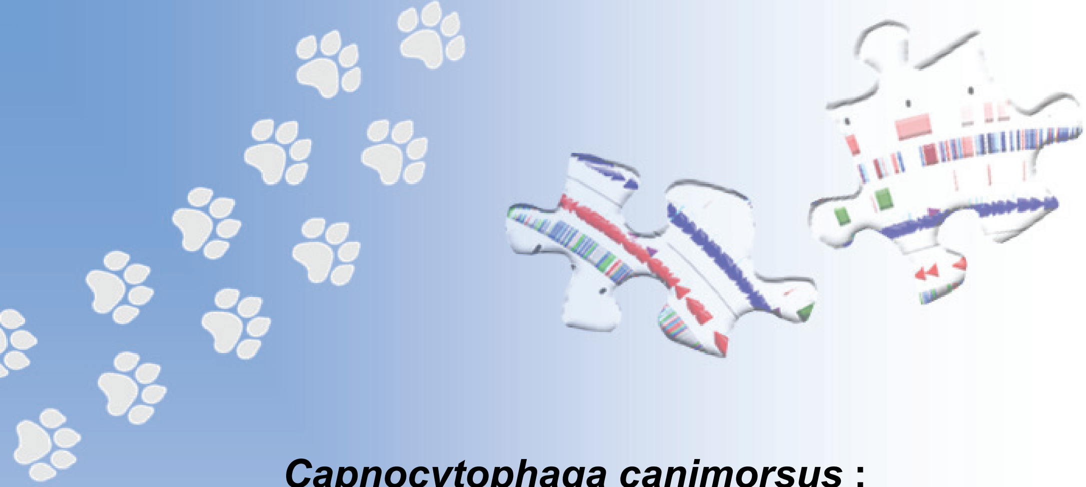
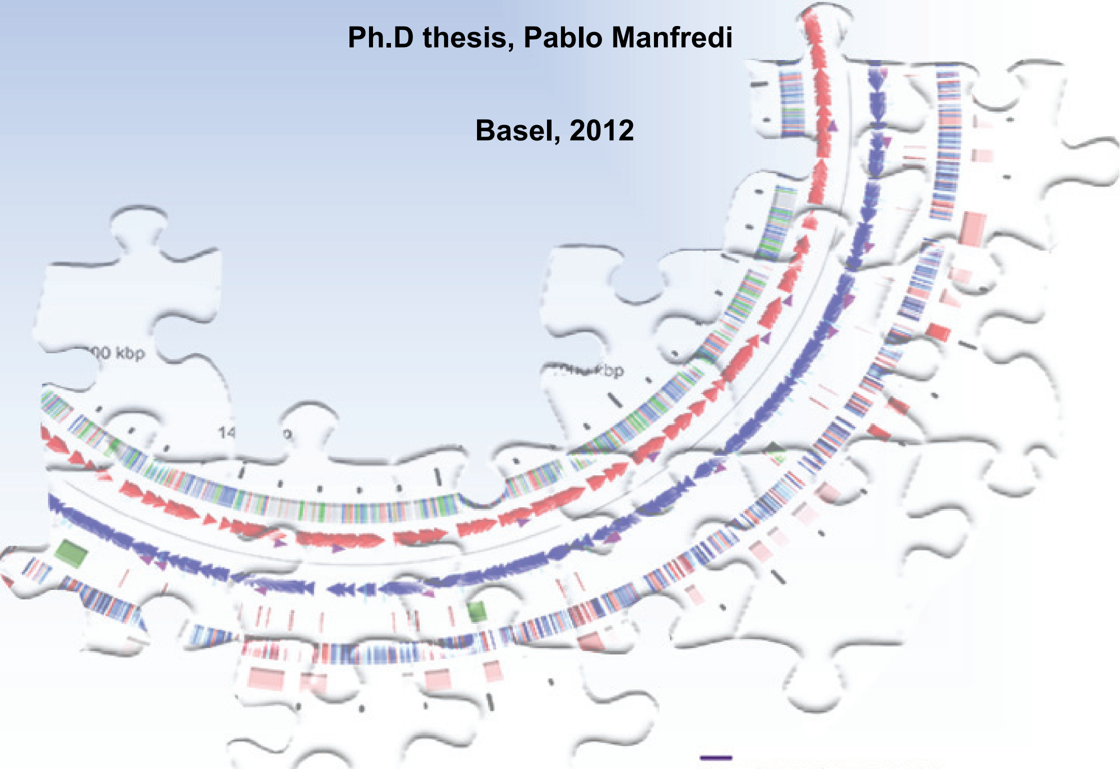


Capnocytophaga canimorsus :

Genomic characterization of a specialised host-dependent lifestyle and implications in pathogenesis

Ph.D thesis, Pablo Manfredi

Basel, 2012



BIOZENTRUM

Universität Basel
The Center for Molecular Life Sciences

Capnocytophaga canimorsus:

Genomic characterization of a specialised host-dependent lifestyle and implications in pathogenesis

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

von
Pablo Manfredi
aus Toulouse, France

Basel, March 2012

UNI BASEL
FOCAL AREA
INFECTION BIOLOGY
BIOZENTRUM
Universität Basel

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von :

Prof. Dr. G. R. Cornelis

Prof. Dr. C. Dehio

Basel, den 22. Februar 2011

Prof. Dr. M. Spiess, Dekan.

UNI BASEL

FOCAL AREA
INFECTION BIOLOGY

BIOZENTRUM
Universität Basel

UNI
BASEL

FOCAL AREA
INFECTION BIOLOGY

BIOZENTRUM
Universität Basel

Cover page : the “Cc5 genome jigsaw” by Pablo Manfredi.

FOCAL AREA
INFECTION BIOLOGY

BIOZENTRUM
Universität Basel

UNI
BASEL

1. Content

1. Contents **5**
2. Summary **7**
3. Introduction **11**
3.1. Capnocytophaga canimorsus 12
3.2. The *Capnocytophaga genus* 13
3.3. *C. canimorsus* is member of the *Bacteroidetes* phylum 14
3.4. *C. canimorsus* is a canine oral commensal 18
3.5. *C. canimorsus 5* and the molecular basis of its way of life 23
4. Genomics of *C. canimorsus* 5 **25**
4.1. Submitted manuscript 26
4.2. Additional data 61
4.2.1. Genome assembly and restriction fragment profile 61
4.2.2. Semi automated genome annotation pipeline 62
4.2.3. Genome scale metabolic modeling 65
4.2.4. Genomic codon usage analysis 68
5. The polysaccharide utilization locus 5 **71**
5.1. Publication 72
5.2. Additional data: New promoters with diverse expression levels 99
6. Additional unpublished data **101**
6.1. Additional genomes sequencing 102
6.2. Genomes sequencing and Assembly 102
6.2.1. Development of a microreads assembly pipe 104
6.2.2. Preliminary hybrid assembly of the *Cc2*, *Cc11* and *Cc12* 108
6.3. Genomics of *Capnocytophaga* 110
6.3.1. Mapping of the X2E4 transposon mutant 112
6.3.2. Genomics of *C. canimorsus* 113
6.3.3. C. canimorsus a dog commensal & an opportunistic pathogen? 114
6.3.4. C. canimorsus and O_2 utilization 118
6.3.5. Taxonomic conservation of polysaccharide utilization loci 119
7. Conclusions and perspectives **123**
8. References **127**
9. Appendix **137**
10. Acknowledgments **143**
11. *Curriculum vitae* **145**
12. Supplementary data **149**

2. Summary

Here is presented the complete 2,571,405-bp genome sequence of *Capnocytophaga canimorsus* strain 5 (*Cc5)*, a strain that was isolated from a fatal septicaemia. Phylogenetic analysis of conserved genes supports the inclusion of *C. canimorsus* into the *Cytophaga-Flavobacteria-Bacteroides* (CFB) phylum and indicates close relationships with environmental *flavobacteria* as *Flavobacterium johnsoniae* and *Gramella forsetii*. In addition, relative phylogenetic topology of *Capnocytophaga* species shows that *C. canimorsus* share more sequence similarities with human host associated *Capnocytophaga* species than species from the latter group among themselves (*e.g. C. gingivalis* and *C. ochracea*).

As compared to other *Capnocytophaga*, *C.canimorsus* seems to have differentiated by large-scale horizontal gene transfer compensated by gene losses. Consistently with a relatively reduced genome size, genome scale metabolic modelling suggested a reduced global pleiotropy as it is illustrated by the presence of a split TCA cycle or by the metabolic uncoupling of the hexoses and N-acetylhexosamines pathways. In addition and in agreement with the high content in HCO_3^- and Na^+ ions in saliva, we predicted a CO_2-dependent fumarate respiration coupled to a Na^+ ions gradient based respiratory chain in *Cc5*. All together these observations draw the picture of an organism with a high degree of specialization to a relatively homeostatic host environment.

Unexpectedly, the genome of *Cc5* did not encode classical complex virulence functions as T3SSs or T4SSs. However it exhibits a very high relative number of predicted surface-exposed lipoproteins. Many of them are encoded within 13 different putative polysaccharide utilization loci (PULs), a hallmark of the CFB group, discovered in the gut commensal *Bacteroides thetaiotaomicron*. When *Cc5* bacteria were grown on Hek293 cells, at least 12 PULs were expressed and detected by mass spectrometry. Semi-quantitative analysis of the *Cc5* surfome identified 73 surface exposed proteins among which 40 were lipoproteins and accounted for 76% of the total quantification. Interestingly, 28 proteins (38%) were encoded by 9 different PULs and corresponded to more than 54% of total MS-flying peptides detected. A systematic knockout analysis of the 13 PULs revealed that 6 PULs are involved in growth during cell culture infections with most dramatic effect observed for ΔPUL5. Proteins encoded by PUL5, one of the most abundant PULs (12%), turned out to be devoted to foraging glycans from N-linked glycoproteins as fetuin but also IgG. It was not only essential for growth on

cells but also for survival in mice and in fresh human serum therefore representing a new type of virulence factor.

Further characterization of the PUL5 deglycosylation mechanism revealed that deglycosylation is achieved by a large surface complex spanning the outer membrane and consisting of five PUL5 encoded Gpd proteins and the Siac sialidase. GpdCDEF contribute to the binding of glycoproteins at the bacterial surface while GpdG is a β-endo-glycosidase cleaving the N-linked oligosaccharide after the first N-linked GlcNAc residue. We demonstrate that GpdD, -G, -E and -F are surface-exposed outer membrane lipoproteins while GpdC resembles a TonB-dependent OM transporter and presumably imports oligosaccharides into the periplasm after cleavage from glycoproteins. Terminal sialic acid residues of the oligosaccharide are then removed by SiaC in the periplasm. Finally, degradation of the oligosaccharide proceeds sequentially from the desialylated non reducing end by the action of periplasmic exoglycosidases, including β-galactosidases, β-N-Acetylhexosaminidases and α-mannosidases.

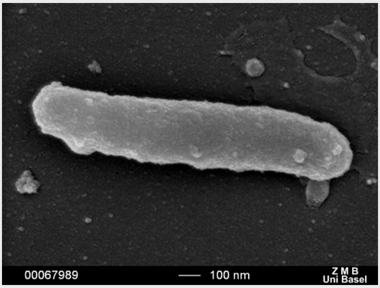
Genome sequencing of additional *C. canimorsus* strains have been performed with the only use of second generation sequencing methods (Solexa and 454). Two assembling approaches were developed in order to enhance assembly capacities of pre-existing tools. Draft assemblies of the three pathogenic human blood isolates *C. canimorsus 2* (three contigs), *C. canimorsus 11* (152 contigs) and *C. canimorsus 12* (63 contigs) are presented here. Comparative genomics including genomes of four available human hosted *Capnocytophaga* species stressed *C. canimorsus* exclusively conserved features as an oxidative respiratory chain and an oxidative stress resistance or the presence of a *Cc5* specific PULs content. Therefore we propose these features as potential factors involved in the pathogenesis of *C. canimorsus*.

Pablo Manfredi

3. Introduction

3.1. *Capnocytophaga canimorsus*

Figure 3.1 ***C. canimorsus***



SEM of a thin Rod-shape *C. canimorsus strain 5* (*Cc5*). (Chantal fitcher, 2007)

Capnocytophaga canimorsus (**Figure 3.1**), formerly *DF-2* (*dysgogenic fermentator 2*), is a fastidious Gram negative commensal bacterium from the normal canine oral flora. It is responsible for rare but life-threatening zoonoses that occur after close contact with dogs (91%) and cats (9%) with a higher frequency for bites (54%), scratches (8.5%) or simply licks [1]. Such infections can lead to affections ranging from very mild flu like symptoms to fulminant sepsis potentially leading to multiple organ failure ([2] and [3]). Alternatively and in a minority of cases, meningitis, endocarditis or myocarditis can be observed. Fastidious growth of the pathogen and lack of symptoms during the initial stages of infection often lead to unattended wound [4]. Mortality is highest in case of sepsis (30%) [1], while it only reaches 5% for meningitis [5]. Reported predisposing factors are splenectomy (33% of sepsis cases), alcohol abuse (24%) or other immunosuppression (5%) but 41% of the patients do not show any other obvious risk factors [1].

C. canimorsus has first been described in 1976 [6] and assigned to the *Capnocytophaga* genus in 1989 [7]. Since then, it is regularly isolated from dog or cat bite infections [8]. Nowadays, *C. canimorsus* infections are well known by clinicians and more than 200 cases have been reported so far [9]. Apparent *C. canimorsus* infection incidence in Denmark encloses 1 case annually per million [3]. However several reasons would explain a significant underestimation of the factual infection frequency: 1) Systematic prophylactic antibiotic treatments after most categories of bites related injuries [10]; 2) sensitivity of *C. canimorsus* to most widely used antibiotics [11]; 3) an extended and variable incubation period (from 5 to 15 days) [5] with a large range of symptoms [3] [12]; 4) And fastidious growth specially in inappropriate routinely used blood culture conditions [12]. It is likely that generalization of clinical nucleotide sequence determination methods will afford a better assessment of the *C. canimorsus* infection incidence [13].

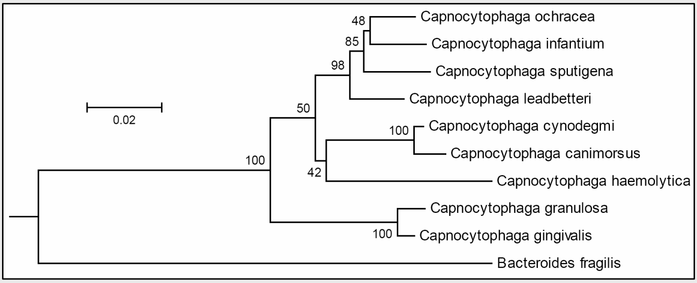
3.2. The *Capnocytophaga genus*

The *Capnocytophaga* genus exclusively includes a variety of fusiform commensals found in the oral flora of humans and other mammalians (**Figure 3.2**). Often co-isolated with *C. canimorsus, Capnocytophaga cynodegmi* (*DF-2 like*) is also found in dogs and cats with a significantly higher prevalence [13]. It occasionally leads to local wound infections in humans and animals with no obvious predisposing factors [7]. Seven *Capnocytophaga* species (formerly *DF-1* group) are found in humans (*Capnocytophaga ochracea, Capnocytophaga sputigena, Capnocytophaga gingivalis, Capnocytophaga haemolytica, Capnocytophaga granulosa, Capnocytophaga infantium*, *Capnocytophaga leadbetteri*) [14]. Human Infections with human-associated *Capnocytophaga* species are extremely rare and only few cases have been reported mostly in immunocompromised patients [15-21].

The *Capnocytophaga* genus has first been thoroughly characterized in 1979 [22-25]. It forms a functionally homogeneous taxon of capnophilic (greek: carbon dioxide (καπνος : smoke) loving), gliding, strict fermentators [7]. These bacteria are able to grow in aerobic or anaerobic conditions provided an elevated level of carbon dioxide is present (5-10% v/v). They are positive to the benzidine assay suggesting presence of iron-porphyrin compounds as cytochromes or other particular respiratory chain components. Acetate and succinate are the major or sole metabolic end products. G+C contents are rather low and range from 33-41%.

Figure 3.2 Phylogenetics of the *Capnocytophaga* genus

Type strains 16S rRNA phylogenetic tree using the Weighbor weighted neighbor-joining algorithm. Bootstrap values are represented on their corresponding nodes; branch length is scaled in terms of mutation rate per site.



http://rdp.cme.msu.edu/treebuilder/viewer.spr

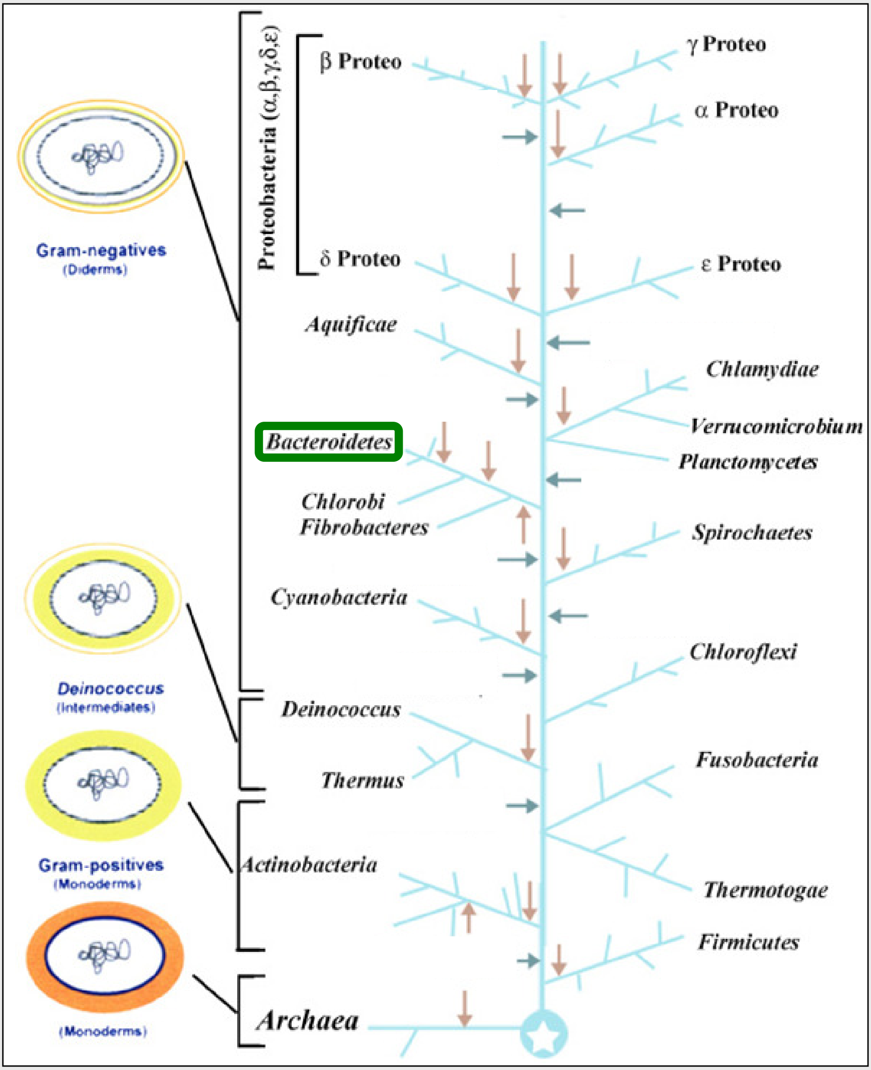
3.3. *C. canimorsus* is member of the *Bacteroidetes* phylum

Capnocytophaga belong to the *Flavobacteriaceae* family from the *Bacteroidetes* phylum. *Bacteroidetes* are remotely related to *Proteobacteria* and to most commonly studied human pathogens. They are taxonomically close to the environmental aquatic phylum *Chlorobi* (Green sulfur bacteria) and to the major rumen commensals *Fibrobacteres* (**Figure 3.3**). *Bacteroidetes* phylum currently ramifies into *Bacteroidia*, *Sphingobacteria*, *Flavobacteria* and *Cytophagia* classes. So far, only 34 *bacteroidetes* have their chromosome(s) completely sequenced (**Table 3.3**).

The phylum exhibits a wide range of habitats and includes free-living and host-associated organisms. Several extremophiles belong to this phylum , for example the thermohalophilic and halophilic *Rhodothermus marinus* that colonize very narrow zones around submarine hot springs [26], the psychrophilic (or cryophilic) *Flavobacteriaceae 3519-10* isolated in Antarctica from deep glacial ice that is able to grow at -8 ℃ by both producing an ice-binding protein and an ice recrystallization inhibitor [27], or the hyperhalophilic *Salinibacter rubber* from saltern crystallizer ponds whose proteins make up has adapted to strong ionic conditions [28]. Nevertheless, *Bacteroidetes* are not restricted to hyperspecialized niches and several ubiquitous environmental organisms are commonly found in soil and freshwater like *Flavobacterium johnsoniae,* the main model system for studies of gliding motility [29] or the pleomorphic *Spirosoma linguale* originally isolated from a laboratory water bath [30]. Host associated *Bacteroidetes* also display strong diversity. Several arthropods and protists endosymbionts have been described among *Bacteroidetes* to date. For instance, the *Blattabacterium spp.* (*Flavobacteriales*) are maternally inherited major endosymbiont of the cockroach and thought to support metabolic nitrogen recycling [31, 32], the N_2-fixing endosymbiont *Azobacteroides pseudotrichonymphae* (*Bacteroidales*) lives in the termite's gut protist P*seudotrichonympha grassii*'s, ensures optimal lignocellulose fermentation and prevents nitrogen deficiencies [33], another example is *Amoebophilus asiaticus*, an obligate endoparasite of the free living *Acanthamoeba sp*. [34].

Figure 3.3 Prokaryotic Phylogeny Webpage (April 2007).

Large DNA Insertion / deletion events (blue and brown arrows) are of high interest in phylogeny determination. **(http://www.bacterialphylogeny.com/index.html)**



Extracellular host associated *Bacteroidetes* are by far the most studied organism of the phylum mainly because of the specialized relationship they share with human hosts. *Bacteroides spp.* are dominant members of the major human microflora community, the colonic microbiota (*e.g. Bacteroides fragilis* [35], *B. thetaiotaomicron* [36], *B. vulgatus and B. distasonis* [37]). They are also considered as opportunistic pathogen as they can severely limit the success of gastro-intestinal surgery, and are repeatedly been associated with extraintestinal infections in animals and humans. Specialized pathogens among *Bacteroidetes* have also been reported and are of high interest in odontology like the highly proteolytic *Porphyromonas gingivalis* that initiates periodontal disease, one of the most frequently occurring infectious diseases in humans [38]. Other members of this phylum, particularly from the *Flavobacteriaceae* family (as *C. canimorsus*), are also renowned for the damages they can cause in the zootechnical field. The worldwide respiratory avian pathogen *Ornithobacterium rhinotracheale* typically causes airsacculitis symptoms leading to millions of dollars losses to the poultry industry annually [39]. *Riemerella anatipestifer*, a contagious septicemia agent in various birds also accounts for major economic losses in industrialized duck production [40]. Another example is the facultative intracellular pathogen of trouts and salmons *Flavobacterium psychrophilum*. it is currently one of the most devastating fish pathogens due to horizontal and vertical transmission and to the gravity of symptoms it generates (septicemia and extensive necrotic lesions) [41].

Table 3.3 Completely sequenced genome within *Bacteroidetes*

Class	Genus	genomes	DNA source Isolation
Bacteroidia	*Bacteroides*	4	Human intestinal microflora
	Azobacteroides	1	Termite gut protest associated
	Parabacteroides	1	Human intestinal microflora
	Porphyromonas	2	Human oral microflora
	Prevotella	2	Cattle rumen flora / Human oral microflora
Cytophagia	*Dyadobacter*	1	Plant stems
	Spirosoma	1	laboratory water bath
Flavobacteria	*Blattabacterium*	2	cockroachs
	Capnocytophaga	1	Human oral microflora
	Croceibacter	1	Bermuda Atlantic
	Unknown Flavobacteriales	2	Antarctica subglacial lake / Coastal Pacific Ocean
	Flavobacterium	2	Soils & fresh waters / Salmon infection
	Gramella	1	Sea waters
	Robiginitalea	1	Sea waters
	Sulcia	4	sap-feeding insects
	Zunongwangia	1	deep-sea waters
Sphingobacteria	*Chitinophaga*	1	pine litter
	Cytophaga	1	soil
	Pedobacter	1	dry soil
	Rhodothermus	1	submarine hot springs, Iceland
	Salinibacter	2	saltern crystallizer pond
unclassified	*Amoebophilus*	1	Acanthamoeba sp.

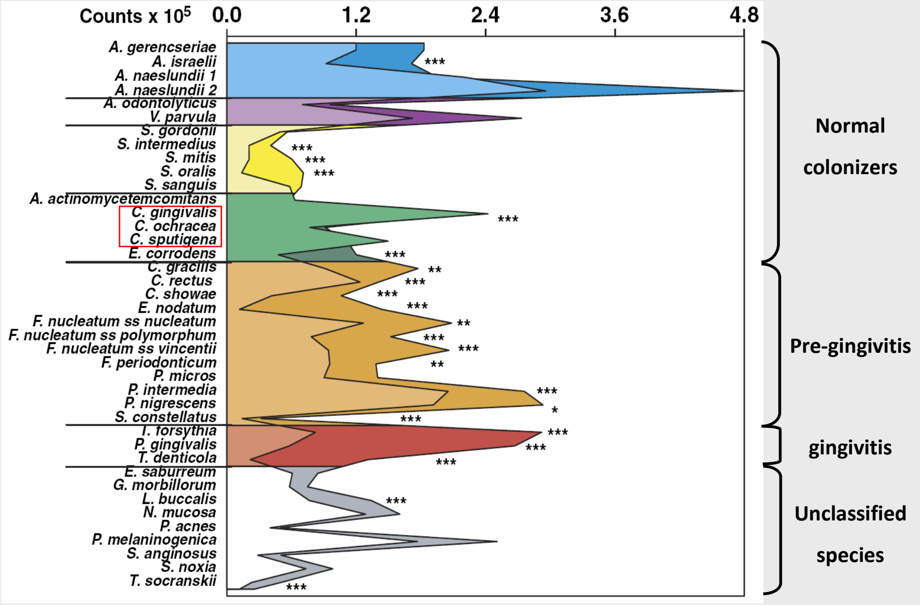
3.4. *C. canimorsus* is a canine oral commensal

Mammalians' oral cavity is a highly heterogeneous environment made of different tissular compartments that exhibits strong composition and structural differences (*e.g.* mucosa, dental surfaces, gingival epithelium, lingual surfaces, saliva, crevicular fluids…). Each microenvironment is colonized by a specifically associated microbial biofilm as the so called dental plaque that cover several oral surfaces including the dental enamel layer. However, despite such a micro-environmental diversity, microflora is not well compartmented in the oral cavity. Several attempts to identify microbial composition bias according to oral localization failed to define specific site associated bacterial communities and it is currently accepted that "everything is everywhere" [42, 43]. In total, human oral microbiota is composed of up to 700 bacterial phylotypes that alternatively become dominant according to the on going physiological state (*e.g.* gingivitis, tooth decay, early/late colonization stages or stable and self-sustained climax communities) [43, 44].

