K. Expression of Major Cold Shock Proteins and Genes by *Yersinia enterocolitica* in Synthetic Medium and Foods. J Food Prot. 2005;68(11):2454–8.

283. Neuhaus K, Francis KP, Rapposch S, Görg A, Scherer S. Pathogenic *Yersinia* species carry a novel, cold-inducible major cold shock protein tandem gene duplication producing both bicistronic and monocistronic mRNA. J Bacteriol. 1999;181:6449–55.

284. Söderholm H, Lindström M, Somervuo P, Heap J, Minton N, Lindén J, et al. cspB encodes a major cold shock protein in *Clostridium botulinum* ATCC 3502. Int J Food Microbiol. 2011;146(1):23–30.

285. Derman Y, Söderholm H, Lindström M, Korkeala H. Role of csp genes in NaCl, pH, and ethanol stress response and motility in *Clostridium botulinum* ATCC 3502. Food Microbiol. 2015; 46:463–70.

286. Dahlsten E, Isokallio M, Somervuo P, Lindström M, Korkeala H. Transcriptomic Analysis of (Group I) *Clostridium botulinum* ATCC 3502 Cold Shock Response. PLoS One. 2014;9(2):e89958.

287. Reuter S, Connor TR, Barquist L, Walker D, Feltwell T, Harris SR, et al. Parallel independent evolution of pathogenicity within the genus *Yersinia.* Proc Natl Acad Sci U S A. 2014;111(18):6768–73.

288. Bresolin G, Neuhaus K, Scherer S, Fuchs TM. Transcriptional analysis of long-term adaptation of *Yersinia enterocolitica* to low-temperature growth. J Bacteriol. 2006;188(8):2945–58.

289. Bresolin G, Morgan JAW, Ilgen D, Scherer S, Fuchs TM. Low temperature-induced insecticidal activity of *Yersinia enterocolitica*. Mol Microbiol. 2006;59(2):503–12.

290. Setlow P. Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. J Appl Microbiol. 2006;101(3):514–25.

291. Setlow P. I will survive: DNA protection in bacterial spores. Trends Microbiol. 2007;15(4):172–80.

292. Setlow P. Germination of spores of Bacillus species: What we know and do not know. J Bacteriol. 2014;196(7):1297–305.

293. Ando Y, Toshibumi T, Sunagawa H, Oka S. Heat resistance, Spore germination, and enterotoxigenicity of *Clostridium perfringens*. Vol. 29, Microbiol. Immunol. 1985.

294. Raju D, Sarker MR. Comparison of the levels of heat resistance of wild-type, cpe knockout, and cpe plasmid-cured *Clostridium perfringens* type A strains. Appl Environ Microbiol. 2005;71(11):7618–20.

295. Li J, McClane BA. A novel small acid soluble protein variant is important for spore resistance of most *Clostridium perfringens* food poisoning isolates. PLoS Pathog. 2008;48(5).

296. Li J, Paredes-Sabja D, Sarker MR, McClane BA. Further Characterization of *Clostridium perfringens* Small Acid Soluble Protein-4 (Ssp4) Properties and Expression. PLoS One. 2009;4(7):e6249.

297. Raju D, Waters M, Setlow P, Sarker MR. Investigating the role of small, acid-soluble spore proteins (SASPs) in the resistance of *Clostridium perfringens* spores to heat. BMC Microbiol. 2006;6(1):50.

298. Diao MM, André S, Membré JM. Meta-analysis of D-values of proteolytic *Clostridium botulinum* and its surrogate strain *Clostridium sporogenes* PA 3679. Int J Food Microbiol. 2014; 174:23–30.

299. Butler RRI, Schill KM, Wang Y, Pombert J-F. Genetic Characterization of the Exceptionally High Heat Resistance of the Non-toxic Surrogate *Clostridium sporogenes* PA 3679. Front Microbiol. 2017; 0:545.

300. Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL et al. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. Infect Immun. 2011;79:4286–98.

301. Lindström M, Hielm S, Nevas M, Tuisku S, Korkeala H. Proteolytic *Clostridium botulinum* Type B in the Gastric Content of a Patient with Type E Botulism Due to Whitefish Eggs. Foodborne Pathog Dis. 2004;1(1):53–7.

302. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013; 10:563–9.

303. Staden R, Judge DP, Bonfield JK. Managing Sequencing Projects in the GAP4 Environment. In: Introduction to Bioinformatics. Humana Press; 2003. p. 327–44.

304. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J, editor. PLoS One. 2014;9(11):e112963.

305. Myers GSA, Rasko DA, Cheung JK, Ravel J, Seshadri R, DeBoy RT, et al. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. Genome Res. 2006;16(8):1031–40.

306. Batzilla J, Höper D, Antonenka U, Heesemann J, Rakin A. Complete genome sequence of *Yersinia enterocolitica* subsp. *palearctica* serogroup O:3. J Bacteriol. 2011;193(8):2067.

307. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvement to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2016;4(45).

308. Ward N, Moreno-Hagelsieb G. Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? de Crécy-Lagard V, editor. PLoS One. 2014;9(7):e101850.

309. Darling AE, Mau B, Perna NT. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5(6):e11147.

310. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42(D1).