Characterization of the commensal way of life of *C. canimorsus* is crucial in the understanding process of the pathogenic events it can trigger when incidentally introduced into alternative mammalian hosts. Identification of preferentially colonized oral sites or host groups by *C. canimorsus* would be highly informative. It would then be possible to assess possible interactions (with host cells or other bacteria), substrates availability, and sustained immune pressure during commensalism with dogs or cats. However, canine and feline oral microbiology are poorly studied and only few works consider *Capnocytophaga* species in animals [45].

In contrast to *C. canimorsus* and *C. cynodegmi*, human hosted oral *Capnocytophaga* species (HCSs) benefit from sound investigation. *C. gingivalis*, *C. ochracea* and *C. sputigena* belong to the 8% of identified species that normally account for more than half of the total oral microbiota and are therefore considered in most polymicrobial studies [43]. The most obvious feature emerging from literature is an apparent tropism of HCSs for inflammation sites (*i.e.* bacteria is more abundant at gingivitis or periodontitis sites) but this is also observed for the vast majority of oral bacteria [43]. In contrast to suspected periodontal pathogens and most normal colonizers, HCSs have been shown to be significantly more prevalent and abundant in periodontally healthy persons compare to individual exhibiting periodontitis (**Fig. 3.4.1**). Even more, their presence in the oral cavity correlates to lower risks of dental disease progression [43, 46-48].

Figure 3.4.1 Microbial profiles of healthy and periodontitis affected individuals

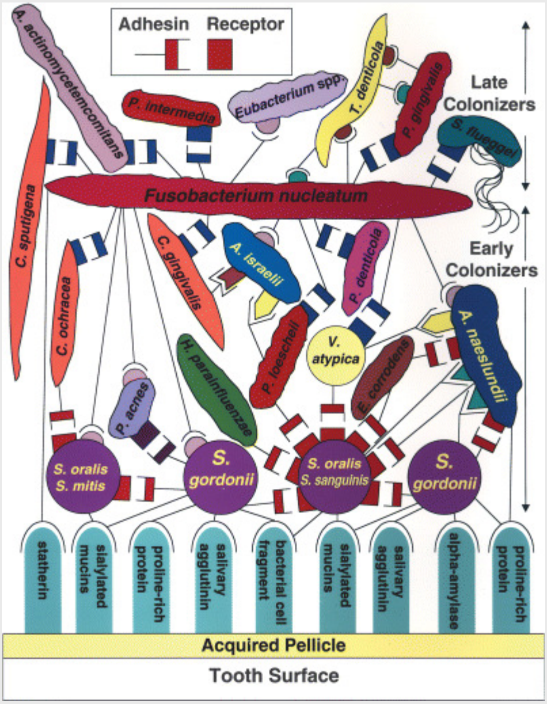


Adapted from Socransky & Haffajee [43]. Mean counts (×105) of 40 microbial taxa in subgingival plaque samples taken from 184 periodontally healthy and 592 subjects with chronic periodontitis. The species are color coded according to their role during biofilm formation and pathogenesis. The darker shade represents the periodontitis subjects, while the lighter shade represents the periodontally healthy subjects.

No direct interaction or adhesion to human host tissues have been observed for HCSs so far. In facts, most polymicrobial subgingival biofilm models consider HCSs as secondary colonizers because of their lectin-based capacity to coaggregates with several gram-positive primary colonizers (*e.g. Streptococcus sanguis, Actinomyces naeslundii, Actinomyces israelii*). HCSs are then thought to bridge primary colonizers to tertiary colonizing species as the highly congregating *Fusobacterium nucleatum* and late colonizers (*e.g. P. gingivalis* or *Treponema denticola*) (**Fig. 3.4.2**) [44].

As for most oral bacteria, several studies failed to find significant differences in HCSs abundances among normal oral compartments. The only repeatedly reported bias in HCSs oral distribution is the increasing bacterial abundance that correlates with increasing subgingival pockets depth [43]. Subgingival pocket is a very important oral compartment responsible for significant serum and leukocyte exchange between the oral cavity and subjacent tissular circulation. The so called crevicular fluid, a constitutive serum exudate, virtually fills the subgingival pocket [49]. Consequentially, primary and adaptative immunity is higher there than in any other oral sub-compartment indicating that immune pressure sustained by subgingivial inhabitants is substantial. Interestingly, this is consistent with the addition of blood or serum to growth media required by all *Capnocytophaga* species to achieve rapid growth [7, 22].

Figure 3.4.2 Model of oral bacterial colonization at the dental surface (Kolenbrander et *al*., 2006)



From the bottom, primary colonizers bind via adhesins (round black lines) to complementary salivary receptors (round-topped columns) in the acquired pellicle coating the tooth surface. Sequential bacterial binding results in the appearance of nascent surfaces that bridge with the next coaggregating partner. Rectangular symbols represent lactose-inhibitable coaggregations. Other symbols represent components that have no known inhibitor. The bacterial species first mentioned here are *Actinobacillus actinomycetemcomitans*, *Eikenella corrodens*, *Eubacterium spp.*, *Haemophilus parainfluenzae*, *Prevotella denticola*, *P. intermedia*, *P. loescheii*, *Propionibacterium acnes*, *Selenomonas flueggei*, *Streptococcus gordonii*, *Streptococcus mitis*, *Streptococcus oralis*, and *Veillonella atypica.*

It has been estimated by cultural means that more than every second dog carries *C. canimorsus* in its normal oral flora [50]. Because of the very specific cultural conditions required by *C. canimorsus* strains, prevalence has often been underestimated in previous studies. A recent study using a PCR based method reported up to 74% of dogs carrying *C. canimorsus* in their mouth [13].

Correlation between *C. canimorsus* occurrence and different host factors as lifestyle, health, breed or species have been poorly investigated. A recent study describes a higher occurrence of *C. canimorsus* among small breeds and male or neutered dogs [45]. A few studies reported other oral hosts than dogs and cats. So far, *C. canimorsus* has only been isolated from nutrition specialized mammalian species as carnivores or herbivores where amylase activity and dental decay are hardly observed (**Table 3.4**) [51-54]. One could speculate that *Capnocytophaga* species benefit from a relative independence from host diet uptake as it could be in the case of individuals with good dental hygiene practices or for carnivores that quickly swallow their food without any oral preprocessing. In other hosts, nutrient particles remaining in the oral cavity would support a different microbiotic profile and an increased competition for *Capnocytophaga* species.

Table 3.4 Occurrence of cultured *C.canimorsus* in mammalian oral cavities

(Mally *et al.*, 2009; Blanche *et al.*, 1998, Westwell *et al.*, 1989, HJ Lipner 1947 and Chauncey *et al.*, 1963.)

Species	**Dog**	**Cat**	**Human**	**Pig**	**Rodents**	**Sheep**	**Cattle**
effectives	**376**	**369**	**140**	**13**	**35***	**12**	**15**
***C.canimorsus* positive cultures**	**128**	**60**	**0**	**0**	**0***	**3**	**5**
Amylase activity	**0**	**0**	**+++**	**++**	**+++***	**0**	**0**

* extrapolated from *C.canimorsus* counts in Hamsters and amylase activity from Rabbits and Rats

3.5. *C. canimorsus 5* and the molecular basis of its way of life

Capnocytophaga canimorsus 5 (*Cc5*) is a strain isolated from a patient with fatal septicemia and is used as a model to understand the high aggressiveness of *C. canimorsus* for humans. Recently, a number of clues showed that *Cc5* does not exhibit any strong pathogen-associated molecular pattern. Unusual features of its LPS provide *C. canimorsus* with the capacity to resist to killing by human complement as well as to escape phagocytosis by human polymorphonuclear leukocytes (PMNs) [55]. They are also able to evade detection and phagocytosis by macrophages, which results in a lack of release of pro-inflammatory cytokines [56]. Despite such anti-inflammatory mechanisms, *C. canimorsus* are not cytotoxic for macrophages and it has even been shown that they remain undetected by Toll like receptor 4 (TLR4) analogs. In addition to this passive evasion from innate immunity, some strains, including *Cc5*, even actively block macrophage’s pro-inflammatory responses: Despite stimulation by an endotoxic *Escherichia coli* lipopolysaccharide (LPS), macrophages fail to release nitric oxide (NO), TNF and other cytokines if they are pre-incubated in presence of *C. canimorsus* [56]. Moreover, when challenged by *Escherichia coli*, these macrophages can no longer kill phagocytosed *E. coli* [57]. The molecular bases of these active immunosuppressive mechanisms are not understood yet. However, their study led to the serendipitous discovery that the fastidious *C. canimorsus* grow readily upon direct contact with mammalian cells including phagocytes. This property was found to be dependent on a peripheral sialidase allowing *C. canimorsus* to harvest amino sugars of glycan chains from host cell glycoproteins [58]. Interestingly, sialidase was also found to contribute to bacterial persistence in a murine infection model [58]. Thus, the feeding system that *C. canimorsus* evolved in its extremely competitive niche -the canine oral cavity-, could be seen as an essential virulence factor.

Despite extended characterization of *C. canimorsus* behavior in presence of diverse mammalian cells, molecular mechanisms of host interaction involved in bacterial growth and in immunity control remains poorly understood. To this purpose, deciphering of the *C. canimorsus* genome consequently became a priority. This thesis describes assembly, annotation and analysis of the *Cc5* genome and follow-up experiments enabling further understanding of the *C. canimorsus* life style.

4. Genomics of *C. canimorsus* 5

The *Capnocytophaga canimorsus* Genome and Surfome reveal a key role of glycan foraging systems in its specialized host-dependent lifestyle.

4.1. Submitted manuscript

The *Capnocytophaga canimorsus* Genome and Surfome reveal a key role of glycan foraging systems in its specialised host-dependent lifestyle

Pablo Manfredi[1], Francesco Renzi[1], Manuela Mally[1], Loïc Sauteur[1], Hwain Shin[1], Simon Ittig[1], Cécile Pfaff[1], Mathias Schmaler[2], Suzette Moes[1], Martin Gentner[1], Stephan Grzesiek[1], Paul Jenö[1], Marco Pagni[3], Guy R Cornelis[1]†

1.Biozentrum der Universität Basel, Basel, Switzerland
2. Department of Biomedicine, University Hospital Basel, Basel, Switzerland
3. Swiss Institute of Bioinformatics, Lausanne, Switzerland
† Corresponding author

Statement of authors' work.

PM performed all genomic and bioinformatics analysis with the support of MP and HS.

FR performed surface proteins identification with **PM's** assistance for data analysis.

SM and PJ performed all mass spectrometry experiments.

PM, LS and FR performed mutagenesis with the technical assistance of CP.

LS performed cell culture growth assays and serum sensitivity assays with the support of **PM** and MM.

FR performed fetuin deglycosylation assays.

MS and **PM** performed mice infection experiment with the help of FR, LS and SI.

NMR analysis of *C. canimorsus* culture supernatant has been done by SI, MG and SG.

ABSTRACT

Capnocytophaga canimorsus are commensal Gram-negative bacteria from dog's mouth that cause rare but dramatic septicaemia in humans. *C. canimorsus* escape innate immune defenses and have the unusual property to feed on cultured mammalian cells, including phagocytes. Here we present the complete 2,571,405-bp genome sequence and the surface proteome of strain *Cc5*. Genome analysis highlighted a close relationship between *Capnocytophaga* and *Flavobacteria* among *Bacteroidetes*. Functional annotation and metabolic modeling consistently reflect adaptation to the canine oral environment. The genome of *Cc5* does not encode any classical complex virulence system but a very high relative number of lipoproteins. Many of these belong to 13 surface exposed feeding complexes encoded by polysaccharide utilization loci (PULs), a hallmark of the *Flavobacteria-Bacteroides* group. When *Cc5* bacteria were grown on Hek293 cells, at least 12 PULs were expressed and their products represented more than half of the total peptides from the surface proteome. Systematic mutagenesis revealed that half of these complexes contributed to growth on cells. The complex encoded by PUL5, one of the most abundant ones, turned out to be devoted to foraging glycans from N-linked glycoproteins. It was not only essential for growth on cells but also for survival in mice and in fresh human serum. It thus represents a new type of virulence factor.

Author Summary

Capnocytophaga canimorsus are Gram-negative commensal bacteria from the oral flora of dogs and cats, which cause rare but severe infections in humans that have been bitten or simply licked by a dog/cat. Fulminant septicemia and peripheral gangrene are the most common syndromes. Here we present the first genome sequence of a *C. canimorsus* strain and we analyze the proteins anchored at the bacterial surface. The genome analysis underlines the proximity of *C. canimorsus* with *Bacteroides spp*, the main commensals of the human colon, and also with *Flavobacteria*, saprophytes from aquatic environments. Like the others, *C. canimorsus* are dedicated glycophile bacteria. Indeed, we identified 13 surface-exposed protein complexes specialized in foraging diverse polysaccharides and complex glycosides. One of them, abundant at the bacterial surface, turned out to be devoted to the harvest of host glycoproteins. Although its main function must be to sustain commensalism in dog's mouth, we show that it may also contribute to human pathogenesis.

INTRODUCTION

Capnocytophaga canimorsus, formerly *dysgonic fermentor 2 (DF-2),* is a non-haemolytic Gram negative commensal bacterium from dog's mouth responsible for rare but life-threatening zoonoses. The genus *Capnocytophaga* belongs to the phylum *Bacteroidetes,* family of *Flavobacteriaceae*. It includes a variety of commensals found in the oral flora of mammalians. *C. canimorsus* are found in dogs and cats while *Capnocytophaga gingivalis*, *ochracea* and *sputigena* are found in human mouth [7, 14]. Human infections by *C. canimorsus* occur after dog bites, scratches or simply licks. They generally appear as fulminant septicaemia, peripheral gangrene or meningitis, with mortality as high as 40 % [3, 5]. A few recent observations help understanding the high aggressiveness of *C. canimorsus* for humans. First, *C. canimorsus* are able to escape complement killing and opsonization and hence to avoid phagocytosis by human polymorphonuclear leukocytes (PMN's)[55]. They also escape detection and phagocytosis by macrophages, which results in a lack of release of pro-inflammatory cytokines [56]. In addition to this passive evasion from innate immunity, some strains even actively block the onset of pro-inflammatory signalling induced by an *Escherichia coli* lipopolysaccharide (LPS) stimulus [56] and are able to block the killing of phagocytosed *E. coli* by macrophages [57]. The molecular bases of these active immunosuppressive mechanisms are not understood yet. However, their study led to the serendipitous discovery that the fastidious *C. canimorsus* grow readily upon direct contact with mammalian cells including phagocytes. This property was found to be dependent on a sialidase allowing *C. canimorsus* to harvest amino sugars of glycan chains from host cell glycoproteins [58]. Interestingly, sialidase was also found to contribute to bacterial persistence in a murine infection model [58]. Thus, the feeding system that *C. canimorsus* evolved in its extremely competitive niche -the canine oral cavity-, could be seen as an essential virulence factor.

Here, we report the first complete genome sequence and the surface proteome of a *C. canimorsus* strain. These analyses revealed the presence of 13 putative surface exposed polysaccharide utilization systems, typical of the

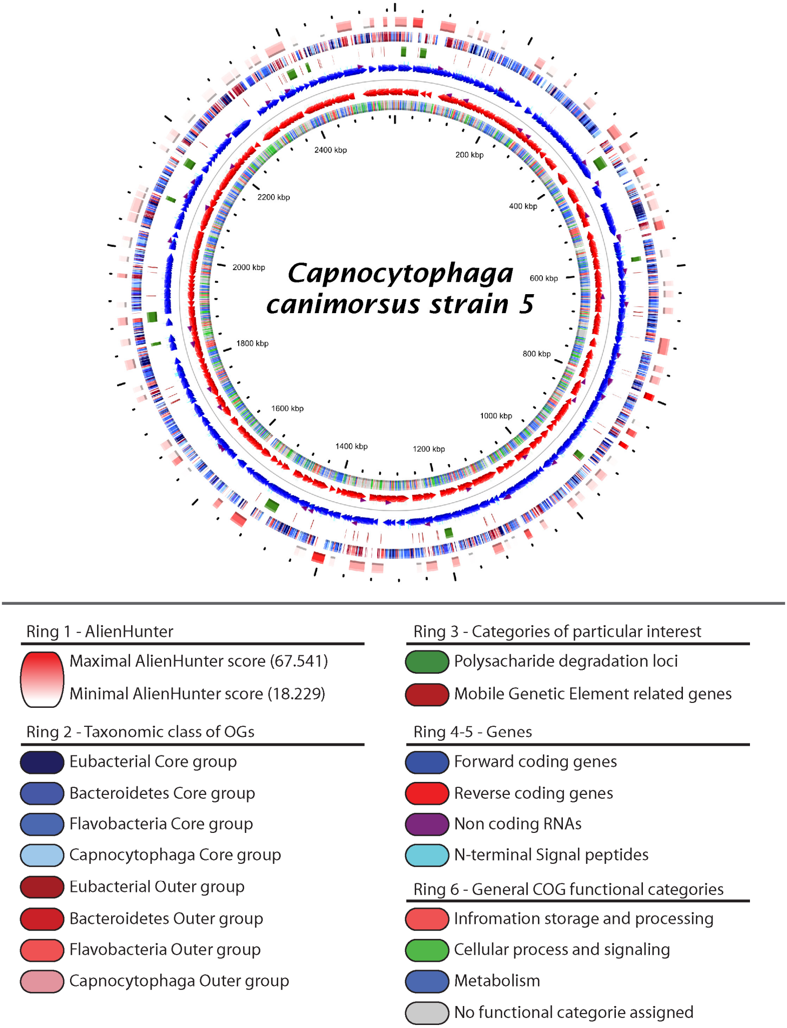
Cytophaga-Flavobacteria-Bacteroides group. Through systematic deletion mutagenesis of the 13 polysaccharide utilisation loci (PULs), we identified a PUL essential for glycoprotein deglycosylation, growth on mammalian cells, growth in human serum and persistence in the mouse. To our knowledge, this is the first report of a coherent foraging system specialized in N-linked surface glycoproteins deglycosylation. It also provides the first evidence that such a foraging system could be a virulence factor.

RESULTS

General Genome features

The genome of *Cc*5 consists of a single circular replicon of 2,571,405 bp with a G+C content of 36.11% (CP002113)(**Fig 4.1.1**). No plasmid was detected during assembly. In total, 2,414 coding sequences (CDSs) were identified, with 1,364 coding for proteins with high similarity to proteins in the non-redundant database (**Table S4.1**). This genome size is similar to those of *C. gingivalis* (NZ_ACLQ00000000, 2.66 Mb, 65 contigs), *C. sputigena* (NZ_ABZV00000000, 3.00 Mb, 37 contigs) and *C. ochracea* (NC_013162, 2.6 Mb, complete genome)[59]. As compared to genomes of other members of the *Bacteroidetes* phylum, such as the 6.1 Mb genome of the free living *Flavobacterium johnsoniae* [60], the 6.25 Mb genome of the commensal *Bacteroides thetaiotaomicron* [36] and the 5.3 Mb genome of *Bacteroides fragilis* [35], the *C. canimorsus* genome is thus rather small but it is still larger than that of *Porphyromonas gingivalis* (2.3 Mb)[38]. The genome encodes 46 tRNAs, three sets of ribosomal RNA genes, and 6 additional non-coding RNAs (an RNaseP, two tmRNAs, a TPP riboswitch, an SRP and one single CRISPR sequence)(**Table S4.1**).

Figure 4.1.1. Circular map of the *Cc5* genome.
From the most outer to the most inner ring (1 to 6). 1) White to red gradient indicates Alien Hunter scores above threshold (ranging from 18.229 to 67.541). 2) Taxonomic class of the cluster of orthologs established during this study. 3) PULs (green) and IS related elements (red). 4-5) Forward strand CDSs (blue), reverse strand CDS (red) and ncRNAs (purple). 6) Color coded COG functional categories.


200 kbp
400 kbp
600 kbp
800 kbp
1000 kbp
1200 kbp
1400 kbp
1600 kbp
1800 kbp
2000 kbp
2200 kbp
2400 kbp
Capnocytophaga canimorsus strain 5
Ring 1 - AlienHunter
Maximal AlienHunter score (67.541)
Minimal AlienHunter score (18.229)
Ring 2 - Taxonomic class of OGs
Eubacterial Core group
Bacteroidetes Core group
Flavobacteria Core group
Capnocytophaga Core group
Eubacterial Outer group
Bacteroidetes Outer group
Flavobacteria Outer group
Capnocytophaga Outer group
Ring 3 - Categories of particular interest
Polysacharide degradation loci
Mobile Genetic Element related genes
Ring 4-5 - Genes
Forward coding genes
Reverse coding genes
Non coding RNAs
N-terminal Signal peptides
Ring 6 - General COG functional categories
Infromation storage and processing
Cellular process and signaling
Metabolism
No functional categorie assigned

Comparison of the *Cc5* genome with 13 genomes from the *Bacteroidetes* phylum and two genomes from the *proteobacteria* phylum (*Escherichia coli* and *N. meningitidis*) (**Fig 4.1.2**) defined a set of 243 orthologous groups (OGs) conserved in every taxon. As expected, most of these (90) are involved in translation, ribosomal structure and biogenesis and represent the vast majority of this functional category within *Cc5's* genome (137 genes). Considering solely members of the *Bacteroidetes* phylum, the number of conserved orthologs only raised to 333. This contrasts with the much higher number of genes shared with *Flavobacteria* (849 *i.e* 35% of Cc5 genome) and with the three *Capnocytophaga* genomes currently available (1,121 *i.e* 46% of the *Cc5* genome)(**Fig 4.1.3.A**). These data indicate that the *Capnocytophaga* have conserved a relatively high number of functions from *Flavobacteria.* Consequently, *Flavobacteriaceae* seem to have a large, specific and conserved core genome despite their capacity to colonize a wide range of habitats. In contrast, the *Bacteroidetes* phylum appears heterogeneous as most conserved genes were also conserved among all 15 Gram-negative bacteria considered. (**Fig 4.1.3.ABC**).

To have a hint as to the evolution of the *C. canimorsus* genome, we computed phylogenetic trees of 209 conserved proteins in the 15 genomes considered (**Fig 4.1.2**), *C. canimorsus* surprisingly clustered in between the three *Capnocytophaga* species colonizing the human mouth, suggesting that diversification of the *C. canimorsus* branch occurred after adaptation to the oral environment.

Figure 4.1.2. Phylogenetic tree of *Bacteroidetes.*
Consensual phylogenetic tree based on 209 proteins from 13 representatives of the *Bacteroidetes* phylum. Two *proteobacteria* were taken as outgroup (*E. coli* and *N. meningitidis*). Numbers on the branches indicate the % of the 209 trees in which the species were separated by that branch. Branch length is scaled in terms of expected amino acid substitutions per position.

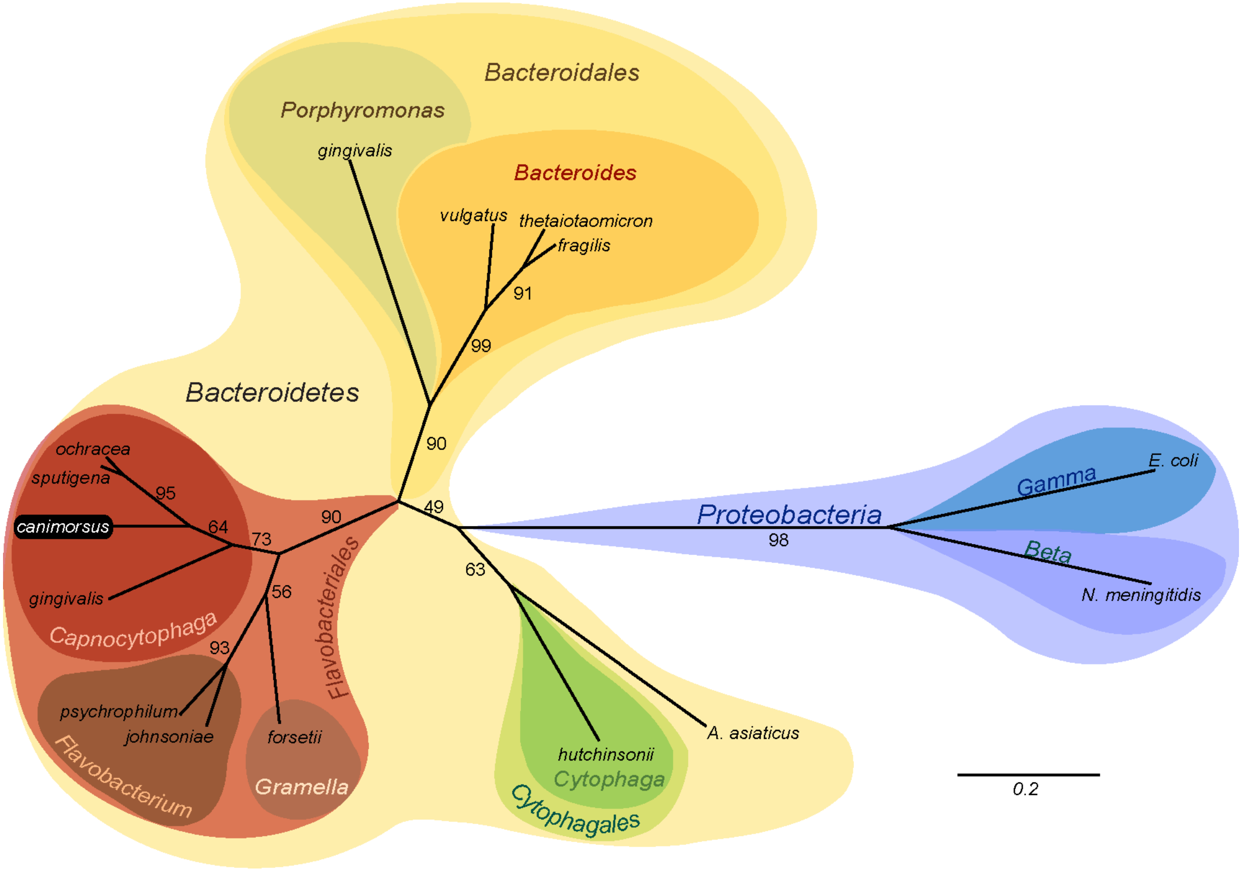
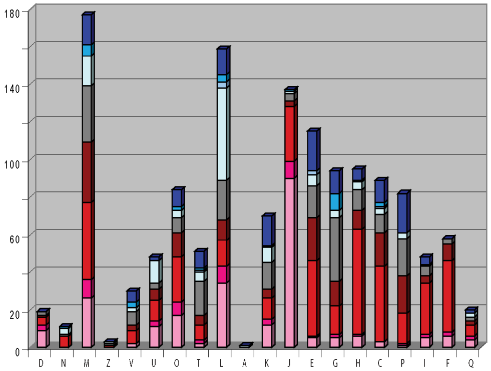


Figure 4.1.3. Orthologous groups distribution at different taxonomic levels or in respect to their functional categories (COG).
A) Taxonomic classes among *orthologous groups* (OGs) including *Cc5* genes. Core groups correspond to OGs with at least one occurrence in all the bacteria from the corresponding taxon (15 genomes considered here, **Fig. 4.1.2**) while Outer groups correspond to OGs where no ortholog was found among genomes from the associated phylotype. B) Histogram representing the genomic distribution of COG functional categories (horizontal axis, D to Q code as in panel C) with color coded taxonomic distribution categories (vertical axis, number of genes). C) Percentage of genes assigned to functional COG categories in the *Cc5* genome. D) Distribution of orthologs and paralogs among the four *Capnocytophaga* considered in this study. Species specific CoDing Sequences (CDS) are exclusively found in the corresponding C*apnocytophaga* genome. Missing genes are defined as CDS found in three C*apnocytophaga* species but missing in the one considered. E) Histogram representing the distribution of the COG functional categories (horizontal axis, D to Q as in B and C) with color coded (as in D) four species (vertical axis, number of genes). F) Groups of Orthologs and close paralogs populating the four *Capnocytophaga* genomes Venn diagram.

A

	Class Specific	Total Class	Total paralogs
▬ Eubacterial Core group	243	243	6
▬ Bacteroidetes Core group	90	333	6
▬ Flavobacteria Core group	516	849	17
▬ Capnocytophaga Core group	272	1121	35
▬ Eubacterial Outer group	623	623	26
▬ Bacteroidetes Outer group	9	632	28
▬ Flavobacteria Outer group	53	685	28
▬ Capnocytophaga Outer group	234	919	41
▬ Unassigned	374		33

B



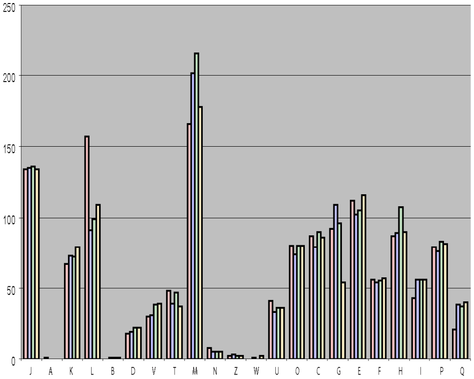
C

Code	Description	%age
D	Cell cycle control, cell division, chromosome partitioning	0.79
N	Cell motility	0.46
M	Cell wall/membrane/envelope	7.33
Z	Cytoskeleton	0.12
V	Defense mechanisms	1.24
W	Extracellular structures	0.00
U	Intracellular trafficking, secretion and vesicules	1.99
Y	Nuclear structure	0.00
O	Posttranslational modification, protein turnover, chaperones	3.48
T	Signal transduction	2.11
B	Chromatin structure and dynamics	0.00
L	Replication, recombination	6.59
A	RNA processing	0.04
K	Transcription	2.90
J	Translation, ribosomal structure and biogenesis	5.68
E	Amino acid metabolism	4.76
G	Carbohydrate metabolism	3.89
H	Coenzyme metabolism	3.94
C	Energy production	3.69
P	Inorganic ion metabolism	3.40
I	Lipid metabolism	1.99
F	Nucleotide metabolism	2.40
Q	Secondary metabolites metabolism	0.83
R	Function unknown	39.52
S	General function prediction	7.66

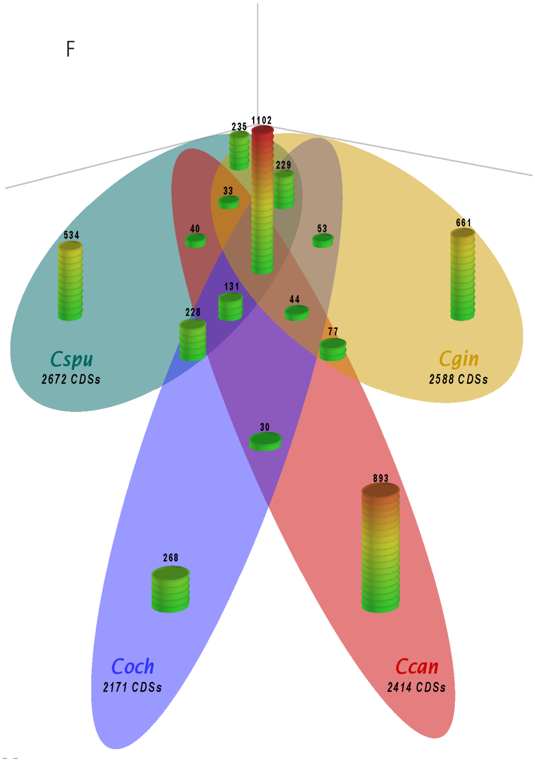
D

	Proteome	Close paralogs	OGs	CDSs in OGs	CDS not Clustered	Total Specie specific	missing CDS present in the 3 other genomes
▬ *C.canimorsus*	2414	109	1753	1817	597	893	229
▬ *C.ochracea*	2171	153	1886	1972	199	268	33
▬ *C.sputigena*	2672	241	2127	2267	405	534	44
▬ *C.gingivalis*	2588	260	1912	2066	522	661	131

E



F



Adaptation to the canine oral environment

89 regions accounting for 0.95 Mb of the *Cc5* genome exhibited significant bias in DNA composition (**Fig 4.1.1**) and most of them encoded mobile genetic elements related genes (**Fig 4.1.1 and Table S4.1**). In addition, 893 *Cc5* genes (36% of the genome) did not match any ortholog in the three other *Capnocytophaga* genomes available and are referred to as the "*Capnocytophaga* outer group" (**Fig 4.1.3.AF**). Within this group of genes, 623 (26.1% of Cc5 genome) even failed to cluster with any homolog at all during OG analysis of 15 genomes (*i.e. Eubacteria* outer group) (**Fig 4.1.3.A**). Hence, during its speciation and adaptation to the mouth of carnivores, *C. canimorsus* acquired a significant number of genes, by horizontal transfer. Some of these genes could originate from other bacteria as illustrated by several successive best blast hits (BHs) from other members of the oral microflora like *Neisseria lactamica,* or *Propionibacterium*. Eukaryotic BHs were also found and often exhibited N-terminal bacterial export sequences suggesting functional selection pressure (**Table S4.1**). The *Cc5* genome contains 157 genes involved in DNA replication, recombination and repair (COG category L) while the 3 other *Capnocytophaga* contain only between 91 and 109 CDSs in this category (**Fig 4.1.3.CDE**). In spite of significant horizontal gene transfer, the genome of *Cc5* (2.57Mb) remains slightly smaller than the genome of the three *Capnocytophaga* colonizing the human mouth (see before). Hence, the genome of *C. canimorsus* has counter-balanced the acquisitions by losses and this is revealed by (i) a low redundancy level (lowest number of paralogs in the *Capnocytophaga* genus (**Fig 4.1.3.**D)), (ii) the absence of many genes conserved in the three other Capnocytophaga (**Fig 4.1.3**.DF) and (iii) a high number of ISs (**Table S4.1** and **Fig 4.1.1**)[61].

Like the other *Capnocytophaga, C. canimorsus* are capnophilic bacteria, meaning that they require a CO_2-enriched atmosphere (>5%) for their growth [7, 22]. This requirement is consistent with the adaptation to the oral environment, known to contain high concentrations of the bicarbonate anion (HCO_3^-)[62]. In *C. ochracea*, HCO_3^--derived carbon has been shown to end up in succinate [62], a major final metabolite [22, 24]. Consistently, *C. ochracea* synthesizes high amounts of phosphoenolpyruvate carboxykinase (PEPCK), an enzyme which catalyzes the conversion of the glycolytic pathway intermediate phosphoenolpyruvate (PEP) and HCO_3^- to oxaloacetate and ATP. Oxaloacetate is then converted in a two-steps reaction to the anaerobic final electron acceptor fumarate (**Fig 4.1.4**). The *Cc5* genome encodes all the enzymes of this pathway as well as a respiratory quinol:fumarate reductase (QFR) membrane protein complex [63] that completes the anaerobic respiratory pathway (**Fig 4.1.4**). To validate these *in silico* findings, we analyzed the culture supernatant of *Cc*5 grown on Raw 264.7 macrophages, by Nuclear Magnetic Resonance. Consistently, the only products released in mM concentrations were acetate (1.75 mM) and succinate (1.82 mM), the reduced product of fumarate respiration (**Fig 4.1.5**).

Diheme-containing QFR based fumarate respiration indirectly generates a proton motive force [64]. However, interestingly enough, *Cc5* metabolism modeling strongly suggests a Na^+ cycle based respiratory chain as observed in marine and pathogenic bacteria such as *Vibrio cholerae*. Accordingly, the two components of the respiratory complex I (Nqr and Mrp), nine solute transporters, three H^+-efflux antiporters and potentially the ATP-synthase appear to be also Na^+-dependent (**Fig 4.1.4**).

Figure 4.1.4 Model of terminal energy catabolism and respiratory chain of *C. canimorsus 5*.
The high potential energy metabolism (e.g. glycolysis) produces pyruvate, oxaloacetate and fumarate (curved red arrows). A main metabolic pathway (Bold black arrows) leads to production of the two major fermentation products succinate and acetate. As shown for *C. ochracea,* the energy metabolism requires a CO_2 dependent PEP carboxylation that produces oxaloacetate (Ccan_10960) and ATP (Ccan_15480) [62, 65]. Oxaloacetate is metabolized into malate, fumarate and succinate. Released succinate could be metabolized by cross-feeding bacteria from the oral polymicrobial community [66, 67]. Like *C. ochracea*, *C. canimorsus* would also form acetate from PEP and increase the ATP yield as compared to succinate formation. Fumarate reduction to succinate is mediated by a Diheme-containing menaquinol-fumarate reductase (QFR) and indirectly contributes to the proton gradient (white arrows) through fumarate respiration [64]. Respiratory complex I is represented by two putative NADH dependent Na^+ pumps, namely Mrp like complex and NQR (NADH:quinone oxidoreductases) that reduce menaquinones (K_2) to menaquinols (K_2H_2). This suggests that the respiratory system of *C. canimorsus* primarily generates a Na^+ gradient in addition to the H^+ gradient. Accordingly, nine solute transporters and three H^+-efflux antiporters appear to be also Na^+-dependent. Two menaquinol oxidative complexes NrfHA and NrfBCD (initially named for nitrate reduction by formate) oxidize menaquinols and indirectly contribute to the H^+ gradient by ammonium formation or oxidized (OCc) cytochrome c reduction (RCc). The NrfBCD complex is genetically associated to a cytochrome c oxidase complex (Cco 1) that could directly interact with RCc generated by NrfBCD. An additional locus coding another Cco complex has been identified in the *Cc5* genome (Cco 2). The specificity to Na^+ or/and H^+ gradients of the F0F1 ATPase is not clearly predicted. However, the γ-subunit (Ccan_01890) hits the ATP synthase γ-chain, Na^+specific model (PTHR11693:SF10). OM: outer membrane, IM: plasma membrane. Doted lines represent hypothetical reactions.

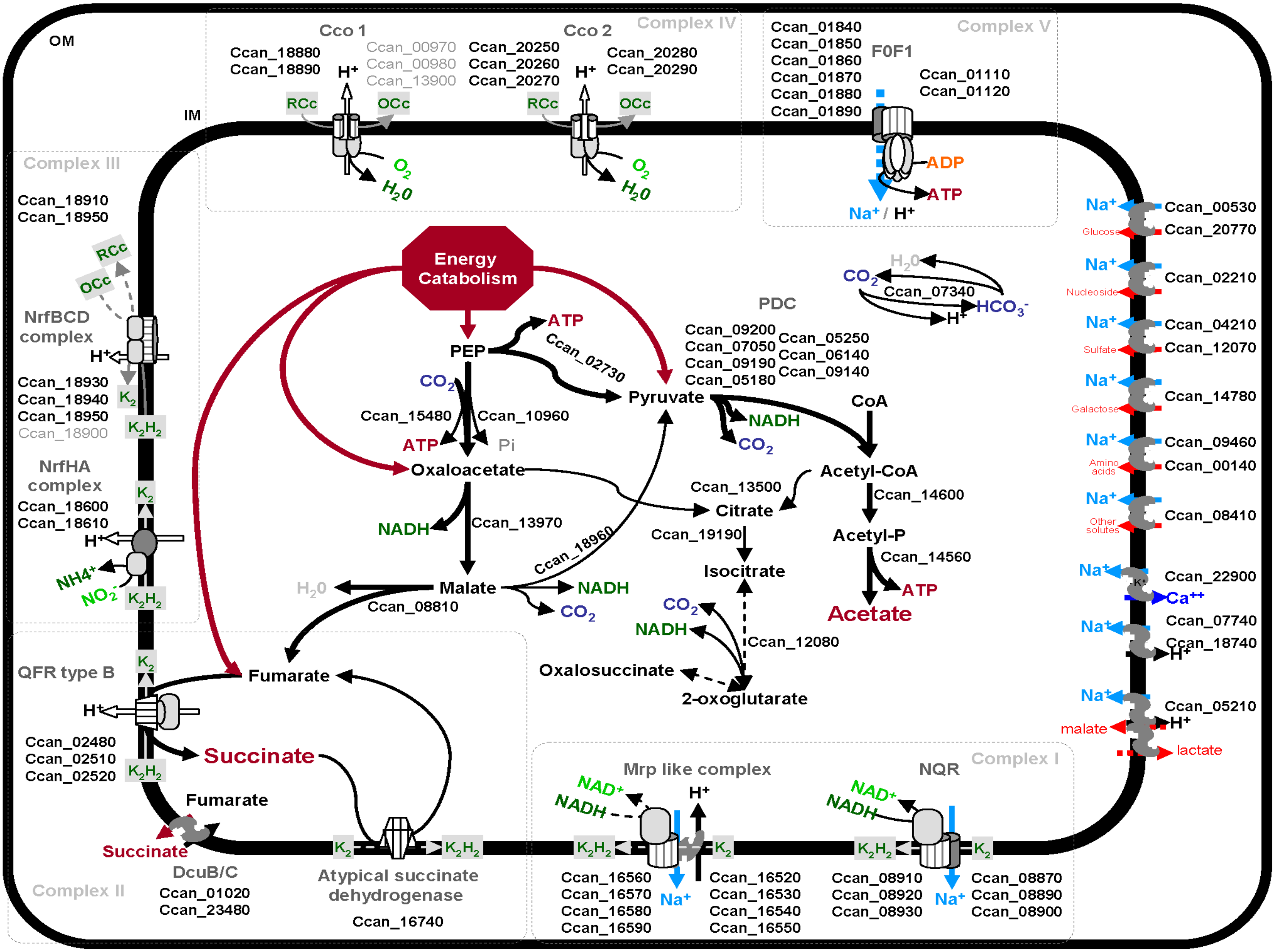
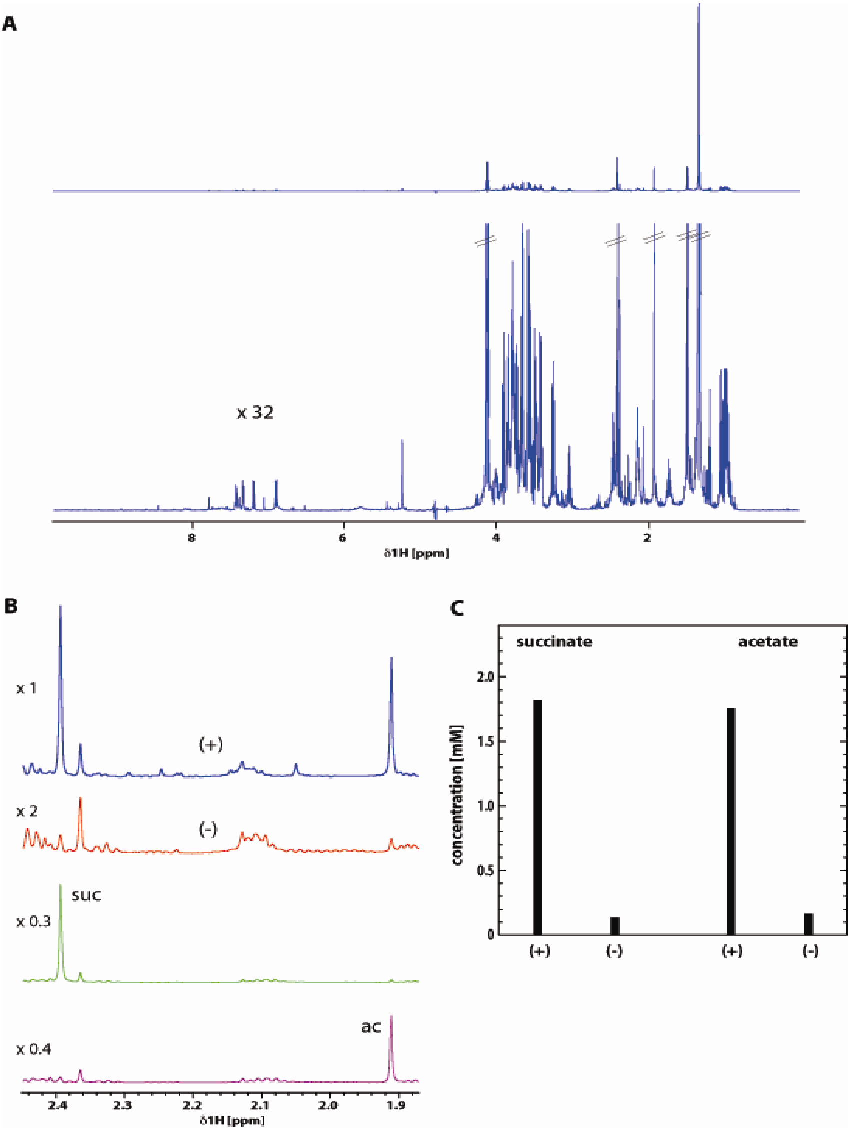

OM
IM
Complex IV
Cco 1
Ccan_18880
Ccan_18890
Ccan_00970
Ccan_00980
Ccan_13900
H+
RCc
OCc
O2
H20
Cco 2
Ccan_20250
Ccan_20260
Ccan_20270
Ccan_20280
Ccan_20290
H+
RCc
OCc
O2
H20
Complex V
F0F1
Ccan_01840
Ccan_01850
Ccan_01860
Ccan_01870
Ccan_01880
Ccan_01890
Ccan_01110
Ccan_01120
ADP
ATP
Na+ / H+
Complex III
Ccan_18910
Ccan_18950
RCc
OCc
NrfBCD complex
H+
Ccan_18930
Ccan_18940
Ccan_18950
Ccan_18900
K2
K2H2
NrfHA complex
Ccan_18600
Ccan_18610
K2
H+
NH4+
NO2-
K2H2
QFR type B
K2
H+
Ccan_02480
Ccan_02510
Ccan_02520
K2H2
Succinate
Fumarate
Succinate
DcuB/C
Ccan_01020
Ccan_23480
Complex II
K2
K2H2
Atypical succinate dehydrogenase
Ccan_16740
Energy Catabolism
PEP
ATP
Ccan_02730
Pyruvate
CO2
Ccan_15480
Ccan_10960
ATP
Pi
Oxaloacetate
NADH
Ccan_13970
Malate
H20
Ccan_08810
Ccan_18960
NADH
CO2
Fumarate
PDC
Ccan_09200
Ccan_07050
Ccan_09190
Ccan_05180
Ccan_05250
Ccan_06140
Ccan_09140
NADH
CO2
CoA
Acetyl-CoA
Ccan_14600
Acetyl-P
Ccan_14560
ATP
Acetate
Ccan_13500
Citrate
Ccan_19190
Isocitrate
CO2
NADH
Ccan_12080
Oxalosuccinate
2-oxoglutarate
H20
CO2
Ccan_07340
HCO3-
H+
Complex I
Mrp like complex
NAD+
NADH
H+
K2H2
K2
Ccan_16560
Ccan_16570
Ccan_16580
Ccan_16590
Na+
Ccan_16520
Ccan_16530
Ccan_16540
Ccan_16550
NQR
NAD+
NADH
K2H2
K2
Ccan_08910
Ccan_08920
Ccan_08930
Na+
Ccan_08870
Ccan_08890
Ccan_08900
Na+
Glucose
Ccan_00530
Ccan_20770
Na+
Nucleoside
Ccan_02210
Na+
Sulfate
Ccan_04210
Ccan_12070
Na+
Galactose
Ccan_14780
Na+
Amino acids
Ccan_09460
Ccan_00140
Na+
Other solutes
Ccan_08410
Na+
K+
Ca++
Ccan_22900
Na+
Ccan_07740
Ccan_18740
H+
Na+
Ccan_05210
H+
malate
lactate

Figure 4.1.5 NMR analysis of the supernatant of Raw 264.7 macrophages cultures infected or not with *Cc5*.

A) overview spectrum of the supernatant from infected cultures. Resonances close to water (4.78 ppm) are obscured due to solvent suppression. B) selected regions from the spectra from the infected (+) and not-infected (-) cultures, as well as of 3 mM succinate (suc) and 3 mM acetate (ac) dissolved in (-) medium. In the infected sample (+), two resonances (2.39 ppm and 1.91 ppm) are more intense than in the non-infected control (-). Data from *C. ochracea* [62] indicate that succinate and/or acetate are the metabolites most likely to have higher concentrations. This assumption was confirmed by the observation of the respective resonances (2.39 ppm, suc) and (1.91 ppm, ac) in the control samples prepared from succinate (suc) and acetate (ac) dissolved in (-) medium. C) Using the NMR peak intensities of the supernatant and control spectra, the following concentrations of these metabolites are determined: 1.82 mM (suc,+), 0.14 mM (suc,-), 1.75 mM (ac,+), and 0.17 mM (ac,-).



Gliding motility and export/import systems

In good agreement with the early observation that *C. canimorsus* exhibits gliding motility [7], the *Cc5* genome contains 20 homologs to the *gld/spr/por* genes encoding the archetypal gliding motility system from *Flavobacterium johnsoniae* [68] (**Table 4.1**).

Table 4.1 Genes involved in gliding motility and the related protein export apparatus

	F. joh	F. psy	C. hut	P. gin	P. int	P. dis	B. fra	B. the	C.can
gldA	Fjoh_1516	FP0252	CHU_1545	PGN_1004	PIN_A1093	BDI_1335	BF2629	BT_0562	**Ccan_13070**
gldB	Fjoh_1793	FP2069	CHU_3691	PGN_1061	PIN_A1414	BDI_1780	BF0973	BT_4189	**Ccan_17700**
gldC	Fjoh_1794	FP2068	CHU_0945						**Ccan_17690**
gldD	Fjoh_1540	FP1663	CHU_3683			BDI_1991			**Ccan_01250**
gldF	Fjoh_2722	FP1089	CHU_1546						**Ccan_07670**
gldG	Fjoh_2721	FP1090	CHU_1547						**Ccan_07660**
gldH	Fjoh_0890	FP0024	CHU_0291	PGN_1566		BDI_1879	BF4095	BT_3818	**Ccan_01070**
gldI	Fjoh_2369	FP1892	CHU_3665	PGN_0743					**Ccan_11090**
gldJ	Fjoh_1557	FP1389	CHU_3494	PGN_1676	PIN_A0879	BDI_3324	BF2407		**Ccan_02810**
gldK(porK)	Fjoh_1853	FP1973	CHU_0171	PGN_1676	PIN_A0879	BDI_3324	BF2407		**Ccan_01610**
gldL(porL)	Fjoh_1854	FP1972	CHU_0172	PGN_1675	PIN_A0878	BDI_3323	BF2931		**Ccan_01620**
gldM(porM)	Fjoh_1855	FP1971	CHU_0173	PGN_1674	PIN_A0877	BDI_3322	BF2932		**Ccan_01630**
gldN(porN)	Fjoh_1856	FP1970	CHU_2610	PGN_1673	PIN_A0876	BDI_3321			**Ccan_01640**
sprA(sov)	Fjoh_1653	FP2121	CHU_0029	PGN_0832	PIN_A1146	BDI_2659			**Ccan_21890**
sprB	Fjoh_0979	FP0016	CHU_2225	PGN_1317	PIN_A1872				**Ccan_06770**
sprE(porW)	Fjoh_1051	FP2467	CHU_0177	PGN_1877	PIN_A2099	BDI_3149			**Ccan_01790**
porP	Fjoh_3477	FP2412	CHU_0170	PGN_1677	PIN_A0880	BDI_3325			**Ccan_00610** **Ccan_03400** **Ccan_03990**
porQ	Fjoh_2755	FP1713	CHU_2991	PGN_0645	PIN_0248	BDI_3738			**?**
porT(sprT)	Fjoh_1466	FP0326	CHU_2709	PGN_0778	PIN_A1079	BDI_1856			**Ccan_09030**
porU	Fjoh_1556	FP1388	CHU_3237	PGN_0022	PIN_A0180	BDI_2576			**?**
porX	Fjoh_2906	FP1066	CHU_1040	PGN_1019	PIN_A2097	BDI_3342	BF2968	BT_0818	**?**
porY	Fjoh_1592	FP2349	CHU_0334	PGN_2001	PIN_A0086	BDI_2438	BF0583	BT_1470	**?**

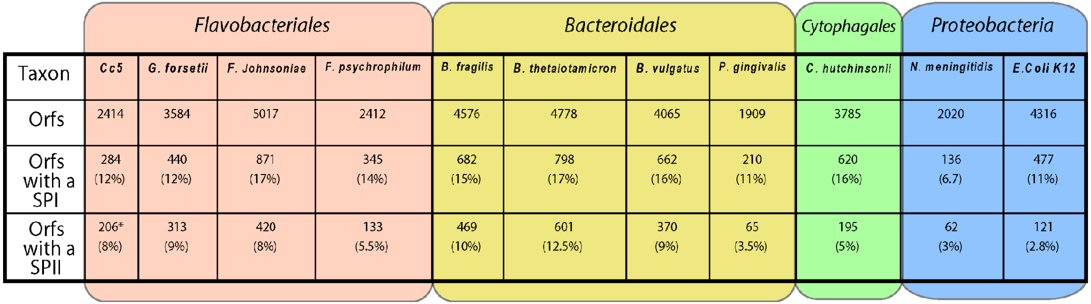
Table modified from [69]. Orthologous genes were defined as reciprocal best-hits. *F. joh*, *F. johnsoniae* UW101 (NC_009441); *F. psy*, *Flavobacterium psychrophilium* JIP02/86 (NC_009613); *C. hut, C. hutchinsonii* ATCC 33406 (NC_008255); *P. gin, P. gingivalis* ATCC 33277 (NC_010729); *P. dis, Prevotella intermedia* 17 (J. Craig Venter Institute); *Parabacteroides distasonis* ATCC 8503 (NC_009615); *B. fra, B. fragilis* YCH46 (NC_006347); *B. the,* and *B. thetaiotaomicron* VPI-5482 (NC_004663). *C. canimorsus* (*C.can*), has been added on the basis of ortholog group analysis with ORTHOMCL.