311. Duncan CL, Strong DH. Improved medium for sporulation of *Clostridium perfringens*. Appl Microbiol. 1968;16(1):82–9.

312. Korkeala HJ, Mäkelä PM, Suominen HL. Growth temperatures of ropy slime-producing lactic acid bacteria. J Food Prot. 1990;53(9).

313. Pitcher DG, Saunders NA, Owen RJ. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett Appl Microbiol. 1989;8(4):151–6.

314. Smyth G, Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, et al. LIMMA: linear models for microarray data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health. 2005.

315. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. Bioinformatics. 2007;23(20):2700–7.

316. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. TM4 microarray software suite. Methods Enzymol. 2006; 411:134–93.

317. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12(2):115–21.

318. Gaidatzis D, Lerch A, Hahne F, Stadler MB. QuasR: quantification and annotation of short reads in R. Bioinformatics. 2015;31(7):1130–2.

319. Johnson SL, Daligault HE, Davenport KW, Jaissle J, Frey KG, Ladner JT, et al. Thirty-two complete genome assemblies of nine *Yersinia* species, including *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*. Genome Announc. 2016;3(2).

320. Hardcastle TJ, Kelly KA. BaySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics. 2010;11.

321. Paredes CJ, Alsaker K V., Papoutsakis ET. A comparative genomic view of clostridial sporulation and physiology. Nat Rev Microbiol 2005 312. 2005;3(12):969–78.

322. Raphael BH, Lautenschlager M, Kalb SR, de Jong LIT, Frace M, Lúquez C, et al. Analysis of a unique *Clostridium botulinum* strain from the Southern hemisphere producing a novel type E botulinum neurotoxin subtype. BMC Microbiol. 2012;12(1):1–10.

323. Palonen E, Lindström M, Korkeala H. Adaptation of enteropathogenic *Yersinia* to low growth temperature. Crit Rev Microbiol. 2010;36(1):54–6.

324. Siddiqui KS, Cavicchioli R. Cold-adapted enzymes. Annu Rev Biochem. 2006;75:403–33.

325. Dimitriu T, Marchant L, Buckling A, Raymond B. Bacteria from natural populations transfer plasmids mostly towards their kin. Proc R Soc B Biol Sci. 2019;286(1905).

326. Xiao Y, Wagendorp A, Moezelaar R, Abee T, Wells-Bennik MHJ. A wide variety of *Clostridium perfringens* type a food-borne isolates that carry a chromosomal cpe gene belong to one multilocus sequence typing cluster. Appl Environ Microbiol. 2012;78(19):7060–8.

327. Murros-Kontiainen A, Johansson P, Niskanen T, Fredriksson-Ahomaa M, Korkeala H, Björkroth J. *Yersinia pekkanenii* sp. nov. Int J Syst Evol Microbiol. 2011;61(Pt 10):2363–7.

328. Schaake J, Kronshage M, Uliczka F, Rohde M, Knuuti T, Strauch E, et al. Human and animal isolates of *Yersinia enterocolitica* show significant serotype-specific colonization and host-specific immune defense properties. Infect Immun. 2013;81(11):4013–25.

329. Uliczka F, Pisano F, Schaake J, Stolz T, Rohde M, Fruth A, et al. Unique Cell Adhesion and Invasion Properties of *Yersinia enterocolitica* O:3, the Most Frequent Cause of Human Yersiniosis. Galán JE, editor. PLoS Pathog. 2011;7(7):e1002117.

330. Valentin-Weigand P, Heesemann J, Dersch P. Unique virulence properties of *Yersinia enterocolitica* O:3 – An emerging zoonotic pathogen using pigs as preferred reservoir host. Int J Med Microbiol. 2014;304(7):824–34.

331. Rakin A, Batzilla J, Garzetti D, Heesemann J. Gains and Losses in *Yersinia enterocolitica* subsp. *palearctica* Genomes. Adv Exp Med Biol. 2012; 954:23–9.

332. Schmühl C, Beckstette M, Heroven AK, Bunk B, Spröer C, McNally A, et al. Comparative Transcriptomic Profiling of *Yersinia enterocolitica* O:3 and O:8 Reveals Major Expression Differences of Fitness- and Virulence-Relevant Genes Indicating Ecological Separation. mSystems. 2019;4(2).

333. Uliczka F, Dersch P. Unique Virulence Properties of *Yersinia enterocolitica* O:3. Adv Exp Med Biol. 2012;954:281–7.

334. Ortiz Martínez P, Fredriksson-Ahomaa M, Sokolova Y, Roasto M, Berzins A, Korkeala H. Prevalence of Enteropathogenic *Yersinia* in Estonian, Latvian, and Russian (Leningrad Region) Pigs. Foodborne pathogens and disease, 2009, 6.6: 719-724.

335. Wouters JA, Kamphuis HH, Hugenholtz J, Kuipers OP, De Vos WM, Abee T. Changes in glycolytic activity of *Lactococcus lactis* induced by low temperature. Appl Environ Microbiol. 2000;66(9):3686–91.

336. Monedero V, Mazé A, Boël G, Zúñiga M, Beaufils S, Hartke A, et al. The Phosphotransferase System of *Lactobacillus casei*: Regulation of Carbon Metabolism and Connection to Cold Shock Response. Microb Physiol. 2007;12(1–2):20–32