Regarding protein export, besides the Sec and the Tat protein secretion systems, the genome encodes 6 major facilitators, 20 putative ABC transporters and 4 type I secretions systems but no type II, type III, type IV or type VI secretion systems (**Table S4.1**). However, like the flagellum, the gliding motility was recently shown to include a protein export apparatus [60].

Genome annotation predicts 206 lipoprotein genes, which corresponds to 8.5 % of the total coding capacity (**Fig 4.1.6.A**). This content of lipoproteins is relatively high as compared to *Eubacteria* in general but it is standard among *Bacteroidetes* (**Fig 4.1.6.A**). In agreement with the predicted synthesis of many lipoproteins, the LolACDE lipoprotein export system was identified (**Table S4.1**) but, as for all *Bacteroidetes* currently studied, LolB could not be identified on the basis of the sole sequence. The very high number of lipoproteins suggests that the lipoprotein export pathway could be used as a common protein export pathway as shown for *P. gingivalis* which uses lipoproteins to build surface filaments [70].

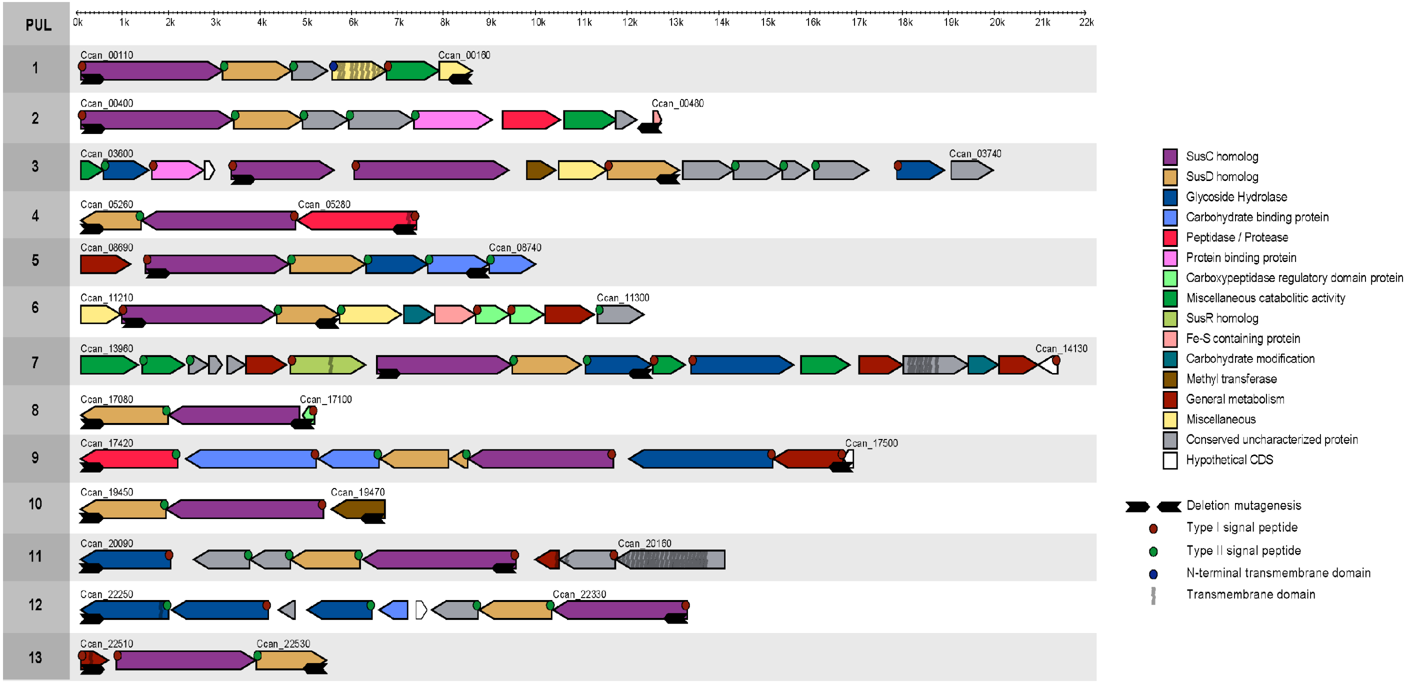
Figure 4.1.6 Bacterial lipoprotein contents comparison and their distribution among the 13 Polysaccharide Utilization Loci of *Cc5.*
A) Genomic content of genes encoding signal peptides I (SPI) or signal peptides II (SPII, lipoproteins) for 11 bacterial genomes. * indicates that 7 lipoprotein annotation tags were manually added to the *Cc5* genome during semi manual curation and were not detected by the LipoP software used here. B) The 13 PULs identified by the presence of SusC-like and SusD-like genes. Putative functions are color coded as indicated in the key. The black arrows show the range of the deletion in the various knockout mutants engineered. Dots and waves give indications concerning the cellular localization of the protein.

A



	Flavobacteriales				Bacteroidales				Cytophagales	Proteobacteria	
Taxon	***Cc5***	***G. forsetii***	***F. Johnsoniae***	***F. psychrophilum***	***B. fragilis***	***B. thetaiotamicron***	***B. vulgatus***	***P. gingivalis***	***C. hutchinsonii***	***N. meningitidis***	***E.Coli K12***
Orfs	2414	3584	5017	2412	4576	4778	4065	1909	3785	2020	4316
Orfs with a SPI	284 (12%)	440 (12%)	871 (17%)	345 (14%)	682 (15%)	798 (17%)	662 (16%)	210 (11%)	620 (16%)	136 (6.7)	477 (11%)
Orfs with a SPII	206* (8%)	313 (9%)	420 (8%)	133 (5.5%)	469 (10%)	601 (12.5%)	370 (9%)	65 (3.5%)	195 (5%)	62 (3%)	121 (2.8%)

B



The genome of *C. canimorsus* 5 contains 13 polysaccharide-utilization loci (PULs)

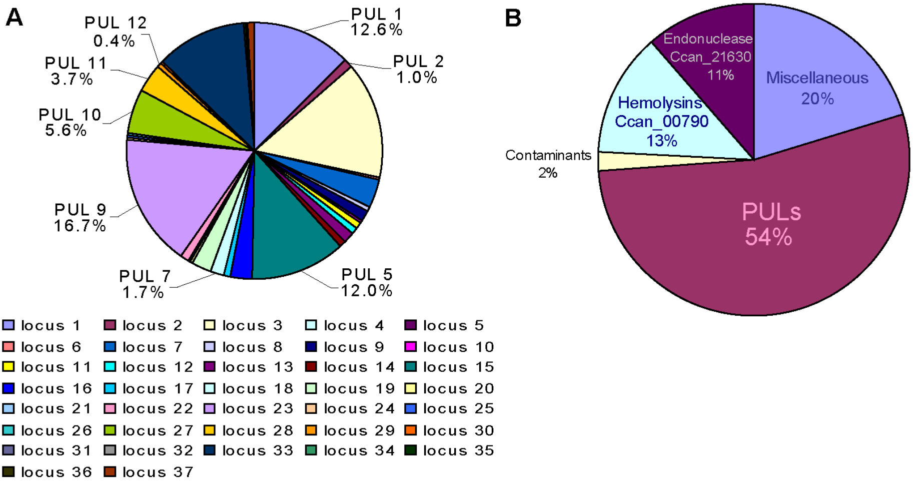
A. Salyers and co-workers discovered that *B. thetaiotaomicron* is endowed with a cell envelope-associated multiprotein system that enables the bacterium to bind and degrade starch [71]. A key feature of this starch utilization system (Sus) is the coordinated action of several gene products involved in substrate binding and degradation. Interestingly, some of the Sus components are predicted to be lipoproteins and have been shown to be surface exposed [72, 73]. Subsequent microbial genome sequencing projects revealed the presence of many polysaccharide utilization loci (PULs) encoding “Sus-like systems” in the genome of *B. thetaiotaomicron* and other saccharolytic *Bacteroidetes* [36, 73, 74]. Sus-like systems target all major classes of host and dietary glycans [75]. Thus, PUL-mediated glycan catabolism is an important component in gut colonization and ecology, but the genome of saprophytic *Bacteroidetes* like *F. johnsoniae* also contains a high number of PULs [60], indicating that PULs are a hallmark of the *Bacteroidetes* phylum rather than of commensal *Bacteroides* only. Since the genome of *C. canimorsus* also encodes a high number of lipoproteins and since *C. canimorsus* can harvest glycan moieties from mammalian surface glycoproteins [58], we paid particular attention to two conserved archetypal outer membrane (OM) proteins (SusC and SusD) [76, 77]. SusC resembles a TonB-dependent transporter and is essential for energy-dependent import of starch oligosaccharides into the periplasm [76] while SusD is a α-helical starch-binding lipoprotein. Iterative Hidden Markov Model screens based on *susD* and *susC* homologs identified 13 hypothetical PULs, which could encode surface feeding machineries (**Fig 4.1.6.B**). This number of PULs is significant but nevertheless much lower than the number found in *B. thetaiotaomicron* (88) [73] and in *F. johnsoniae* (44)[60], which presumably reflects the specialization to the oral cavity niche. As a matter of comparison, we found that the genome from the human *C. ochracea* exhibits 20 PULs.

Within the 13 PULs from *Cc5*, *susC* and *susD* homologs show strong synteny conservation among *Bacteroidetes* (*eg*. between Ccan_14040-14030 and gi:29348720-gi:29348719 from *B. thetaiotaomicron*). However, even though other PUL genes from *Cc5* have orthologs among other *Bacteroidetes*, PULs are poorly conserved, suggesting a reshuffling of these loci during evolution as nicely shown recently for *Bacteroides plebeius* [78].

Glycan-foraging complexes are predominant at the bacterial surface

The genome of *C. canimorsus* encodes a high proportion of predicted lipoproteins and a significant proportion of them are part of Sus-like systems. Consequently, we hypothesized that most of these lipoproteins would be surface exposed and not periplasm-exposed like in enterobacteria for instance. To test this hypothesis, we carried out a proteomic analysis of the surface of *Cc*5 bacteria cultivated onto HEK293 cells. The first approach was a mild tryptic shaving followed by MS/MS analysis (**Table S4.2**). Excluding 6 clear cytosolic contaminants (2 elongation factors and 4 ribosomal proteins), this approach identified a total of 62 putative surface-exposed proteins, including 59 where the peptide detection signal was strong enough to allow a relative quantification. As a control, we applied our shaving protocol to a corresponding lysed bacterial preparation and samples were analyzed by MS/MS (**Table S4.2**). The two lists of proteins were significantly different and, as expected, the contaminants from the shaving ranked high in the list of total proteins. Among the 62 hypothetical surface proteins, 38 were predicted to be lipoproteins processed by signal peptidase-II, 18 had a classical signal peptide and 6 had no characteristic peripheral feature (**Table S4.2**). The second approach was a surface biotinylation followed by avidin purification. It gave only 24 hits with only 3 clear cytosolic contaminations (1 elongation factor already contaminating the list of shaved proteins and 2 ribosomal proteins) but nevertheless added 13 new proteins to the surfome list, among which 3 predicted outer membrane proteins (OMPs) without SP, 8 with a SPI and two lipoproteins. Interestingly, 4 of the new proteins, including 3 *susC* homologs, happened to be encoded by PULs previously detected by the shaving method. In total, the list of surface proteins came thus to 75 (**Table S4.2**). Interestingly, the predominant proteins from the surfome were those encoded by PUL9 (16.6 %), PUL1 (12.6 %), PUL5 (12.0 %) as well as a putative thiol-activated cytolysin (Ccan_00790) (12.8 %) and a putative endonuclease (Ccan_21630) (11.3 %). Products of PUL2, -6, -10, -11 and -12 were minor components of the surfome. Finally, products of PUL3, -4, -7 and -13 were detected in purified outer membranes. Thus, when *Cc5* bacteria were grown on Hek293 cells, at least 12 PULs were expressed and their products contributed 53.5 % of the total peptides from the surfome (**Fig 4.1.7**). All this indicates that surface-exposed complexes specialized in foraging complex glycans or other macromolecules play a central role in the biology of *C. canimorsus*.

Figure 4.1.7. Genetic and Functional distribution of the surfome of *C. canimorsus 5*



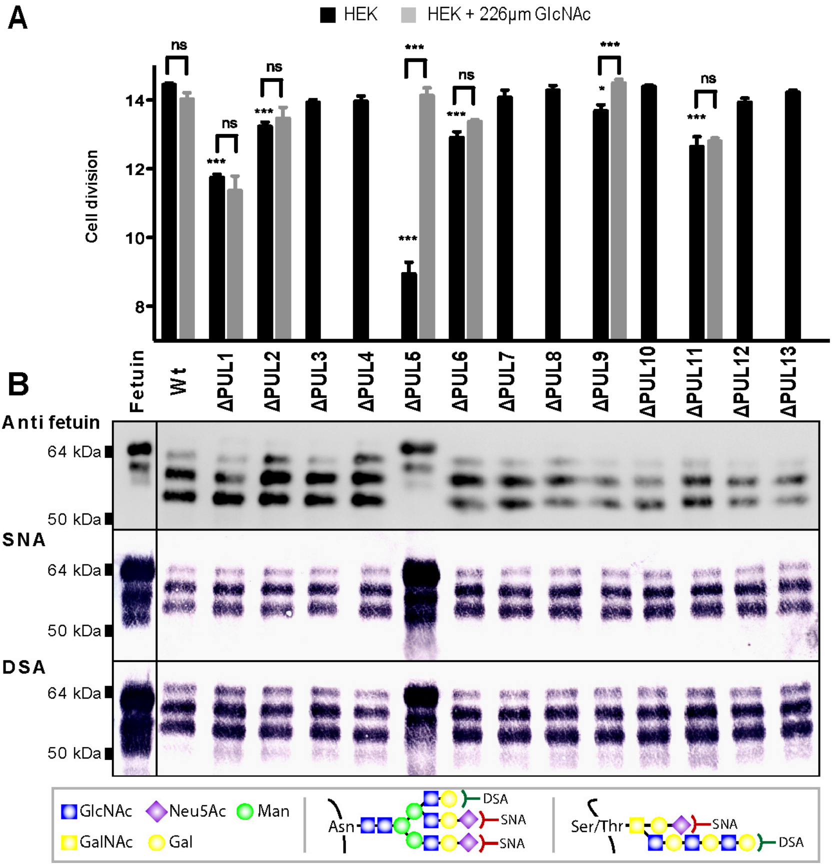
59 surface-exposed proteins are encoded by only 34 loci, suggesting that most of these proteins form functional complexes. In agreement with this, these loci include 8 out the 13 PULs identified in the genome. Proteins were quantified by MS-MS peptide intensity. Panel A: % of the surface proteome encoded by the 37 loci (including 3 ribosomal contaminant loci). Panel B: Functional distribution of surface protein highlighting the predominance of PUL-encoded feeding complexes at the bacterial surface (53.5%). The endonuclease Ccan_21630 and the surface exposed putative hemolysin Ccan_00790 respectively accounted for 11% and 13% of the total surfome.

PULs contribute to growth on cells, host protein deglycosylation and survival in human serum and in a murine model.

In order to assess the impact of these feeding complexes on growth at the expenses of mammalian cells, we undertook to independently knockout each of the 13 PULs. Removal of some PULs had a clear impact on growth on Hek293 cells but not on growth in blood agar plates. Deletion of PUL5 alone led to a severe reduction of growth at the expenses of Hek293 cells (**Fig 4.1.8.A**) but deletion of PUL1,-2,-6,-9 or -11 also had a lower but significant impact. In the case of PUL5 and PUL9, the growth deficiency could be suppressed by the addition of N-Acetylglucosamine (GlcNAc) to the culture medium (**Fig 4.1.8.A**), suggesting that these PULs do indeed encode glycan foraging systems.

In order to confirm that *C. canimorsus* grow at the expenses of cellular glycoproteins, wt *Cc5* bacteria and PUL deletion mutants were incubated with fetuin, a standard serum glycoprotein and the glycosylation state was monitored by lectin staining and immuno blotting. As shown in **Fig 4.1.8.B**, fetuin was deglycosylated by wt *Cc5* bacteria and by all the PUL deletants, except by PUL5 deletants.

Figure 4.1.8 Contribution of the different PULs to feeding on HEK293 cells and to fetuin deglycosylation.
A. The 13 PUL knockout mutant strains were inoculated on HEK293 cells at moi=0.2, with (grey) or without (black) supplemented N-Acetyl glucosamine (GlcNAc) and grown for 23 hours. Significance is assessed by T-test of wt vs. ΔPUL deletants and GlcNAc complementation vs. its corresponding non complemented ΔPUL (n=3). B. Deglycosylation of fetuin. top, western blot with anti-fetuin; middle: staining with the *Sambucus nigra* lectin (SNA) that binds preferentially to terminal Gal(α2-6)Sialic acid; bottom, staining with *Datura stramonium* lectin (DSA) that recognises (β-1,4) linked N-Acetylglucosamine oligomers.



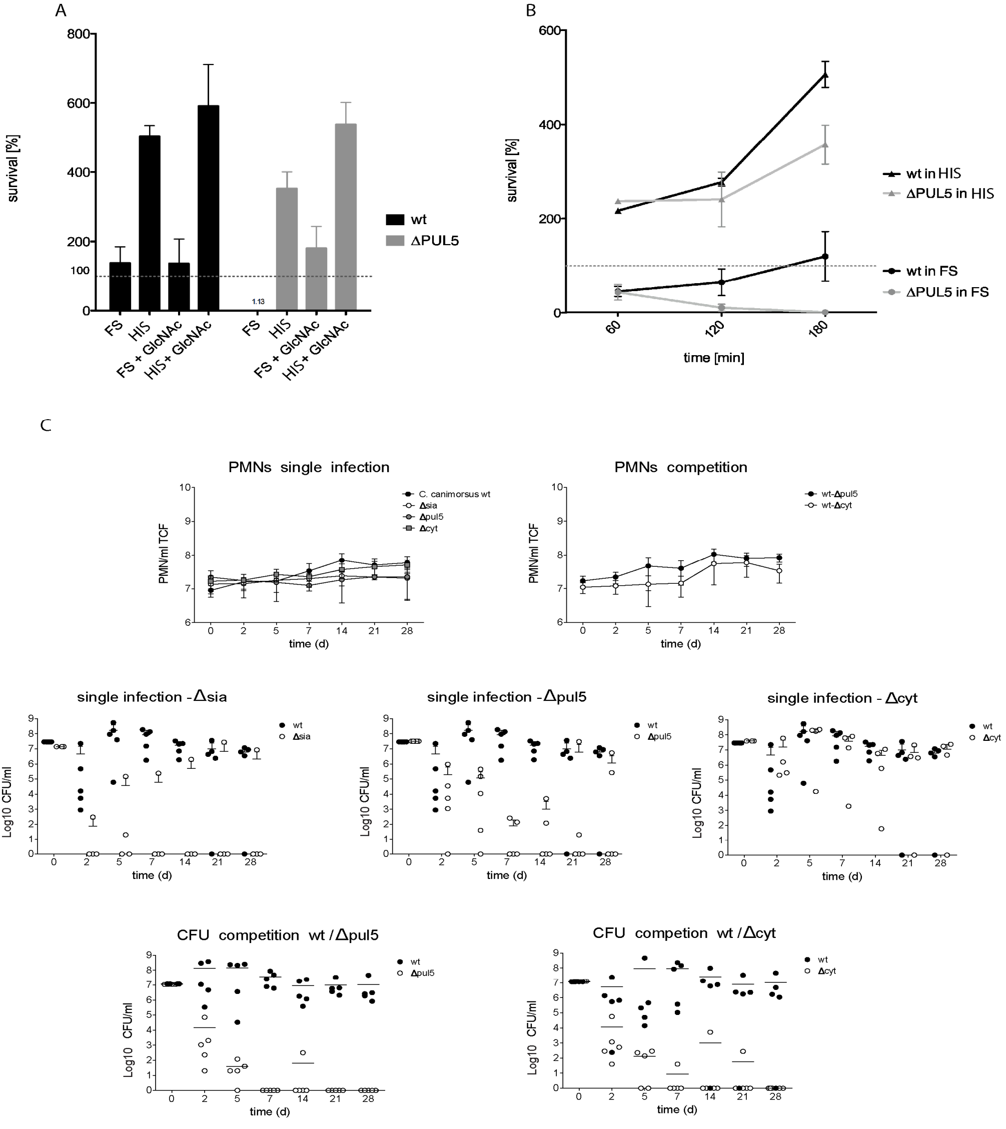
We conclude from all these observations that PUL5 plays a major role in the capacity of *C. canimorsus* to feed on live host cells by deglycosylating surface glycoproteins. The locus, which is among the most expressed PULs, (encoding 12% of the surfome, see previous section) consists of six genes. The SusC-like integral OMP represents the porin of the system, three lipoproteins presumably involved in substrate binding and a forth one predicted to be an endoglycosidase (**Fig 4.1.6.B**).

Since deglycosylation of host proteins could also contribute to growth during septicemia, we compared the growth of wt and ΔPUL5 bacteria in fresh and heat inactivated human serum. As shown in **Fig 4.1.9.AB**, while wt bacteria could grow even in fresh serum, the ΔPUL5 bacteria were significantly impaired in their growth. They even showed some sensitivity to the bactericidal activity of fresh human serum, although not to the same extend as a mutant affected in LPS synthesis [55]. Interestingly, serum sensitivity exclusively resulted from growth impairment in human serum as it was complemented by GlcNac.

Finally, we compared the survival of wt and ΔPUL5 bacteria in teflon cages implanted into mice, the only reported animal model for *C. canimorsus* [58]. We also included in this study, the sialidase mutant known to persist less than wt [58] and a mutant affected in the thiol-activated cytolysin (Ccan00790). As shown in **Fig 4.1.9.C**, in each experiment, only 1 out of 5 mice cleared wt *Cc5* bacteria after 28 days. In contrast, 4 mice cleared the sialidase mutant and 3 mice cleared the ΔPUL5 mutant. Only one mouse cleared the cytolysin mutant. In competition experiments, ΔPUL5 and cytolysin mutants were cleared. We infer from all these data that PUL5 contributes to the survival in mice and in fresh human serum and hence that PUL5 can be considered as a virulence factor [58].

Figure 4.1.9. Survival and growth of *wt* and Δ*PUL5 Cc5* in murine tissue cages and in serum.

A) *Cc5* bacteria were injected into tissue cages implanted into mice and bacterial loads were inferred from the number of colony forming units after plating tissue cage fluid. *Cc5 wt* and knockout for PUL5 5 (ΔPUL5), sialidase (Δsia, Δ*Ccan_04790::ermF*) and cytolysin (Δcyt, Δ*Ccan_00790::ermF*) were tested. Polymorphonuclear neutrophils (PMNs) populations were monitored during infection with no significant increase observed (two top graphs). Single infections and competition assays were followed during 28 days. B-C) 10^7 *Cc5* bacteria were suspended in 1 ml of 10% human fresh serum (FS) or heat inactivated serum (HIS). In panel B, bacteria were counted by plating after 3 h of incubation in presence or absence of N-Acetylglucosamine (GlcNAc). In panel C, samples were counted after 1, 2 and 3 hours of incubation.



DISCUSSION

Our genome analysis confirms the relatedness between the mouth commensals from the *Capnocytophaga* genus and the gut commensals from the *Bacteroides* genus but it also shows that the *Bacteroidetes* phylum is heterogeneous, suggesting that intermediate clades or taxa remain unknown. The genome analysis also shows that *Capnocytophaga* are closer from *Flavobacteriaceae* such as the marine *G. forsetii* [79] and the soil and lake saprophytic bacterium *F. johnsoniae* [60] than from *Bacteroides*. With *F. johnsoniae*, *C. canimorsus* shares the whole set of 13 gliding motility genes (*gldA-N*) (**Table 4.1**) agreeing with its initial description as a gliding bacterium [7]. During growth on mammalian cells, *Cc5* bacteria produced large amounts of succinate. Genome-based metabolic modeling suggests that succinate was generated by CO_2-dependent fumarate respiration coupled to Na^+ gradient based respiratory chain. This model is consistent with the capnophilia of *C. canimorsus* and with the relatively high concentration of HCO_3^- in saliva (25 mM).

The genome of *Cc*5 did not encode any of the complex secretion pathways commonly found in the α and γ proteobacteria like T2S, T3S, T4S and T6S. In contrast, *C. canimorsus* was found to encode an unusually high proportion of predicted lipoproteins, like several other members of the BFC group. However, analysis of the *Cc*5 surface proteome indicated that, in contrast to what is seen in proteobacteria, a significant part of these lipoproteins are surface exposed. This property, suggests that these bacteria expose a number of proteins on their surface rather than secreting them. In *P. gingivalis*, it has even been shown that major structural components of two cell surface filaments are matured through lipoprotein precursors [70]. A substantial routing of proteins through the lipoprotein pathway could thus be central to the biology of the whole BFC group. The abundance of these surface exposed lipoproteins coupled to the fact that *C. canimorsus* was shown to deglycosylate mammalian lipoproteins hinted that *C. canimorsus* is endowed with foraging systems like the archetypal starch utilization system (Sus) of *B. thetaiotaomicron* which also includes predicted lipoproteins [74]. This system consists of several lipoproteins with capacities to bind (SusD-like) or to hydrolyse complex polysaccharides and of a TonB-dependent porin (SusC-like), which are thought to form a complex [73, 74, 80]. A screen for homologs of SusC and SusD confirmed the presence of 13 putative PULs, encoding Sus-like systems. This number of PULs is significant but nevertheless much lower than the number found in *B. thetaiotaomicron* (88) [73] and in *F. johnsoniae* (44)[60] but similar to the number found in *G. forsetii* (14), a marine bacterium adapted to the degradation of high molecular weight organic matter with a predicted preference for polymeric carbon sources [79]. The low number of PULs reflects the specialization to the oral cavity niche rather than a reduced importance of the complexes encoded by these loci. Indeed, PUL-encoded proteins represent more than half of the surface-exposed proteins and hence the most important protein class at the bacterium-host interphase. The low number of PULs found in *C. canimorsus* compared to *Bacteroides spp.* suggests that *C. canimorsus* feeds less from the host diet and more from the host itself and from the rest of the complex mouth flora [81]. Besides the homologs to SusC and SusD, most of these 13 PULs encode putative glycan hydrolases. Six PULs turned out to be involved in the capacity of *C. canimorsus* to grow at the expenses of mammalian cells [58]. One of them, PUL5 was found to encode a complex involved in N-linked glycoprotein deglycosylation and this complex turned out to be the most abundant at the bacterial surface, underlying the importance of protein deglycosylation for these bacteria. Interestingly, *B. thetaiotaomicron* has already been shown to deglycosylate mucin O-glycans from the gut [73]. The observation that PUL5-encoded complex deglycosylates N-linked glycoproteins nicely fits with the previous report showing that sialidase is key to growth of *C. canimorsus* at the expenses of cells and their persistence in the mouse [58]. Sialidase presumably cooperates with the PUL5 proteins in spite of the fact that it is encoded outside any of the 13 PULs. Not surprisingly, like the sialidase gene, the *PUL5* genes were also found to be necessary for survival and growth in human serum as well as persistence in the mouse. In conclusion, although the genome of *Cc5* does not encode any classical virulence function, it encodes a surface-exposed glycoproteins foraging system which can be considered as a new type of virulence factor.

Methods

Ethics statement:

Animal experiments were performed in strict accordance with institutional and guidelines of the Swiss veterinary law (article 13a TSchG; 60-62 TSchV). The protocol was reviewed and approved by the veterinary office of the canton Basel (Permit Number: 1397-Inflammation and mouse peritonitis model in mice, valid until 2010-12-31).

Human serum samples used for this study were provided by the “Blutspendezentrum SRK beider Basel”. Samples were taken from healthy volunteer blood donors after obtaining written informed consent, in agreement with the guidelines of the “Ethikkommission beider Basel EKBB”.

Bacterial growth conditions: *C. canimorsus* bacteria were routinely grown on heart infusion agar supplemented with 5% sheep blood at 37°C in the presence of 5% CO_2. For growth on cells, $4x10^4$ bacteria were incubated with $2x10^5$ HEK293 cells or Raw 264.7 macrophages in a final volume of 1ml DMEM with 10% (v/v) fetal calf serum and 1mM sodium pyruvate for 23h (DETAILED MATERIAL AND METHODS).

Genome sequencing and annotation: Genomic DNA of *C. canimorsus 5* was isolated by using the QIAGEN Genomic-tip 500/G and corresponding buffers followed by Phenol / Chlorophorm purification to achieve even higher DNA purity. Sequencing of the *Cc5* chromosome integrated four different sequencing approaches corresponding to more than 80X read coverage in total (see DETAILED MATERIAL AND METHODS). Assembly and annotation of the genome are described in the DETAILED MATERIAL AND METHODS**.**

Proteome: For the surface-exposed proteome, bacteria were grown on HEK293 cells, harvested by carefully washed twice with 10mM Hepes and trypsinized for 30 min at 37 °C. The supernatant was then filtered through 0.20 µm pore size filters, reaction was stopped with formic acid (0.1% final) and peptides were stored at – 20 °C until further analysis. Alternatively, the surface-exposed proteins were biotinylated with Sulfo-NHS-SS-Biotin (0.02 g/L) after bacteria were first incubated with regular biotin (0.2 g/L) in order to saturate the transport systems. The bacterial lysate was then cleared by centrifugation and the labeled proteins were immobilized on avidin. Finally, bound proteins were released by incubating the resin with SDS-PAGE sample buffer containing 50 mM DTT and analyzed by MS-MS. For the OM proteome, bacteria were collected from blood agar plates, resuspended at OD_{600}=1 and sonicated. Membrane pellets were resuspended in HEPES 10mM with 1% Sarkosyl incubated at room temperature for 30 minutes and re-centrifuged. The pellet was resuspended and analyzed by MS-MS. More details are given in DETAILED MATERIAL AND METHODS.

Identification of the main metabolic end product in Cc5 culture supernatants:

Cc5 were grown in the presence of Raw 264.7 macrophages. 0.1% NaN_3 was added to the supernatant and pH adjusted to 7.5. The medium was finally filter sterilized and the macromolecules discarded by a 3 kDa cut-off filter. Following steps were carried out on samples containing 5% D_2O in 5 mm standard NMR tubes and samples were measured with a spectrometer equipped with a triple resonance pulse field gradient probehead. The temperature of 297.18 K was determined according to the splitting (1.675

ppm) of a 100% ethylene glycol temperature calibration sample. Spectra were processed and evaluated using the software Topspin 2.1.6. 1D proton NMR spectra were recorded with the excitation sculpting scheme achieving water suppression by gradient dephasing of the water resonance. The proton carrier was set to the water frequency for solvent suppression. Spectra were recorded with 57344 complex points and acquisition times of 1.99 seconds (DETAILED MATERIAL AND METHODS).

Mutagenesis and allelic exchange was performed has described in ref [82] with slight modifications (DETAILED MATERIAL AND METHODS).

Survival and growth in human serum: bacteria were harvested from blood agar plates. A total of 10^7 bacteria were incubated in 10% NHS PBS with or without 0.005% GlcNAc (w/v) at 37℃ in a heating block. Serial dilutions were plated onto blood plates, and viable colonies were counted after 48h of incubation in a humidified atmosphere supplemented with 5% CO_2 at 37℃ (DETAILED MATERIAL AND METHODS).

Tissue cages infection in mice were performed has described in ref [82](DETAILED MATERIAL AND METHODS).

Detailed material and methods

Conventional bacterial growth conditions and selective agents

The strains used in this study are listed in *Appendix*. *Escherichia coli* strains were routinely grown in LB broth at 37℃. *C. canimorsus* bacteria were routinely grown on heart infusion agar (Difco) supplemented with 5% sheep blood (Oxoid) for 2 days at 37℃ in the presence of 5% CO_2. To select for plasmids, antibiotics were added at the following concentrations: 10 µg/ml erythromycin (Em), 10 µg/ml cefoxitin (Cf).

Growth of Cc5 bacteria on HEK293 cultured cells

Human Embryonic Kidney 293 cells (HEK293) were cultured in DMEM (Invitrogen) with 10% (v/v) fetal calf serum and 1mM sodium pyruvate. Cells were grown in medium without antibiotics in a humidified atmosphere enriched with 5% CO_2 at 37℃. Bacteria were harvested by gently scraping colonies off the agar surface and resuspended in PBS to an OD_{600} of 0.0008. A total of $4x10^4$ bacteria were incubated with $2x10^5$ HEK293 cells in a final volume of 1ml medium with or without 0.005% GlcNAc (w/v) devoid of antibiotics for 23h, resulting in a multiplicity of infection of 0.2. Serial dilutions were plated onto blood plates, and viable colonies were counted after 48h of incubation in a humidified atmosphere enriched with 5% CO_2 at 37℃.

Genomic DNA preparation

Genomic DNA of *C. canimorsus 5* was isolated by using the QIAGEN Genomic-tip 500/G (Cat.No.10262) and corresponding buffers (Cat.No.19060) followed by Phenol / Chlorophorm purification to achieve even higher DNA purity.

Global sequencing strategy, Assembly

Sequencing and assembly of the Cc5 chromosome included i) pair-end reads from a ~4 kb inserts plasmid library of ~25 000 clones representing ~10X physical Coverage, ii) pair-end reads from a ~ 40 kb inserts fosmid library of ~ 4600 clones corresponding to ~60X physical Coverage, iii) A run of 454 pyrosequencing corresponding to 20X read Coverage and iv) a set of 33 nucleotides microreads generated with Solexa sequencing technology corresponding to ~49X read coverage. In addition, targeted sequencing has been performed on weakly covered regions. Assembly has been done with Phred/Phrap/Consed package [83-85]. Short reads (454) have been preassembled and condensed into pseudoreads using Newbler assembler (http://www.454.com/). Pseudoreads were then integrated to the Sanger data using Phrap. After gaps closure, micro reads (Solexa) have been aligned with the circular chromosome of Cc5 using MAQ [86] to increase coverage and base call confidence particularly on homopolymeric tracts.

CDS Annotation

Glimmer 3.02 [87] was run with default settings. Predicted coding sequences (CDSs) were then considered for possible alternative starting codons. Briefly, a score based in-house Perl script compiled i) the distance of the considered CDS from the initial CDS prediction by Glimmer, ii) the bacterial frequency of the starting codon considered, iii) the possible presence of an N-terminal signal peptide computed by LipoP [87], iv) and the N-terminal alignment of the current CDS with its best blast hit [88] against the GenBank's non-redundant database (NR, at the NCBI). C-terminal properties as possible early stop codons (pseudogenes) or fusion/deletion events were also inferred from such alignments. Finally, CDS overlaps were monitored and CDSs eventually shortened. Best scored CDSs were then screened with EMBOSS:pepstats [89] for physico-chemical inferences, with InterProScan [90] for domain identification and PRIAM [91] for accurate EC annotation. For each CDS, a position-specific matrix has been computed for 5 cycles against the uniref90 using a size adapted initial matrix with PSI-BLAST [88] (cutoff: $10e^{-5}$ e-value). Matrices were then used during a one-iteration PSI-BLAST *vs.* Swiss-Prot, TrEMBL [90] or STRING Orthologous Groups [92] for COG assignment.

Non coding RNAs

The complete chromosome has been scanned against all Rfam CMs using the INFERNAL software [93] with default options and stringent bit score cutoff (40) has been applied. rRNAs have been predicted with RNAMMER [94] and tRNA with tRNAscan-SE [95].

Genomic DNA sequence features

The chromosomal origin of replication has been suggested based on the location of lowest cumulative GC skew value and presence of DnaA boxes clusters. The first T of the AT rich region was proposed as +1.
Alien Hunter v1.7 has been used to spot bias in DNA composition often due to recent DNA acquisition or very high transcriptional levels [96].

Orthologs groups

15 predicted proteomes were clustered in ortholog groups using Orthomcl v1.4 [97] with the following settings: OrthoMCL Mode 1, P-value Cut-off 1e-05,

Percent Identity Cut-off 30, Percent Match Cut-off 50, MCL Inflation 1.5, and Maximum Weight 316. Predicted proteomes used in the present study include those from *Capnocytophaga canimorsus 5* (CP002113), *Capnocytophaga ochracea DSM 7271* (NC_013162), *Capnocytophaga gingivalis ATCC 33624* (NZ_ACLQ00000000), *Capnocytophaga sputigena ATCC 33612* (NZ_ABZV00000000), *Flavobacterium johnsoniae UW101* (NC_009441), *Flavobacterium psychrophilum JIP0286* (NC_009613), *Gramella forsetii KT0803* (NC_008571), *Bacteroides fragilis YCH46* (NC_006347), *Bacteroides thetaiotaomicron VPI5482* (NC_004663), *Bacteroides vulgatus ATCC 8482* (NC_009614), *Porphyromonas gingivalis W83* (NC_002950), *Cytophaga hutchinsonii ATCC33406* (NC_008255), *Amoebophilus asiaticus 5a2* (NC_010830), *Escherichia coli K-12 W3110* (AC_000091) and *Neisseria meningitidis 053442* (NC_010120).

Phylogenic analysis

Consensual phylogenetic tree of 13 *Bacteroidetes* and two proteobacteria (mentioned here above) has been computed using the PHYLIP package 3.6 [98]. Among 243 orthologous groups (OGs) conserved in every taxon, 209 were exempt of any paralog and were used to compute single protein phylogenies with Maximum Likelihood. Amino acid sequences from the same OGs were first aligned with ClustalW (default settings) [99]. Alignment files were then used as input for Proml (PHYLIP 3.65) with following settings: S, o, 15, o, m, d, 21, 3, 1, Y (http://evolution.genetics.washington.edu/phylip/doc/proml.html) and 209 single protein Maximum Likelihood phylogenetic trees were generated. A Consensus tree has been inferred with Consense (PHYLIP 3.65) following the extended Majority rule (default settings) and species partition scores were kept as confidence estimates. Topology restricted comparisons between the consensus and the 209 single protein trees have been performed with treedist in Symmetric Difference mode (PHYLIP 3.65). Finally, the 21 OGs exhibiting best scoring trees (closest topology from consensus) have been used for branch length estimation using Proml (settings: s, g, o, 15, Y) on the concatenated corresponding alignments (14,130 amino acids).

Identification of SusC/SusD homologs in the genome of *C.canimorsus* 5

SusC (gi|29341017|gb|AAO78807.1) and SusD (gi|29341016|gb|AAO78806.1) from *Bacteroides thetaiotaomicron* VPI-5482 were blasted against the nr70 subset. Hits above the threshold (Hsp_evalue < 10e-5 & Hsp_align_len/ORF_Length > 0.6 & Hsp_align_len/Hit_len > 0.6 & Hsp_identity/Hsp_align_len > 0.4) were aligned with clustalW from the MEGA 4 software (default settings). Alignments were used to build HMMs with HMMER.2.3.2. Models were calibrated and *C. ochracea* and *C.canimorsus 5* homologs screened out. In the case of SusD, an arbitrary initial cutoff “e-value” (0.25) was chosen so that all predicted hits from the first cycle were fished in the vicinity of TonB-dependent outer membrane proteins. Concerning SusC, an arbitrary initial cutoff “e-value” (10e-14) was chosen so that all predicted hits from the first cycle were fished in the vicinity of the previously detected SusD homologs. The newly identified protein sequences were then integrated into the HMM and the procedure has been repeated with the same cutoff e-value (0.25 or 10^{e-14}) until no new hit was detected.

Identification of the main metabolic end product in Cc5 culture supernatants

Cc5 were grown (24h) in the presence of murine macrophages (Raw 264.7) in Dulbecco's modified eagle medium supplemented with 1mM Na-Pyruvate and 10% v/v fetal calf serum. Medium was collected and the bacteria pelleted by centrifugation (5 minutes, 15000rcf, 4°C). 0.1% NaN3 was added to the supernatant and pH adjusted to 7.5 with phosphate-buffer (500mM, pH8). The medium was finally passed through a 0.22 um filter and a 3 kDa cut-off filter (vivaspin, Sartorious). Following steps were carried out on samples containing 5% D2O in 5 mm standard NMR tubes and samples were measured with a Bruker Avance DRX 600 spectrometer equipped with a triple resonance pulse field gradient probehead. The temperature of 297.18 K was determined according to the splitting (1.675 ppm) of a 100% ethylene glycol temperature calibration sample. Spectra were processed and evaluated using the software Topspin 2.1.6 (Bruker). 1D proton NMR spectra were recorded with the excitation sculpting scheme (pulseprogram zgesgp in the standard Bruker library) as described previously [100] achieving water suppression by gradient dephasing of the water resonance. The proton carrier was set to the water frequency for solvent suppression. Spectra were recorded with 57344 complex points and acquisition times of 1.99 seconds. With 64 scans, the total experimental time was 3 minutes and 26 seconds.

Bacterial Surface Digestion

The surface-exposed proteins from *C. canimorsus 5* bacteria were digested essentially as described in ref *[101]* and [102]. Bacteria were grown on heart infusion agar plates (Difco) supplemented with 5% sheep blood (Oxoid) (SB plates) for 2 days at 37°C in the presence of 5% CO_2. They were then suspended in PBS and used to infect 7.5 x 10^6 HEK293 cells at an moi of 10 (≈10^8 bacteria). Infected cells were incubated for 15h at 37 °C in DMEM (Invitrogen) medium supplemented with 10% (v/v) fetal bovine serum (FBS). The medium and bacteria were collected taking care not to detach the HEK293 cells and centrifuged at 1000 g for 5 min at 4°C to get rid of the HEK293 cells eventually present. The supernatant was then centrifuged at 3500 g for 10 min at 4 °C to harvest bacteria. The bacterial pellet was gently resuspended in 10mM Hepes (pH 7.4) and then washed twice with 10mM Hepes (pH 7.4). Cells were resuspended in 1 ml of 10mM Hepes (pH 7.4) and 10 µg trypsin (Roche) was added. Digestion was carried out for 30 min at 37 °C. Bacterial cells were then spun down at 3.500 g for 10 min at 4 °C and the supernatant was filtered through 0.20 µm pore size filters (Millex, Millipore, Bedford, MA). Protease reaction was stopped with formic acid (0.1% final concentration) and the solution containing the peptides was stored at – 20 °C until further analysis.

Biotinylation of the bacterial surface

The surface-exposed proteins from *C. canimorsus 5* strain were biotinylated with the "Pierce Cell Surface Protein Isolation Kit" with adaptation of the protocol. *Cc5* bacteria were grown on SB plates and then on HEK293 cells exactly as described here above. The bacterial pellet was gently suspended in 10mM Hepes (pH 7.4), washed twice with 10mM Hepes (pH 7.4) and resuspended in 10 ml of 10mM Hepes (pH 7.4). Since biotin can be taken up by *Flavobacteria* [103]**,** bacteria were first incubated with regular biotin (0.2 g/L) in order to saturate the transport systems. After 5 min Sulfo-NHS-SS-Biotin (0.02 g/L) was added. After 2 min at RT, the reaction was stopped by the addition of 0.5 ml of Quenching Solution (Pierce) and 1ml 10X TBS (pH 7.4). Bacteria were harvested by centrifugation at 5000 g for 10 min at 4 °C, washed twice in TBS (pH 7.4) and then lysed in 1mL according to the manufacturer's protocol. The bacterial lysate was then cleared by centrifugation at 16000g for 10 min at 4 °C and the labeled proteins were immobilized on the NeutrAvidin Gel according to the manufacturer's protocol. Finally the bound proteins were released by incubating the resin with SDS-PAGE sample buffer containing 50 mM DTT.

Identification of the main metabolic end product in Cc5 culture supernatants

Cc5 were grown (24 h) in the presence of murine macrophages (Raw 264.7) in Dulbecco's modified eagle medium supplemented with 1 mM Na-Pyruvate and 10% v/v fetal calf serum. Medium was collected and the bacteria pelleted by centrifugation (5 minutes, 15000 rcf, 4 °C). 0.1% NaN_3 was added to the supernatant and pH adjusted to 7.5 with phosphate-buffer (500 mM, pH 8). The medium was finally passed through a 0.22 um filter and a 3 kDa cut-off filter (Vivaspin, Sartorius). NMR samples were prepared from 400 µl of this medium by adding 5% D_2O and placed into 5 mm standard NMR tubes. NMR measurements were carried out at 24 ˚C on a Bruker Avance DRX 600 spectrometer equipped with a triple resonance pulse field gradient probe. 1D proton NMR spectra were recorded with the excitation sculpting scheme (pulseprogram zgesgp in the standard Bruker library) as described previously [100] achieving water suppression by gradient dephasing of the water resonance. The proton carrier was set to the water frequency for solvent suppression. Spectra were recorded with 57344 complex points and acquisition times of 1.99 seconds. The total experimental time was 3 minutes and 26 seconds for the accumulation of 64 transients. Spectra were processed and evaluated using the software Topspin 2.1.6 (Bruker).

Mutagenesis and allelic exchange

Mutagenesis of *Cc5* Wt has been performed has described in ref [82] with slight modifications. Briefly, replacement cassettes with flanking regions spanning approximately 500 bp homologous to direct PULs framing regions were constructed with a three-fragment overlapping-PCR strategy. First, two PCRs were performed on 100 ng of of *Cc5* genomic DNA with primers A and B (*c.f. Appendix*) for the upstream flanking regions and with primers C and D for the downstream regions. Primers B and C contained 20 bp of sequence homology to the *ermF* insertion cassette. The *ermF* resistance cassette was amplified from pMM106 with primers 5502 and 5503. All three PCR products were cleaned and then mixed in equal amounts for PCR using Phusion polymerase (Finnzymes). The initial denaturation was at 98°C for 2 min, followed by 12 cycles without primers to allow annealing and elongation of the overlapping fragments (98°C for 30 s, 50°C for 40 s, and 72°C for 2 min). After the addition of external primers (A and D), the program was continued with 20 cycles (98°C for 30 s, 50°C for 40 s, and 72°C for 2 min 30 s) and finally 10 min at 72℃. Final PCR products consisted in PUL::*ermF* insertion cassettes and were then digested with *Pst*I and *Spe*I for cloning into the appropriate sites of the *C. canimorsus* suicide vector pMM25. Resulting plasmids were transferred by RP4-mediated conjugative DNA transfer from *E. coli S17-1* to *C. canimorsus 5* to allow integration of the insertion cassette. Transconjugants were then selected for presence of the *ermF* cassette, checked for sensitivity to cefoxitin and the deleted regions were sequenced.

Fetuin deglycosylation analyses and lectin stainings

Bacteria were collected from blood agar plates and resuspended in PBS at OD_{600}=1. 100 µl of bacterial suspensions were then incubated with 100 µl of a fetuin solution (0.1 $g.l^{-1}$) for 120 minutes at 37℃. As negative control, 200 µl of 1:2 diluted fetuin solution alone was incubated for 120 minutes at 37℃. Samples were then centrufiged for 5 min at 13000 RCF, supernatant collected and 3 µl (and 12 µl SDS buffer) were loaded in a 12% SDS gel. Samples were analyzed by immunoblotting (Fetuine, Rabbit anti-Bovine RIA, UCBA699/R1H, ACCURATE CHEMICAL & SCIENTIFIC CORPORATION) and lectin stainings were performed with *Sambucus nigra* lectin (SNA) and *Datura stramonium* lectin (DSA) according to manufacturer recommendations (DIG Glycan Differentiation Kit, 11210238001, Roche).

Outer Membrane Protein purification

Bacteria were collected from blood agar plates and resuspended in 3ml ice Cold HEPES 10mM (pH7.4) at OD_{600}=1. Bacterial suspensions were then sonicated on ice until they turned clear and spined at 15600g for 2 minutes at 4℃. Supernatants were transferred and centrifuged again for 30 minutes at 15600g at 4℃. Pellets were resuspended in 2 ml HEPES 10mM with 1% sarkosyl and Incubated at room temperature for 30 minutes. Finally, samples were centrifuged at 15600g for 30 min at 4℃ and pellet resuspended in 0.1 ml HEPES. Samples were checked for quality and quantity on silver stained SDS-PAGE and analysed by MS/MS.

Survival and growth in human serum

Bacteria were harvested by gently scraping colonies off the blood agar surface, washed twice (5000g for 7 min) and resuspended in PBS to an OD_{600} of 0.2. Normal human serum (NHS) from healthy volunteers was pooled, aliquoted, and stored at -80℃. Serum was heat-inactivated at 56℃ for 2h. A total of 10^7 bacteria were incubated in 1 ml of 10% NHS in PBS with or without 0.005% GlcNAc (w/v) at 37℃ in a heating block. Serial dilutions were plated onto blood plates, and viable colonies were counted after 48h of incubation in a humidified atmosphere supplemented with 5% CO_2 at 37℃.

Mice and tissue cage infection model

12 week-old male C57BL/6 mice were maintained under pathogen-free conditions in the Animal Facility of the Department of Research, University Hospital Basel. Animal experiments were performed in accordance with the guidelines of the Swiss veterinary law. Teflon tissue cages were implanted subcutaneously in the back of anesthetized mice as previously described [104]. The cages consisted of closed Teflon cylinders (10 mm diameter, 30 mm length, internal volume 1.84 ml) with 130 regularly spaced 0.2 mm holes.

2 weeks after surgery, 200 µl of bacterial suspension was injected percutaneously into the cage. Prior to infection, sterility of the tissue cage was verified. Tissue cage fluid (TCF) was sampled at day 2, 5, 7, 14, 21 and 28 and examined for leukocytes and bacterial viable counts. Leukocytes from TCF were quantified with a Coulter counter (Coulter Electronics). Survival of *Cc5* mutants in the competition experiments were directly compared with wt Cc5 in individual animals giving a 1:1 ratio of wt to mutant bacteria. The number of mutant (Em resistant) and wt bacteria recovered from the TCF of animals was established by plating to media with and without Em. The competitive index was calculated as the (number of mutant/wild-type bacteria recovered from animals)/(number of mutant/wild-type bacteria in the inoculum).

Acknowledgments

We thank Philip Engel for critical reading of the manuscript and assistance in the phylogenetic analysis and Stephanie Wagner for graphical design assistance.

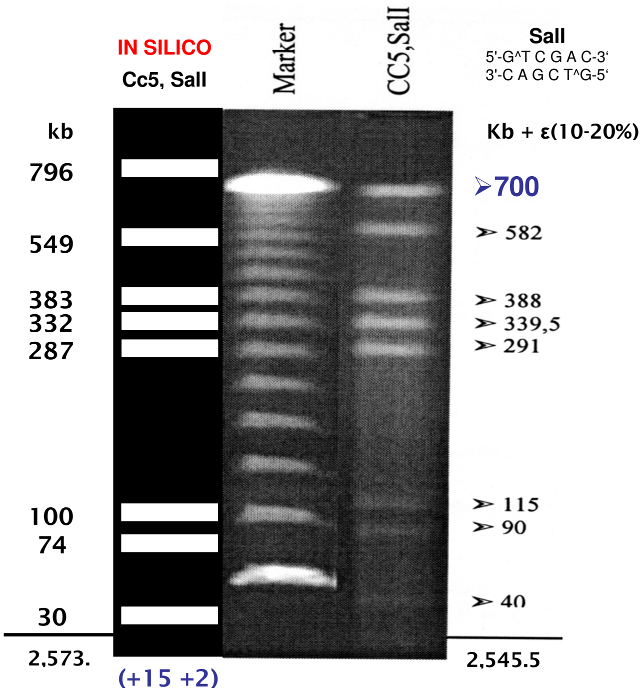
This work was supported by the Swiss National Science Foundation (grant 3100A0-128659/1).

4.2. Additional data

4.2.1. Genome assembly and restriction fragment profile

Genome assembly quality has been assessed by comparing *in silico* predicted restriction profile of the chromosomal sequence by a rare cutter SalI to the actual *in vitro* complete restriction reaction. As represented on **Figure 4.2.1**, *in silico* length are in the experimental tolerance error range (ε = 10-20%) of the observed values. In addition, tow short fragments of 15 and 2 kb were out of the focus of the pulsed field gel electrophoresis (PFGE).

Figure 4.2.1 *in silico* versus *in vitro* restriction profiles of the *Cc5* genome



SalI restriction and PFGE performed by Stephan C. Schuster.

4.2.2. Semi automated genome annotation pipeline

Genome annotation has been performed in a semi automated way using a set of in-house Perl scripts presented on this chapter (**Figure 4.2.2** and supplementary data, *Chapter_4.2_In_House_Scripts* folder). Perl scripts were used to loop single gene analysis software over the whole genome by using local CPU or the BC2 CPU cluster if split work was considered as beneficial (ref.BC2).

First step: open reading frames and coding sequences identification.

BaseCount.pl gives an overview of the assembly file (nucleotides statistics and contigs statistics). *SerialGlimmer3.pl* Integrates the CDS predictor (or gene finder) GLIMER.3 into a loop considering all contigs from a multiple fasta file of an incomplete draft assembly (it output a single file per contig). *Translator.p t*ranslates the Multifasta file of CDS in a protein multiple fasta file. *Super_script_For_Alternative_CDS_Determination.4.pl* has been used to redefine N-terminal boundaries of the genes predicted by GLIMMER as briefly discuss in **chapter 4.1.**

Parallel run of several functional prediction programs.

WWW_InterProScan_PsiBlast_Annotation.pl connects to the European Bioinformatics Institute (EBI) server at http://www.ebi.ac.uk/Tools/InterProScan/ and submits a certain number of concomitant jobs to the InterProScan domain analysis meta-search tool [90]. Each submission corresponds to a single gene and is monitored by a single job in a specific BC2 cluster nod. The number of jobs submitted to InterProScan server is intentionally limited to avoid overloads or queuing issues at the EBI. The script finally generates a single file per sequence with the identified profiles, amino acid coordinates, the name of the software and the databases hitting the current gene with additional cross-references.

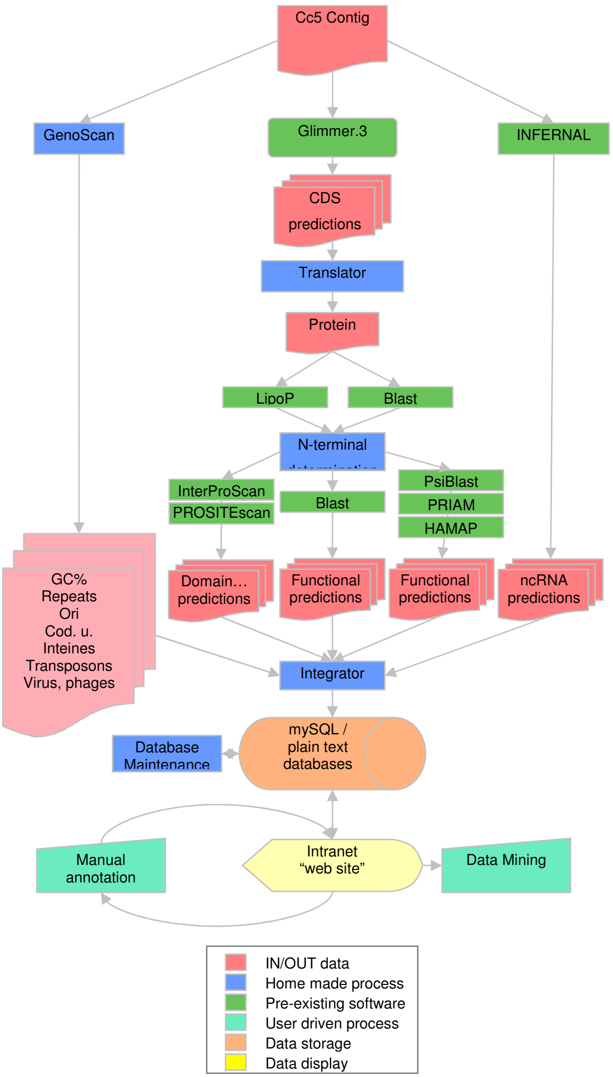
BC2_BlastP_Annotation.pl is used to Psi-blast translated genomes to different databases with the previously reported strategy (*c.f.* chapter 4.1).

20100630_BC2_INFERNAL_Annotation have been designed to optimize genome analysis by the fastidious ncRNA detection software INFERNAL. The script submits a chromosome screening run with each existing model of the RFAM database [93] to a different cluster nod.

Data handling, storing and querying:

Most programs need special input formatting in order to be correctly processed. For this reason, parsing scripts were also created for almost all input or output files used during genome annotation (*e.g. MakeListe.pl, PARSE_.raw_InterProScan_files.pl, PARSE_.XML PsiBlast_files.pl, Fasta2RawTab.pl, PARSE_IntProSca_4_GO.pl*). A MySQL and a plain text database were built to store such generated data. PHP scripts were used for MySQL database management and querry outputting (see supplementary data, *Chapter_4.2_In_House_Scripts* folder). Plain text database has been handled with integrative Perl scripts that fetch data from different data sources (tab delimited or plain text files) (supplementary data, *Chapter_4.2_In_House_Scripts* folder). In addition, a series of html files have been generated with CGview [105] and represent the *Cc5* chromosome with several annotations an interactive display of the functional characterization of CDSs or ncRNAs (limited overview in supplementary data, *Chapter_4.2_In_House_Scripts* folder, Cc5_Chromosome, index.html).

Figure 4.2.2 The annotation pipe



A single consensual sequence is used as starting point.

Coding sequence & ncRNA
An intrinsic method (Glimmer.3) is used to predict coding sequences (CDSs) on genomic DNA. The whole genome is screened by INFERNAL for each non-coding RNA model from the RFAM library. Additional features are directly calculated from the genomic sequence (here, termed GenoScan and mostly supported by EMBOSS package. e.g. Pepstat).

Functional prediction
Each single CDS is translated and undergoes classical functional analysis (InterproScan/Blast/PsiBlast…). Main protein databases (Nr, TrEMBL, KEGG…) as well as the full InterProScan library are used as data providers. Alternative start codons are also considered during this stage.

Sorting, storing, filtering…
An Integrator software is used to collect previously retrieved data and to unify formats in order to integrate it into Cc5 tailored databases. This goes along with a database maintenance tool that updates annotation data on demand.

Display & manipulation
The database has an intranet accessible web-site. This graphical interface was designed in order to provide a manual curation tool and an efficient way to query the databas.

4.2.3. Genome scale metabolic modeling

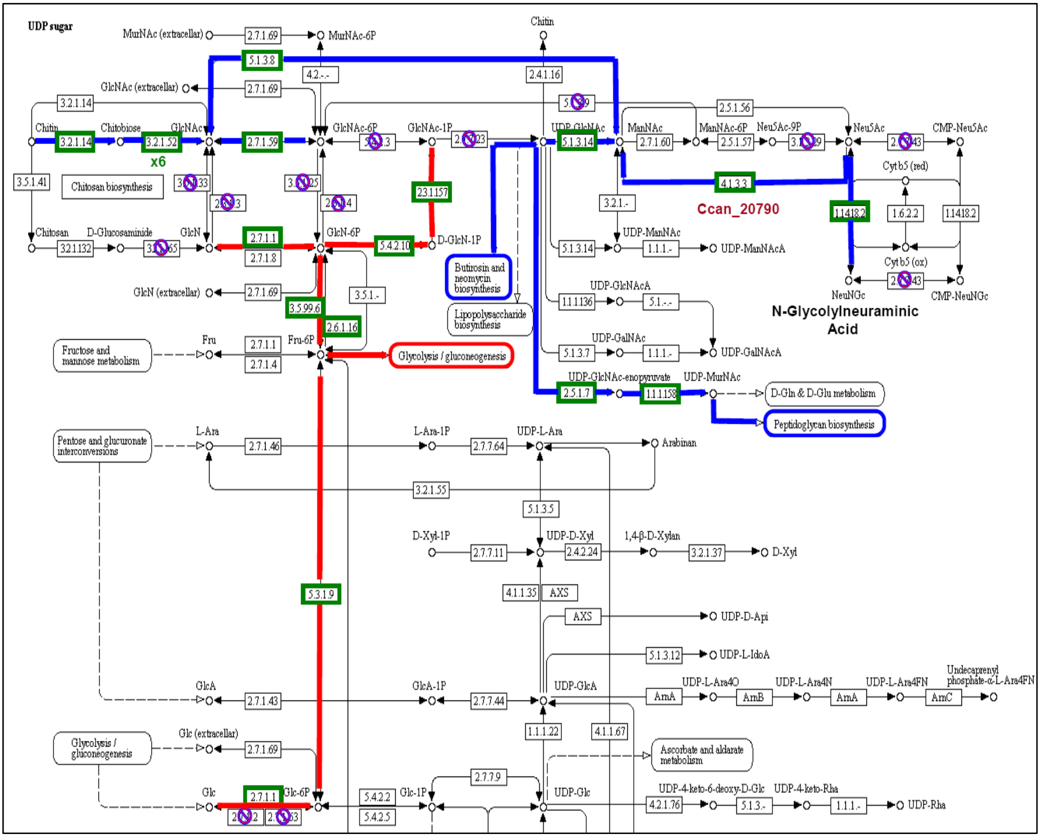
Development of an organism-specific genome scale metabolic databases has been performed with the Pathway Tools package v14.0 [106] and a quick manual curation applied. The software used annotation information (EC number predictions mainly produced by PRIAM and Blast analysis against the Swiss-Prot database). The local database considers 1597 enzymatic reactions, 771 enzymes and 64 transporters out of the 2414 proteins encoded by the *Cc5* genome. Twenty tRNA amino acid ligases were detected and most genes involved in amino acid synthesis were present with the exception of the histidine biosynthesis pathway that was lacking most part of it. When compared to well characterized metabolic schemes from other bacteria (*Agrobacterium tumefaciens C58, Bacillus anthracis Ames, Bacillus. Subtilis subtilis 168, Caulobacter crescentus CB15, Escherichia coli CFT073, Escherichia coli K12, Escherichia coli O157:H7 EDL933, Francisella tularensis subsp. tularensis SCHU S4, Helicobacter pylori 26695, Mycobacterium tuberculosis CDC1551, Mycobacterium tuberculosis H37Rv, Plasmodium. Falciparum 3D7, Shigella flexneri 2a str. 2457T* and *Vibrio cholerae O1 biovar eltor str. N16961*), as expected, the most conserved pathways are the nucleotide and nucleoside biosynthesis pathway together with the glycolysis, the fermentative pathway, a partially conserved split TCA cycle (variation IV) and the pentose phosphate pathways (**Supplementary data, Chapter_4.2_Additional_data, Fig. S4.2.3**). Several genes did encode enzymes with odd activities like members of the mevalonic acid biosynthesis pathway (*Ccan_15750-15760, Ccan_08140*), a high number of enzymes possibly involved in mycolate biosynthesis (**Supplementary data, Chapter_4.2_Additional_data, Fig. S4.2.3**), enzymes involved in putrescine biosynthesis (*Ccan_14980* and *Ccan_15000*), all specific genes requiered for the autoinducer AI-2 production (*Ccan_20040* and *Ccan_17230*), a glucuronosyltransferase (*Ccan_1938*), or enzymes involved in UDP-D-xylose, UDP-D-galacturonate and CMP-N-glycoloylneuraminate biosynthesis. However no *Cc5* specific coherent pathway has been identified by this mean.

Interestingly, metabolism analysis also suggests the presence of an uncoupled metabolism for glucose and N-Acetylglucosamine (**Fig. 4.2.3**) and this is currently supported by previous works [7]. In one hand, glucose fermentation has been reported for most *C. canimorsus* strains tested by Brenner *et al.* and glucose utilization by *Cc5* has been confirmed by formazan assays in G.R. Cornelis' lab (L. Sauteur, master thesis). In another hand, ΔPUL5 *C. canimorsus* bacteria are unable to grow in glucose-rich medium and this growth defect can be fully rescued by addition of N-Acetylglucosamine even at low concentrations (226 µM) (L. Sauteur, master thesis). All these suggest that *C. canimorsus 5* may have split its amino sugar metabolic pathway in an energy providing route (*e.g.* Glycolysis) and a structural biosynthesis route (*e.g.* LPS or peptidoglycan synthesis). Split metabolic pathways tend to reduce metabolic redundancy and to increase the number of compounds required by the bacterium for growth. Consequently substrates are restricted to more specialized purposes (*i.g.* Hexoses for energy, N-Acetylhexoses for structural biosynthesis). Similarly, in the spit TCA cycle of *C. canimorsus 5*, the CO_2 dependent carbon integration route that feeds bacterial respiration with fumarate might be uncoupled to the energy providing side of the TCA cycle (the acetate forming path). In concordance with the relatively reduced genome size of *Cc5*, all these observations may illustrate a reduced metabolic pleiotropy. In such case, the environment has to provide certain amounts of multiple indispensable substrates that *Cc5* is not be able to synthesize through alternative resources. Thus, dependence on a rich and homeostatic environment would suggest a specific bacterial adaptation to a host associated lifestyle.

In addition, the genome scale metabolic database is a fundamental tool to draw accurate observation as for the previously described respiratory model initially derived from the data presented here. It also gave initial input in the identification of the LPS biosynthesis pathway (S. Ittig, unpublished).

Figure 4.2.3 Amino sugar metabolism

Modified from map00520, 08/05/2010, Kanehisa Laboratory, http://www.genome.jp/kegg/pathway/map/map00520.html.



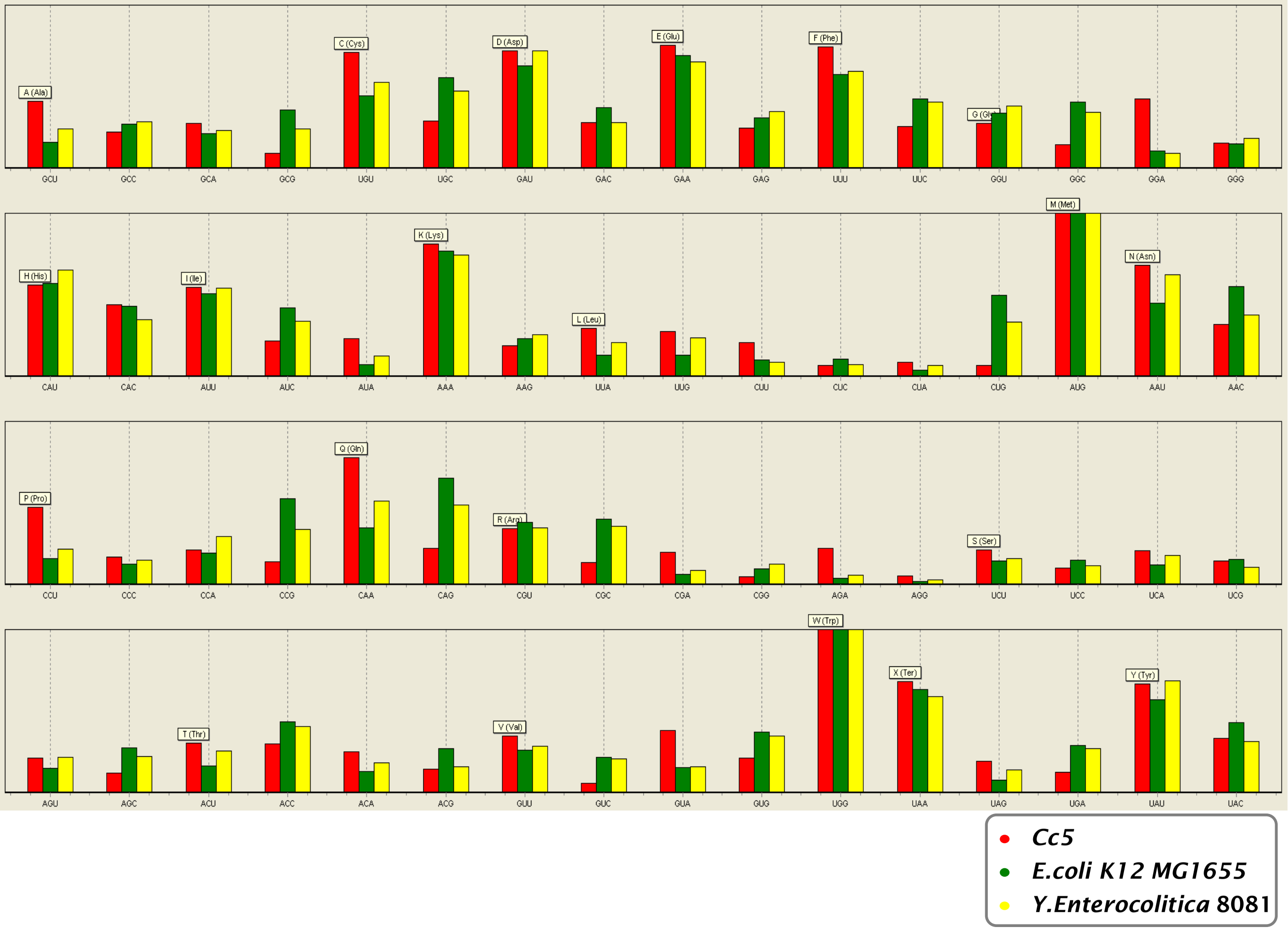
Red and blue lines respectively represent *Cc5*'s glucose and N-Acetylglucosamine pathways. Enzymatic activities predicted in *Cc5* are framed in green. Absence of enzymatic connection is stressed by blue and red symbols.

4.2.4. Genomic codon usage analysis

A genomic codon usage analysis is the assessment of the codon frequency of each amino acid in a given genome. Each gene is then represented by a set of frequencies that can be viewed as an evolutionary hallmark. Optimal codon usage for a given organism is extrapolated from the frequency profiles of a highly conserved set of genes and is therefore a good marker of its vertical evolution. Profiles clustering enables then to group genes that shows common evolutionary features. Such features may depend on different factors as high expression levels that increase selection pressure on certain (important) genes and tend to shift codon usage gene profiles to the optimal one for the considered organism. Inversely, genes that strongly differ in their codon usage from the rest of the genome (or compared to a set of conserved genes) are interesting candidates for recent horizontal gene acquisitions (*i.e.* until recently, under a different codon usage pressure) or pseudogenes (loose of codon usage pressure).

In the present work, another DNA bias analysis performed with Alien hunter (*c.f.* chapter 4.1) out-competes performances of a simple codon usage clustering or a third codon nucleotide analysis [96] (data not shown). However, difficulties encountered during heterologous expression of either *Cc5* functional proteins in *E. coli BL21* or fluorescent proteins in *Cc5* (namely GFP) motivated the identification of the *Cc5* specific codon usage. **Figure 4.2.4** has been generated with INCA [107] and shows the average codon frequencies of *Cc5*, *E.coli K12 MG1655* and *Yersinia enterocolitica 8081* genomes. Obvious discrepancies can be observed for codons encoding alanine, cysteine, glycine, leucine, isoleucine, proline, glutamine, arginine and valine. However expression trials of cytoplasmic GFP and mCherry protein indicated that despite substantial expression of both fluorescent proteins (observed on Coomassie stained SDS-PAGE gels), only mCherry exhibited limited fluorescence levels. Besides, it is known that GFP is more sensitive to oxidative stress during its folding compare to mCherry [108] suggesting that heterologous expression difficulties may originate from a proteins folding incompatibility rather than from protein expression issues.

Figure 4.2.4 Compared average codon frequencies of *Cc5*, *E.coli* and *Y.Enterocolitica* genomes



5. The polysaccharide utilization locus 5

The N-glycan glycoprotein deglycosylation complex (Gpd) from *Capnocytophaga canimorsus.*

5.1. Publication

The N-glycan glycoprotein deglycosylation complex (Gpd) from *Capnocytophaga canimorsus* deglycosylates human IgG

Francesco Renzi, Pablo Manfredi, Manuela Mally, Suzette Moes, Paul Jenö,
Guy R Cornelis †

Biozentrum der Universität Basel, Basel, Switzerland
†. Corresponding author

Statement of authors' work.

PM performed bioinformatics analysis.

FR performed mutagenesis and complementation of PUL5 genes with the help of **PM**.

FR performed fetuin and IgG deglycosylation experiments.

MM performed the lipoproteins radioactive labeling experiments.

FR performed lipoproteins localization experiments.

PM performed the GpdCDEFG complex copurification and mass spectrometry analysis.

FR performed the SiaC copurifications and the mass spectrometry analysis.

SM and PJ performed all mass spectrometry experiments.

FR performed cell culture growth assays with the help of **PM**.

ABSTRACT

C. canimorsus 5 has the capacity to grow at the expenses of glycan moieties from host cells N-glycoproteins. Here, we show that *C. canimorsus 5* has also the capacity to deglycosylate human IgG and we analyze the deglycosylation mechanism. We show that deglycosylation is achieved by a large complex spanning the outer membrane and consisting of the Gpd proteins and sialidase SiaC. GpdD, -G, -E and -F are surface-exposed outer membrane lipoproteins. GpdDEF contribute to the binding of glycoproteins at the bacterial surface while GpdG is a β-endo-glycosidase cleaving the N-linked oligosaccharide after the first N-linked GlcNAc residue. GpdC, resembling a TonB-dependent OM transporter is presumed to import the oligosaccharide into the periplasm after its cleavage from the glycoprotein. The terminal sialic acid residue of the oligosaccharide is then removed by SiaC, a periplasm-exposed lipoprotein in direct contact with GpdC. Finally, degradation of the oligosaccharide proceeds sequentially from the desialylated non reducing end by the action of periplasmic exoglycosidases, including β-galactosidases, β-N-Acetylhexosaminidases and α-mannosidases.

AUTHOR SUMMARY

Capnocytophaga canimorsus are Gram-negative bacteria from the normal oral flora of dogs and cats. They cause rare but severe infections in humans that have been bitten or simply licked by a dog or cat. Fulminant septicemia and peripheral gangrene with a high mortality are the most common symptoms. A surprising feature of these bacteria is their capacity to feed by foraging the glycan moieties of glycoproteins from animal cells, including phagocytes. Here we show that *C. canimorsus* can also deglycosylate human IgGs reinforcing the idea that this property of harvesting host glycoproteins may contribute to pathogenesis. We also unravel the complete deglycosylation system which belongs to a large family of systems devoted to foraging complex glycans, found exclusively in the *Capnocytophaga-Flavobacteria-Bacteroides* group, and whose archetype is the starch harvesting system Sus. It is the first system devoted to deglycosylation of glycoproteins to be characterized.

INTRODUCTION

Capnocytophaga are capnophilic Gram negative bacteria that belong to the family of *Flavobacteriaceae* in the phylum *Bacteroidetes* and colonize the oral cavity of diverse mammals including humans [14, 81]. *Capnocytophaga canimorsus*, a usual member of dog's mouths flora [50, 51] was discovered in 1976 [6] in patients that underwent dramatic infections after having been bitten, scratched or simply licked by a dog. These infections occur, worldwide, with an approximate frequency of one per million inhabitants per year. They generally begin with flu symptoms and evolve in a few days into fulminant septicaemia and peripheral gangrene with a mortality as high as 40 % [3, 5, 6, 54, 109]. A few recent observations help understanding the high aggressiveness of *C. canimorsus* for humans. First, *C. canimorsus* are able to escape complement killing and phagocytosis by human polymorphonuclear leukocytes (PMN's) [55, 57]. They also escape detection and phagocytosis by macrophages, which results in a lack of release of pro-inflammatory cytokines [56]. In addition to this passive evasion from innate immunity, 60 % of the strains are able to block the killing of *Escherichia coli* phagocytosed by macrophages [50, 57] and some strains even block the onset of pro-inflammatory signalling induced by an *E. coli* lipopolysaccharide (LPS) stimulus [56]. The molecular bases of these immunosuppressive mechanisms are not understood yet. However, their study led to the serendipitous discovery that the fastidious *C. canimorsus* grow readily upon direct contact with mammalian cells including phagocytes. This property was found to be dependent on a sialidase (SiaC) allowing *C. canimorsus* to harvest amino sugars of glycan chains from host cell glycoproteins [58]. Recently, we reported the complete 2,571,405-bp genome sequence and the surface proteome of strain *Cc5*. Among others, this study unravelled the existence of 13 complex feeding systems encoded by polysaccharide utilization loci (PULs), a hallmark of the *Cytophaga-Flavobacteria-Bacteroides* (CFB) group [73, 74]. The archetype of these systems is the Sus system, pioneered by the laboratory of A. Salyers and allowing *Bacteroides thetaiotaomicron* to forage starch. It is composed of the surface-exposed SusCDEF protein complex [74, 80] and the SusAB periplasmic proteins [71]. SusC resembles a TonB-dependent transporter essential for energy-dependent import of starch oligosaccharides into the periplasm [76] while SusD is a α-helical starch-binding lipoprotein [77, 110][19,20]. SusE and SusF are other surface-exposed lipoproteins that reinforce starch binding [71]. Finally, the outer membrane α-amylase SusG hydrolyses surface-bound starch [77]. *B. thetaiotaomicron* has 88 of these PULs, identified essentially by the presence of a pair of adjacent *susC*-like and *susD*-like alleles. Interestingly, expression of some PULs is upregulated in the presence of mucin O-glycans or glucosaminoglycans (GAGs), indicating that *B. thetaiotaomicron* also forages on host glycans, primarily the O-glycosylated mucin [73] but these glycoprotein foraging systems have not been characterized so far. Although *Streptococcus oralis*, a firmicute from the human oral flora and *S. pneumoniae* have been shown to remove and metabolize N-linked complex glycans of human glycoproteins [111-113], no PUL-encoded N-linked glycan foraging system has been described in detail. Here, we characterize such a system that was discovered recently in *C. canimorsus 5* (*c.f.* chapter 4.1). It is encoded by chromosome locus *PUL5,* accounts for 12% of the *Cc5* surface proteins and it contributes to survival in mice and in fresh human serum. It thus represents a new type of bacterial virulence factor (*c.f.* chapter 4.1). We show that it deglycosylates human immunoglobulins G (IgG), we present a detailed molecular characterization of this N-linked glycoprotein foraging complex and we show its functional relation with sialidase.

RESULTS

Genetic analysis of the *PUL5* locus.

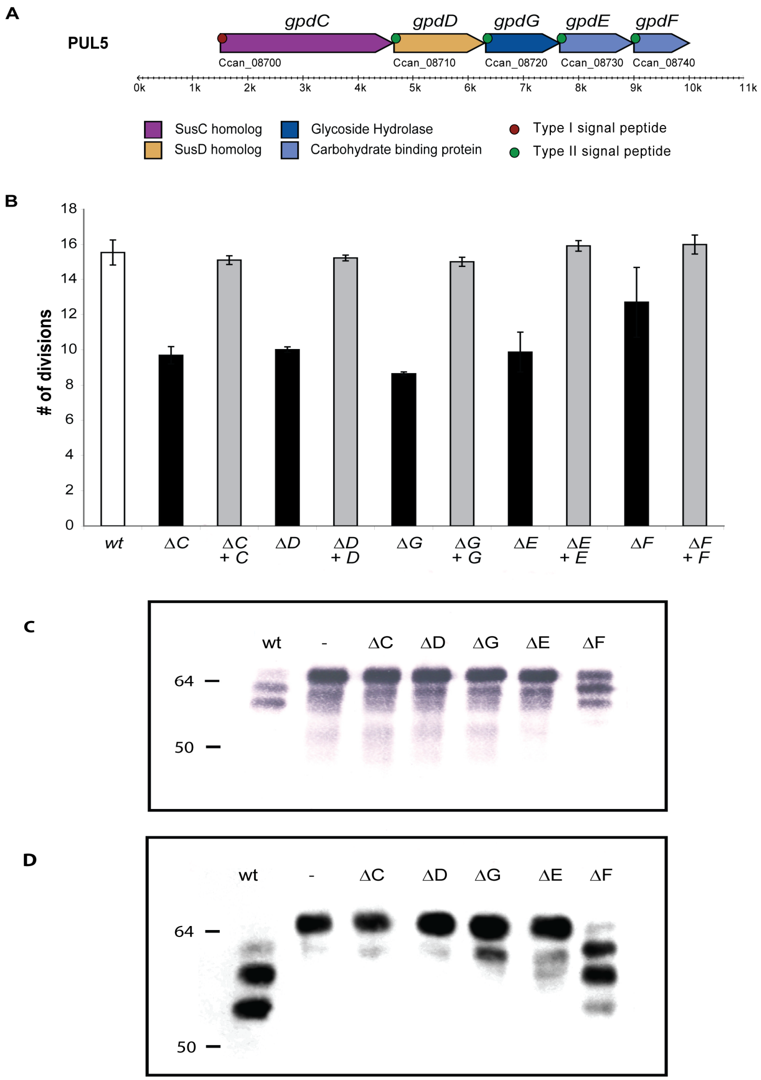
PUL5 consists of the five genes *Ccan*_08700 – *Ccan*_08740. *Ccan*_08700 encodes a SusC-like integral outer membrane (OM) protein presumably forming a pore in the OM while *Ccan*_08710 is a SusD-like protein presumably involved in substrate binding [110]. Since the locus was shown to confer the capacity to deglycosylate proteins (*c.f.* chapter 4.1), we named the five genes *gpd* (for glycoprotein deglycosylation) and we called *gpdC* and *gpdD* the genes encoding homologs to SusC and SusD, respectively. The five *gpd* genes seem to be organized as an operon in the order *gpdC*, *gpdD*, *gpdG*, *gpdE* and *gpdF* (**Fig. 5.1.1A**). GpdG is predicted to be an endo-β-N-acetylglucosaminidase and GpdE has similarities with the Concanavalin A-like lectins/glucanases superfamily on its 108 C-terminal amino acids and could have a substrate-binding role analogous to that of GpdD. Finally, GpdF shows homology to the galactose-binding domain-like superfamily on its 136 C-terminal amino acids suggesting again a role in glycan binding.

In order to investigate what is the function of the individual Gpd proteins we constructed single *gpd* genes knockout strains. None of the knockout mutants was significantly affected in its growth on blood agar plates. In contrast, deletion of any of the *gpdC, -D, -G or -E genes* led to a severe reduction of growth on HEK293 cells while deletion of *gpdF* had only a slight effect **(Fig. 5.1.1.B).** Complementation of the deleted genes with plasmid-borne genes expressed from the natural *gpdC* promoter completely restored growth to the *wt* level indicating that none of the mutation was polar.

In order to determine whether the reduced growth of the mutants was due to a defect in protein deglycosylation, we incubated *wt Cc5* bacteria and the *gpd* mutant bacteria with fetal calf serum protein fetuin, taken as a standard glycoprotein. Fetuin contains 3 O-linked glycans (20 % of the total glycans) and 3 N-linked glycans (80 % of the total glycans)[114]. We monitored glycosylation by staining with *Sambucus nigra* agglutinin (SNA), a lectin that recognizes terminal sialic acids on glycans. As shown in **Fig.**

5.1.1.C, fetuin that had been incubated with *wt Cc5* reacted much less with SNA and appeared as two, still sialylated smaller degradation products. This indicated that partial deglycosylation had occurred and progressed further than a simple desialylation. In contrast, fetuin that was incubated with the *gpdC, -D, -G* and *-E* mutant bacteria was unaffected, indicating that no desialylation occurred in the absence of these *gpd* genes, although sialidase SiaC [58] was unaffected. Fetuin incubated with the *gpdF* mutant showed a slight desialylation indicating that fetuin deglycosylation was not completely abolished as with the other mutants. Fetuin glycosylation was also monitored by immuno-blotting with anti-fetuin antibodies. As shown in **Fig. 5.1.1.D**, the size of fetuin was shifted down after incubation with wt *Cc5* bacteria while the protein migration rate was unchanged after incubation with the *gpdC, -D, -G* and *-E* mutant bacteria. After incubation with *gpdF* mutant bacteria, fetuin did undergo a size shift but not as important as when incubated with wt bacteria. Taken together these results indicate that partial fetuin deglycosylation was strictly dependent on the activity of proteins GpdC, -D, -G, -E and, to a lesser extend -F. Finally, our data strongly suggest that the defect in growth of the *gpd* mutants onto HEK293 cells was completely due to a defect in the ability to deglycosylate host glycoproteins.

Figure 5.1.1. Genetic analysis of the PUL5 locus



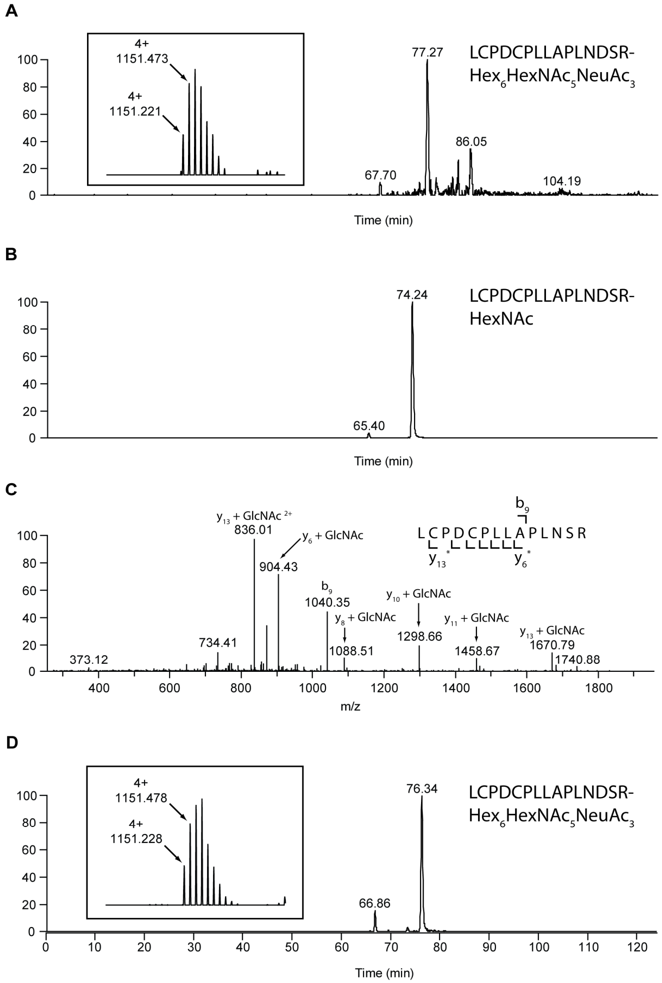
(A). Schematic representation of the PUL5 putative operon (top: new gene designation; below: gene codes derived from the annotation of the genome (*c.f.* chapter 4.1).
(B). Growth of the various individual *gpd* knockout (black) and complemented (grey) mutants on HEK293 cells (moi = 0.2; 23 hours growth).
(C). Glycosylation state of fetuin samples incubated for 3 hours in the presence of the different strains, monitored by staining with SNA that recognizes terminal sialic acid (2-6 or 2-3) linked to Gal or to GalNAc
(D). Western blot analysis with anti-fetuin antibodies of fetuin samples incubated as in (C).

GpdG is an endo-β-N-acetylglucosaminidase.

GpdG is annotated as an endo-β-N-acetylglucosaminidase (*c.f.* chapter 4.1), *i.e* an endo-glycosidase that cleaves N-linked glycan structures at the base of the glycan in between two GlcNAc molecules. Hence, it should leave one GlcNac molecule attached to the protein. Fetuin is reported to be glycosylated on the three asparagine residues Asn99, Asn156 and Asn176 [114]. Analysis by liquid chromatography-mass spectrometry (LC-MS) of trypsin-digested fetuin showed that the main glycosylation site resides on Asn156 and bears a sugar with a $Hex_6HexNAc_5NeuAc_3$ composition (**Fig. 5.1.2.A**). Asn176 was found to carry a sugar with a Hex6HexNAc5NeuAc4 composition, but its site occupancy was much lower than Asn156. Only trace amounts of glycans were found attached to Asn99. After incubation of fetuin with wt *Cc5* bacteria, LC-MS analysis revealed the presence of a peptide whose mass indicated that only one HexNAc moiety remained linked to Asn156 (**Fig. 5.1.2.B**). The fragmentation spectrum of this peptide fully confirmed the presence of the HexNAc moiety on Asn156 (**Fig. 5.1.2.C**). Due to the low site occupancy of Asn176, deglycosylation of Asn176 to the HexNAc moiety was too weak to be detected. The conversion of $Hex_6HexNAc_5NeuAc_3$ to HexNAc on Asn156 suggests an endo-β-N-acetylglycosidase dependent deglycosylation.

To confirm that fetuin deglycosylation was due to the Gpd complex activity and in particular to the GpdG glycosyl hydrolase activity, we then analysed fetuin after incubation with the *gpdG* knockout bacteria. Fetuin incubated in the presence of these mutant bacteria turned out to remain fully glycosylated (**Fig. 5.1.2.D**) indicating that no cleavage occurred in the absence of the enzyme.

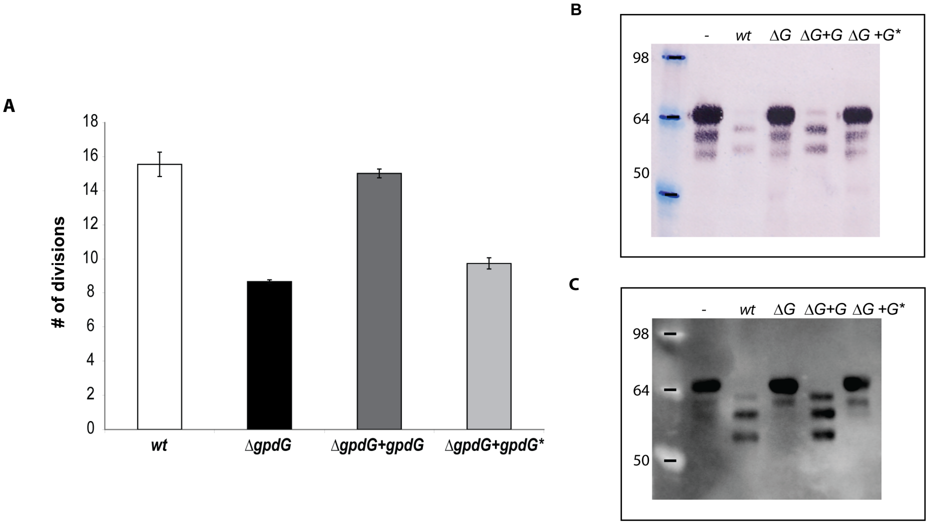
Figure 5.1.2. LC-MS analysis reveals an endo-β-N-acetylglucosaminidase activity of GpdG.



Glycosylation analysis of fetal calf serum fetuin. (A). Asn156 glycosylation of untreated bovine fetuin. Selected ion chromatogram for the quadruply charged tryptic peptide carrying the $Hex_6HexNAC_5NeuAc_3$ glycosyl moiety on the LCPDCPLLAPLNDSR peptide. The inset shows the isotope pattern for the Asn156 glycopeptide. (B). Selected ion chromatogram for the doubly charged Asn156 HexNAc-modified LCPDCPLLAPLNDSR glycopeptide of fetuin that had been incubated with wild-type *Cc5*. (C) Fragmentation spectrum of the Asn156- GlcNAc species with the y- and b-ions that conclusively show the HexNAc modification of Asn156. (D) Asn156 glycosylation of bovine fetuin that had been treated with the *ΔgpdG* strain. Selected ion chromatogram for the quadruply charged tryptic peptide carrying the $Hex_6HexNAC_5NeuAc_3$ glycosyl moiety on the LCPDCPLLAPLNDSR peptide. The inset shows the isotope pattern for the Asn156 glycopeptide.

The sequence of GpdG was then compared to those of two endo-β-N-acetylglucosaminidases, namely EndoS from *Streptococcus pyogenes* capable of deglycosylating N-linked glycans from the γ chain of human immunoglobulins [115], and EndoF from *Flavobacterium meningosepticum* capable of cleaving off high-mannose and complex glycan N-linked from several glycoproteins including immunoglobulins [116]. It appeared that a chitinase motif present in these two enzymes was conserved in GpdG (FDGFDIDWE). In order to further confirm the endo-β-N-acetylglucosaminidase activity of GpdG we substituted the essential E205 residue [116] with a glycine and tested the growth on HEK293 cells of the *gpdG* mutant strain expressing in trans the GpdG catalytic mutant. As shown in **Fig. 5.1.3.A**, the GpdG catalytic mutant was impaired in growth. We then tested the fetuin deglycosylation ability of the GpdG catalytic mutant. As shown by the lectin staining in **Fig. 5.1.3.B** and by the immuno-blotting in **Fig. 5.1.3.C**, bacteria endowed with the GpdG catalytic mutant were completely impaired in fetuin deglycosylation. We conclude from all these experiments that GpdG is an endo-β-N-acetylglucosaminidase.

Figure 5.1.3. The F$_{197}$DGFDIDWE$_{205}$ chitinase motif of GpdG is the catalytic site.

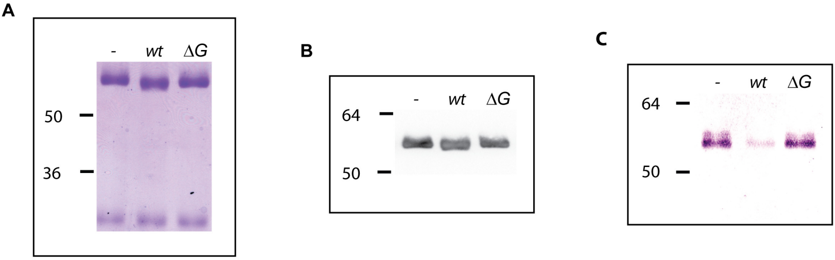


E$_{205}$ from GpdG was substituted with a glycine. (A): Number of divisions after 23 h growth on HEK293 cells of the *ΔgpdG* mutant complemented with gpdG* encoding the catalytic mutant
(B): Fetuin glycosylation state of samples incubated for 3 hours in the presence of the different strains, determined by staining with the *Sambucus nigra* lectin (SNA) that recognizes terminal sialic acid (2-6 or 2-3) linked to Gal or to GalNAc.
(C): same as B after western blot analysis with anti-fetuin antibodies.

The Gpd complex deglycosylates human IgG.

Since GpdG has the same chitinase motif as EndoF and EndoS, known to deglycosylate N-linked glycans from the γ chain of human IgGs *[25,26]*, we tested whether the Gpd complex would also be able to deglycosylate the heavy chain of IgGs. The N297-linked glycan moiety of this chain is biantennary and consists of $Hex_6HexNAc_5NeuAc_2$. Removal of this moiety by EndoS was shown to determine a size shift of ~ 3 KDa [115]. After incubation of purified human IgG with wt *Cc5* bacteria, the molecular mass of the γ chain underwent a slight size shift (**Fig. 5.1.4.A and B**) while the mass of the light chains was unchanged (**Fig. 5.1.4.A**). In contrast incubation with *ΔgpdG* knockout bacteria did not alter the γ chain size indicating that the cleavage was GpdG dependent. To confirm that the size reduction of the γ chain was due to the removal of the glycan moiety, IgG was stained with SNA. As shown in **Fig. 5.1.4.C**, the SNA signal of the γ chain was significantly reduced after incubation with wt *Cc5*. In contrast the γ chains remained fully glycosylated after incubation with *ΔgpdG* bacteria. These data indicated that, like *F. meningosepticum* and *S. pyogenes*, *C. canimorsus* has the capacity to deglycosylate IgGs.

Figure 5.1.4. Human IgG deglycosylation.

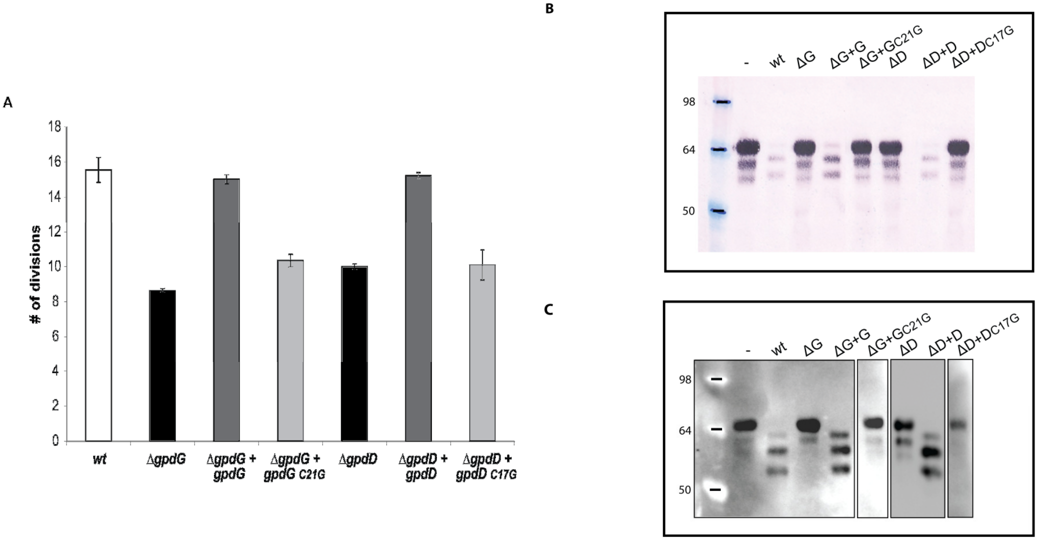


Glycosylation state of human IgG samples incubated for 3 hours in the presence of wt and *ΔgpdG* bacteria monitored by Coomassie staining (A), western blot analysis with anti-IgG antibodies (B) and staining with SNA (C).

GpdD, -G, -E and -F are lipoproteins and lipid modification is fundamental for the complex activity.

The GpdD, -G, -E and –F proteins belong to the OM and surface proteomes of *Cc5* (*c.f.* chapter 4.1). In addition, these proteins are endowed with a signal peptidase II consensus signal peptide. Altogether, this suggests that they could be lipoproteins anchored to the outer leaflet of the outer membrane and exposed at the surface of the bacterium (*c.f.* chapter 4.1). In order to determine whether the lipidation of the Gpd proteins is required for their function, we generated soluble periplasmic versions of GpdD and GpdG by substituting the cystein residue of the lipobox with a glycin. We then tested the ability of the periplasmic variants of GpdD and GpdG to complement the growth deficiency of the *gpdD* and *gpdG* knockout strains on HEK293 cells. As shown in **Fig. 5.1.5,** both the GpdD and GpdG periplasmic variant were unable to complement the growth deficiency indicating that lipid modification is necessary for the proper localization and function of the proteins. This conclusion was reinforced by the fact that bacteria endowed with periplasmic GpdD or GpdG were unable to deglycosylate fetuin **(Fig. 5.1.5).** Hence, we infer that GpdD and GpdG are lipoproteins that are anchored in the outer leaflet of the outer membrane and exposed to the bacterial surface. The same presumably applies to GpdE and GpdF since they have also a lipobox and they are also part of the surface proteome (*c.f.* chapter 4.1).

Figure 5.1.5. Lipid modification of GpdD and GpdG is essential for their activity.



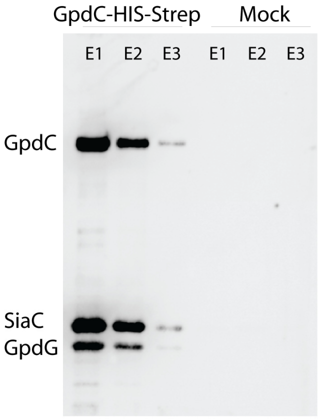
(A) Number of divisions after 23 h growth on HEK293 cells of the *ΔgpdG* bacteria complemented with *gpdD*$_{C17G}$ and *gpdG*$_{C21G}$.
(B) Fetuin glycosylation state of samples incubated for 3 hours in the presence of the different strains, determined by staining with SNA.
(C) Same as B analyzed by western blot with anti-fetuin antibodies.

The Gpd proteins form a deglycosylation complex associated with sialidase.

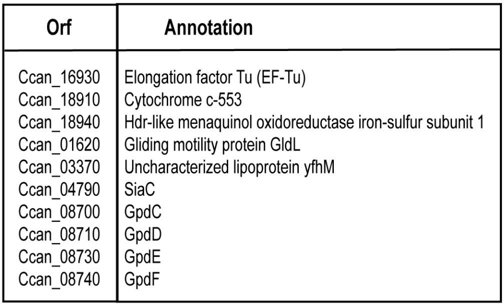
In order to assay whether the five Gpd proteins interact with each other to form a complex at the bacterial surface, we performed a two-step affinity purification with a His-Strep tagged version of GpdC. Analysis by immuno-blot and mass spectrometry **(Fig. 5.1.6)** of the purified fraction revealed the presence, together with GpdC, of GpdD, -G, -E and –F, indicating a stable interaction between all these proteins. Furthermore, six other proteins, among which SiaC (**Fig. 5.1.6)**, co-purified with the complex.

Figure 5.1.6. Gpd proteins form a complex with sialidase

A



B



Orf	Annotation
Ccan_16930	Elongation factor Tu (EF-Tu)
Ccan_18910	Cytochrome c-553
Ccan_18940	Hdr-like menaquinol oxidoreductase iron-sulfur subunit 1
Ccan_01620	Gliding motility protein GldL
Ccan_03370	Uncharacterized lipoprotein yfhM
Ccan_04790	SiaC
Ccan_08700	GpdC
Ccan_08710	GpdD
Ccan_08730	GpdE
Ccan_08740	GpdF

Streptavidine affinity purification of GpdC-His-Strep expressed from its natural promoter in a *ΔgpdC background.* (A) Detection by western blot of GpdC (anti-His antibody), GpdG (anti-GpdG) and Sialidase (anti-SiaC) in the elution fractions. (B) List of protein identified by Mass spectrometry in the elution fractions.

Sialidase is a periplasmic lipoprotein that interacts with GpdC.

SiaC has been previously shown [58] to be essential to sustain growth of *Cc5* in the presence of eukaryotic cells due to its role in the glycoprotein deglycosylation process. We thus focused our attention on the sialidase-Gpd complex interaction. The co-purification of SiaC with GpdC strongly suggested that SiaC is associated to the Gpd complex, although it is encoded far away from PUL5. However, unlike the Gpd proteins, sialidase was not identified in the surface proteome of *Cc5* (*c.f.* chapter 4.1). On the other hand, earlier immunofluorescence assays suggested that sialidase is localized on the bacterial surface and removal of the signal sequence of sialidase prevented growth on cells [58]. In order to better understand the interplay between SiaC and Gpd proteins in the glycoprotein deglycosylation process, we decided to clarify its localization.

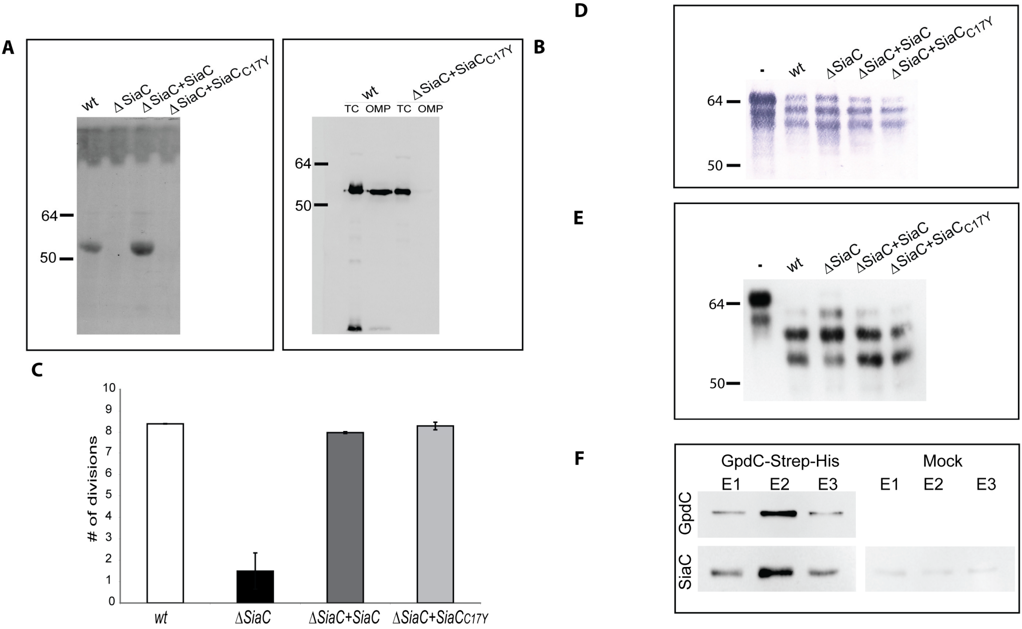
Since the sialidase sequence analysis revealed the presence of a signal peptide with a lipobox in the N-terminal sequence, we first sought to determine whether SiaC is a lipoprotein. We incubated *Cc5* and mutant bacteria encoding SiaC$_{C17Y}$ in the presence of tritiated palmitate and analyzed the total proteins by SDS-PAGE and fluorography **(Fig. 5.1.7.A**). Sialidase appeared indeed to be lipidated and the C17Y mutation completely prevented this lipid modification. The analysis of outer membrane proteins isolated by sarcosyl extraction confirmed that sialidase but not its C17Y variant was associated with the OM **(Fig. 5.1.7.B**). We conclude from these experiments that SiaC is a lipoprotein anchored into the outer membrane.

In order to define whether it is exposed towards the outside like GpdDGEF or towards the periplasm, we tested whether the periplasmic Sia$_{C17Y}$ could restore the growth deficiency of the *siaC* mutant strain. In contrast to what was observed for GpdD and GpdG, expression of Sia$_{C17Y}$ in trans did fully restore the growth defect **(Fig. 5.1.7.C)** indicating that the localization of sialidase in the periplasm and the absence of association with the outer membrane did not prevent its function. This data pointed to the direction of a periplasmic localization of SiaC rather than a surface-exposed localization as was previously suggested [58].

The association between sialidase and the Gpd complex obviously suggests that the two work cooperatively. This was already suggested by the fact that the *gpd* mutant bacteria did not remove the terminal sialic acid residues from fetuin, although SiaC was functional in these mutants **(Fig. 5.1.1.C).** We then tested the ability of the *siaC* knockout bacteria to deglycosilate fetuin. SNA lectin staining **(Fig. 5.1.7.D)** and immuno-blotting **(Fig. 5.1.7.E)** clearly showed the same fetuin deglycosylation pattern for the wt and *siaC* mutant bacteria. These results indicate that the endo-cleavage of fetuin N-glycans, operated by the Gpd complex is completely independent from the activity of SiaC. However, the evidence that SiaC activity is essential for growth on Hek293 cells **(Fig. 5.1.7.C)**, suggests that removal of the glycan terminal sialic acid is nevertheless a crucial step for the subsequent glycan catabolism process. This indicates that the Gpd complex acts upstream of SiaC. Since the Gpd complex includes the GpdC porin-like protein, this sequential order is perfectly compatible with a periplasmic localization of sialidase. Sialic acid removal would thus occur in the periplasm after the glycan has been cleaved off and transported through the GpdC OM channel.

If this model was correct, the interaction between the periplasmic SiaC and the GpdC complex could only occur through a direct interaction with GpdC, since the other Gpd proteins are surface exposed. To test this prediction, we expressed a C-terminally Strep-His double tagged GpdC in a *gpdCDGE* multi knockout strain and we performed a two-step affinity purification of GpdC. The analysis by immuno-blotting **(Fig. 5.1.7.F)** of the fractions eluted after the second purification step showed that SiaC did indeed co-purify with GpdC indicating that SiaC and GpdC do indeed interact directly with each other. The complete deglycosylation complex would thus consist of the surface-exposed lipoproteins GpdDGEF and the periplasm-exposed lipoprotein SiaC, all of them associated to the porin-like GpdC **(Fig. 5.1.7)**.

Figure 5.1.7. Sialidase localization and interaction with GpdC.



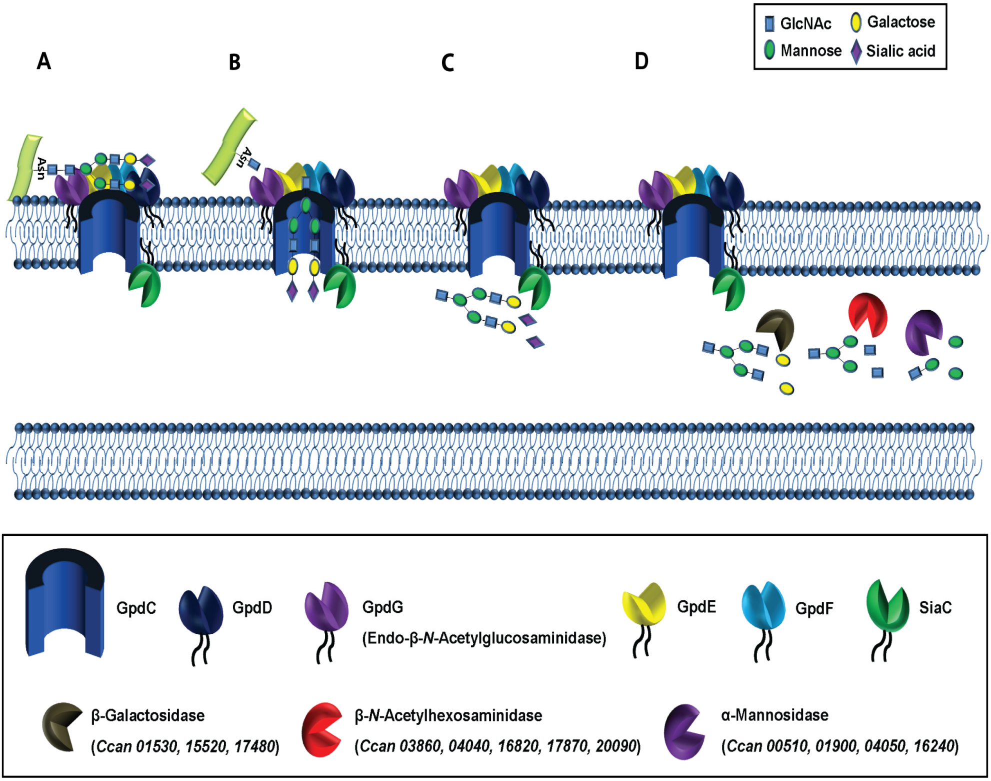
(A) Autoradiography of ^{3}H-palmitate labeled sialidase in different bacteria.
(B) Detection of sialidase by western blot analysis (anti-SiaC antibody) in total cell extracts (TC) and outer membrane protein (OMP) fractions of Cc5 wt and *ΔsiaC* bacteria complemented with the soluble periplasmic sialidase (SiaC$_{C17Y}$).
(C) Number of divisions after 23 hours growth on HEK293 cells of *ΔsiaC* bacteria expressing SiaC or SiaC$_{C17Y}$.
(D) Fetuin glycosylation state after 3 hours of incubation in the presence of the different strains, determined by staining SNA.
(E) Same as D, analyzed by western blot with anti-fetuin antibodies.
(F) Co-purification of SiaC with GpdC-Strep-His produced in a ΔgpdCDGE background. GpdC was detected with anti-Strep antibody and SiaC with anti-SiaC antibodies.

DISCUSSION

Our previous work has shown that *C. canimorsus* deglycosylates surface glycoproteins from the host and sustains its growth on the glycan moieties [58]. Here, we showed that this deglycosylating activity is achieved by the joined action of the PUL5-encoded Gpd complex and sialidase [58]. PUL5 consists of the five *gpdCDGEF* genes. GpdC, an homolog of the archetypal SusC [71] likely represents the specific OM porin of the system. GpdD is an homolog of SusD, a starch-binding protein [16,20] and hence most likely a glycoprotein-binding protein. On the basis of their annotation, we propose that GpdE and GpdF are also glycan-binding proteins. GpdG was annotated as an endo-β-N-acetylglucosaminidase (*c.f.* chapter 4.1) and this annotation was shown to be correct. Indeed mass spectrometry analyses demonstrated that GpdG removes the tribranched complex $Hex_6HexNAc_5NeuAc_3$ glycan structure linked to N156 from the model glycoprotein fetuin, leaving one GlcNac residue to the protein. GpdDGEF were predicted to be lipoproteins (*c.f.* chapter 4.1). Replacement of the critical cysteine of the lipoprotein signal peptide from GpdD and GpdG completely abolished the deglycosylating activity, indicating that a periplasmic location did not sustain the activity. These data, together with the fact that the two proteins belong to the surface proteome indicate that these two lipoproteins are exposed to the surface and not to the periplasm. We assume the same is true for GpdE and F since, like GpdD, they are thought to bind glycans, they contain a lipobox and they belong to the surface proteome. Interestingly, all the five Gpd proteins could be co-purified with the porin-like GpdC, indicating that they all form one single complex at the bacterial surface. Unexpectedly, not only GpdD, -G, -E and -F co-purified with GpdC but also SiaC. Although SiaC was known to be part of the catabolic process, SiaC is not encoded together with GpdCDGEF (*c.f.* chapter 4.1) and it was not anticipated that the interaction would be so close. SiaC turned out to be also a lipoprotein but, unlike GpdD and GpdG, it was still functional when it was directed to the periplasm, unlipidated. We inferred from this observation that, contrary to our initial report, SiaC is a periplasm-oriented lipoprotein. Thus, the observations presented here suggest the model illustrated in **Fig. 5.1.8**: the surface-exposed GpdCDEF complex captures the N-linked complex glycan moieties of glycoproteins, which are then detached from the protein by GpdG and internalized by GpdC. As soon as they reach the periplasm, SiaC removes the terminal sialic acid. This sequence of events is strongly supported by the observation that *gpd* mutant bacteria do not desialylate fetuin, although SiaC is functional in these mutants **(Fig. 5.1.1)**. After desialylation, the oligosaccharide would be sequentially degraded by periplasmic exogalactosidases and the monosaccharides would transferred to the cytosol. This last step of the model is supported by the fact that the genome encodes three putative β-galactosidases (Ccan 01530, Ccan 15520, Ccan17480), five putative β-N-Acetylhexosaminidase (Ccan 03860, Ccan04040, Ccan16820, Ccan17870, Ccan20090) and four putative α-mannosidases (Ccan00510, Ccan01900, Ccan 04050 and Ccan 16220), all of them endowed with a signal peptide I or II, and none of them surface exposed (*c.f.* chapter 4.1). The β-galactosidase and α-mannosidase activities were confirmed in the crude extract (data not shown). The three β-galactosidases seemed actually redundant since they could all be individually knocked out without affecting the growth on cells (data not shown).

This global model strikingly reminds the archetypal Sus system shown to consist of one single complex made of SusCDEF [80]. It is thought that SusG, an *endo*-acting enzyme, generates internal cuts in a bound starch molecule and releases oligosaccharides larger than maltotriose, which are then transported by SusC into the periplasmic compartment. In the periplasm, glycoside hydrolases SusA and SusB then degrade the oligosaccharides into their component sugars prior to final transport to the cytosol [27,28]. The two systems are thus remarkably conserved, although they adapted to different complex saccharides.

Figure 5.1.8. Functional model of complex N-linked glycan moieties deglycosylation processing by *C. canimorsus*.


GlcNAc
Galactose
Mannose
Sialic acid
A
B
C
D
Asn
Asn
GpdC
GpdD
GpdG
(Endo-β-*N*-Acetylglucosaminidase)
GpdE
GpdF
SiaC
β-Galactosidase
(*Ccan 01530, 15520, 17480*)
β-*N*-Acetylhexosaminidase
(*Ccan 03860, 04040, 16820, 17870, 20090*)
α-Mannosidase
(*Ccan 00510, 01900, 04050, 16240*)

Individual glycan processing steps are illustrated. (A) The glycan moiety is bound at the bacterial surface by the Gpd complex. (B) The glycan mopiety is endo-cleaved by GpdG and imported into the periplasm trough the GpdC pore. (C) Terminal sialic acid is cleaved by sialidase (SiaC). (D) The glycan is further processed by the sequencial activity of several periplasmic exoglycosidases.

To our knowledge, the Gpd system is the first PUL-encoded system devoted to foraging N-linked glycoproteins. It contributes to sustain growth of *C. canimorsus* at the expenses of cultured cells (*c.f.* chapter 4.1). Since *C. canimorsus* has 13 PULs (*c.f.* chapter 4.1), it is very likely that some of them could be devoted to the harvest of O-linked glycans, but this activity has not been identified thus far. The best approach would probably be to look for upregulation in the presence of O-linked glycoproteins, as was done in *B. thetaiotaomicron* [74]. Deglycosylation of N-linked glycans is not unprecedented among pathogens and commensals. As mentioned earlier, two *streptococci, S. pyogenes* and *S. oralis* have this remarkable property. In the case of *S. pyogenes*, this activity is exerted towards IgGs by secreted endoglycosidase EndoS and it does not seem to play a major role in nutrient acquisition [115]. In contrast, in *S. oralis*, the activity was shown to sustain growth *[30]*. It is interesting to notice that *S. oralis*, like *C. canimorsus*, is emerging as an important opportunistic pathogen originating from the oral flora. This commonality between two very different bacteria from the same ecosystem suggests first that the capacity to deglycosylate host proteins is a favourable trait in the mouth ecosystem and, second, could favour opportunistic infections. Deglycosylation of IgGs is very likely to contribute to a generalized infection as discussed by Collin and Olsen [115] but, for *C. canimorsus*, one cannot exclude that deglycosylation of other host proteins would also significantly contribute to pathogenesis.

Our data demonstrate that PUL-encoded lipoproteins are surface-exposed. Prolipoproteins are exported through the Sec pathway and then acylated at the periplasmic leaflet of the inner membrane (IM), by the sequential action of glyceryl transferase, O-acyl transferase(s) and prolipoprotein signal peptidase (signal peptidase II). A mature lipoprotein harbours as a first aminoacid a cysteine residue that is lipid modified with a N-Acyl diacyl Glyceryl group which serves to anchor the protein to the IM. In Gram-negative bacteria, some lipoproteins are destined for the OM. These proteins are extracted from the IM, transported across the periplasm and inserted in the inner leaflet of the OM by the Lol pathway (for review see refs [31,32]. Insertion of lipoproteins into the outer leaflet of the OM is however established in some pathogens like *Borrelia* but, the pathway is neither well documented not well understood [117]. Since bacteria from the *Cytophaga-Flavobacteria-Bacteroides* group massively insert lipoproteins in the outer leaflet of the OM, we postulate that they have an original system dedicated to the transport of lipoproteins across the OM but this system still needs to be identified and investigated.

Materials and Methods

Bacterial strains and growth conditions

Conventional bacterial growth conditions and selective agents

The strains used in this study are listed in *Appendix hia coli* strains were routinely grown in LB broth at 37℃. *C. canimorsus* bacteria were routinely grown on heart infusion agar (Difco) supplemented with 5% sheep blood (Oxoid) for 2 days at 37℃ in the presence of 5% CO_2. To select for plasmids, antibiotics were added at the following concentrations: 10 µg/ml erythromycin (Em), 10 µg/ml cefoxitin (Cf), 20 µg/ml gentamicin (Gm), 100 µg/ml ampicillin (Ap) and 50 µg/ml kanamycin (Km).

Growth of *Cc5* bacteria on HEK293 cultured cells

Human Embryonic Kidney 293 cells (HEK293) were cultured in DMEM (Invitrogen) with 10% (v/v) fetal calf serum (Invitrogen) and 1mM sodium pyruvate. Cells were grown in medium without antibiotics in a humidified atmosphere enriched with 5% CO_2 at 37℃. Bacteria were harvested by gently scraping colonies off the agar surface and resuspended in PBS. A total of $4x10^4$ bacteria were incubated with $2x10^5$ HEK293 cells (MOI = 0.2) in a final volume of 1ml medium devoid of antibiotics for 23h.

Mutagenesis and allelic exchange

Mutagenesis of *Cc5* Wt has been performed has described in ref [82] with slight modifications. Briefly, replacement cassettes with flanking regions spanning approximately 500 bp homologous to direct *gpd* framing regions were constructed with a three-fragment overlapping-PCR strategy. First, two PCRs were performed on 100 ng of of *Cc5* genomic DNA with primers A and B (*Appendix*) for the upstream flanking regions and with primers C and D for the downstream regions. Primers B and C contained an additional 5' 20-nucleotide extension homologous to the resistance *ermF* insertion cassette. The *ermF* resistance cassette was amplified from plasmid pMM106 DNA with primers 5502 and 5503. All three PCR products were cleaned and then mixed in equal amounts for PCR using Phusion polymerase (Finnzymes). The initial denaturation was at 98℃ for 2 min, followed by 12 cycles without primers to allow annealing and elongation of the overlapping fragments (98℃ for 30 s, 50℃ for 40 s, and 72℃ for 2 min). After the addition of external primers (A and D), the program was continued with 20 cycles (98℃ for 30 s, 50℃ for 40 s, and 72℃ for 2 min 30 s) and finally 10 min at 72℃. Final PCR products consisted in *gpd::ermF* insertion cassettes and were then digested with PstI and SpeI for cloning into the appropriate sites of the *C. canimorsus* suicide vector pMM25 . Resulting plasmids were transferred by RP4-mediated conjugative DNA transfer from *E. coli* S17-1 to *C. canimorsus 5* to allow integration of the insertion cassette. Transconjugants were then selected for presence of the *ermF* cassette, checked for sensitivity to cefoxitin and the deleted regions were sequenced.

Construction of complementation and expression plasmids

Plasmid pPM1, used for complementation and expression of the Gpd proteins is a derivative of the *E. coli*- *C. canimorsus* shuttle vector pMM47A.1

[82]. pMM47A.1 *ermF* promoter region was cleaved with *Sal*I and *Nco*I and the 117 nucteotides upstream the *gpd*C starting codon sequence, containing the putative *gpdC* promoter, was cloned using the same restriction sites. Full length *gpdC*, *-D*, *-G*, *-E* and *-F* were amplified with the specific primers listed in *Appendix* and cloned into plasmid pPM1 into *Nco*I and *Xba*I restriction sites leading to the insertion of a glycine at position 2.

The E205G substitution inactivating the catalytic site of GpdG was introduced by site directed mutagenesis by overlapping PCR using primers 5008/6061 and 6060/6055 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites leading to plasmid pFR10 (*gpdG**). The C17G substitution of GpdD was introduced by site directed mutagenesis amplifying by PCR using primers 6056 and 6057 and cloning *Nco*I/*Xba*I in pPM1 leading to plasmid pFR8.

The C21G substitution of GpdG was introduced by site directed mutagenesis amplifying by PCR using primers 6054 and 6055 and cloning *Nco*I/*Xba*I in pPM1 leading to plasmid pFR9.

The C17Y substitution SiaC was introduced by site directed mutagenesis amplifying by inverse PCR using primers 5045 and 5046 using as pMM52 as template leading to plasmid pMM121.1.

C-terminal His-Strep double tagged *gpdC* was amplified by two-step overlapping PCR using primers 5081, 5467 and 5530 and cloned in pMM47.A using *Sal*I and *Spe*I restriction sites leading to plasmid pPM3.

Fetuin deglycosylation analyses and lectin stainings

Bacteria were collected from blood agar plates and resuspended in PBS at OD_{600}=1. 100 µl of bacterial suspensions were then incubated with 100 µl of a fetal calf serum fetuin (Sigma F2379) solution (0.1 $g.l^{-1}$) for 120 minutes at 37℃. As negative control, 200 µl of 1:2 diluted fetuin solution alone was incubated for 120 minutes at 37℃. Samples were then centrifuged for 5 min at 13000 x g, supernatant collected and loaded in a 12% SDS gel. Samples were analyzed by immunoblotting (Fetuin, Rabbit anti-Bovine RIA, UCBA699/R1H, ACCURATE CHEMICAL & SCIENTIFIC CORPORATION) and lectin stainings were performed with Sambucus nigra lectin (SNA) according to manufacturer recommendations (DIG Glycan Differentiation Kit, 11210238001, Roche).

Human IgG deglycosylation analyses and lectin stainings

Bacteria were collected from blood agar plates and resuspended in PBS at OD_{600} = 1. 100 µl of bacterial suspensions were then incubated with 100 µl of a purified human IgG (Invitrogen, 02-7102) solution (0.5 $g.l^{-1}$) for 180 minutes at 37℃. As negative control, 200 µl of 1:2 diluted IgG solution alone was incubated for 120 minutes at 37℃. Samples were then centrufiged for 5 min at 13000 x g, supernatant collected and 3 µl (and 12 µl SDS buffer) were loaded in a 12% SDS gel. Samples were analyzed by Coomassie blue staining, immunoblotting (Goat Anti-Human IgG (Fc specific)-FITC antibody, F9512 Sigma)) and lectin stainings were performed with SNA according to manufacturer recommendations (DIG Glycan Differentiation Kit, 11210238001, Roche).

Mass spectrometric analysis of fetuin

Fetuin (Sigma F2379) was reduced with 10 mM TCEP at 37°C for 1 hour and alkylated with 50 mM iodoacetamide for 15 min at room temperature. Fetuin was digested with trypsin at an enzyme to protein ratio of 1:50 (w/w) at 37°C overnight. The peptides were desalted on C18 StageTips (Thermo Fisher Scientific, Reinach, Switzerland) according to the manufacurer's recommendations. The fetuin peptides were analysed on an LTQ Orbitrap instrument (Thermo Fisher, San José, CA, USA) coupled to an Agilent 1200 nano pump according to (*c.f.* chapter 4.1).

Outer Membrane Protein purification

Bacteria were collected from blood agar plates and resuspended in 3 ml ice cold HEPES 10mM (pH7.4) at OD_{600} = 1. Bacterial suspensions were then sonicated on ice until they turned clear and spined at 15600 x g for 2 minutes at 4°C. Supernatants were transferred and centrifuged again for 30 minutes at 15600 x g at 4°C. Pellets were resuspended in 2 ml HEPES 10mM with 1% sarcosyl (N-Lauroylsarcosine sodium salt, Sigma) and incubated at room temperature for 30 minutes. Finally, samples were centrifuged at 15600g for 30 min at 4°C and pellet resuspended in 0.1 ml HEPES. Samples were checked for quality and quantity on silver stained SDS-PAGE and analysed by MS/MS.

Gpd proteins and sialidase co-purification

Cc5 Δ*gpdC* bacteria harbouring plasmid pPM3, expressing a C-terminal His-Strep double tagged GpdC, or harbouring plasmid pPM2, expressing GpdC without any tag (Mock), were grown for 2 days at 37 °C in the presence of 5% CO_2 on sheep blood agar plates. Bacteria from 6 plates were scraped and lysed in 35ml of 25mM Tris-HCl, 150mM NaCl, 0.2% triton, 1% NP-40%, 1% sodium deoxycholate, pH7.6.

For His affinity purification, the lysates were clarified by centrifugation (10 min at 18500g at RT) and the supernatant was diluted 1:2 in PBS, 10 mM Imidazole, in the presence of proteinase inhibitor (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets, Roche). 3.5 ml of 50% slurry Chelating sepharose Fast Flow beads (GE Healthcare) was first coupled to Ni^{2+} according to the manufacturer instructions and then 1.75 ml of resin was added to the solution and incubated overnight at 4 °C on a rotating wheel. The solution was then loaded into a column and the resin washed first with 25 column volumes (CV) of high salt buffer (50mM Tris, 500mM NaCl, pH8) and then with 5 CV of low salt buffer (50 mM Tris, 100 mM NaCl, pH 8). Proteins were then eluted from the resin with 2 CV of elution buffer (50mM Tris, 100mM NaCl, 350 mM Imidazole, pH8). The material eluted from the Ni^{2+} column was then diluted 1:2 in PBS and 1 ml of 50% slurry (0.5 ml CV) *Strep*-Tactin® Superflow® resin (IBA, cat No: 2-1206-002) was added. The solution was then incubated overnight at 4 °C on a rotating wheel. The solution was then loaded into a column and the flow through reloaded into the resin 2 more times. The resin was then washed 4 times with 10 CV of Buffer W (100mM Tris, 150 mM NaCl, 1mM EDTA, pH8) and proteins eluted in 3 steps with 0.5 ml elution buffer (100mM Tris, 150 mM NaCl, 1mM EDTA, 2.5 mM desthiobiotin, pH8). The proteins present in the elution fractions were

identified by MS and immunoblotting, using anti-His for GpdC detection, anti-GpdG and anti-SiaC .
GpdC-sialidase co-purification was performed exactly as described above using *Cc5* Δ*PUL5* bacteria harbouring pPM3 plasmid or harbouring plasmid pPM2 (Mock).Proteins present in the elution fractions were identified by immunoblotting with anti-Strep antibodies to detect GpdC and anti-SiaC.

In vivo radiolabeling with [^{3}H] palmitate, immuno-precipitation and fluorography.

Bacteria were inoculated to HeLa epithelial cells (ATCC CCL-2) in complete DMEM at 37℃ with 5% CO_2 at a moi of 20. 15-16 h post infection, [9,10-^{3}H] palmitic acid (48 Ci/mmol; Perkin-Elmer Life Sciences) was added to a final concentration of 50 µCi/ml and incubation was continued for 8-9 h, by which time the bacterial culture had reached approximately 10^8 bacteria/ml as described elsewhere [58]. Supernatants of 2 x 1 ml were collected without detaching epithelial cells from the wells. Bacteria corresponding to approximately 2x 10^8 cfu were then collected by centrifugation and pellets were combined from 2 ml and stored at -20℃ until they were processed. Pellets were resuspended in 0.1 ml PBS TritonX 1% to lyze bacteria and sialidase was immuno-precipitated by addition of 10 µl rabbit polyclonal anti-SiaC for 1h at RT on a rotating wheel. Protein A agarose slurry (Sigma) was then added in equal amounts for 30 min under constant rotation at RT. Samples were then centrifuged at 14000 x g for 2 min at RT, supernatant was discarded and pellets were washed with 0.5 ml PBS 0.1% Triton which was repeated 4 times. Captured proteins were eluted by addition of 50 µl Lämmli buffer (1% SDS, 10% glycerol, 50 mM dithiothreitol, 0.02% bromophenol blue, 45 mM Tris, pH 6.8) for 5 min at 85℃. Samples were centrifuged again and supernatant was carefully separated from the agarose beads and loaded on SDS PAGE gels using 10% polyacrylamide. After gel electrophoresis, gels were fixed in 25:65:10 isopropanol:water:acetic acid overnight and subsequently soaked for 30 min in AmplifyTM (Amersham). Gels were vacuum dried and exposed to SuperRX^{TM} autoradiography film (Fuji) for 13days until desired signal strength was reached.

Acknowledgements

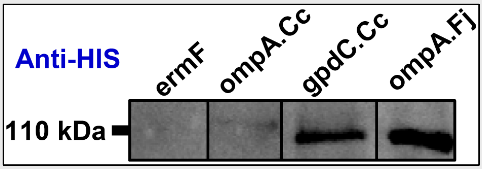
We thank Loïc Sauteur for occasional technical assistance, Simon Ittig and Alexander Titz for stimulating discussions.

5.2. Additional data: New promoters with diverse expression levels

Since Mally *et al.* established expression tools for *C. canimorsus* [82], *ermF* promoter has been intensively and exclusively used in our system. However purification trials of GpdC were both performed under *ermf* and *gpdC*'s native promoter. Interestingly, *gpdC* promoter showed significantly stronger protein expression than *ermF* promoted constructs under our growth conditions (**Fig. 5.2**). In addition, the previously reported strong *ompA* promoter from *Flavobacterium johnsoniae* [118, 119] has been tested together with its *C. canimorsus 5* homolog for GpdC expression (**Fig. 5.2**). All constructs shown here were able to complement growth phenotype of the *gpdC* deletants strain when cultured in presence of cells and even display slightly faster growth on blood agar plates when GpdC was expressed under its native promoter or under the *F. johnsoniae's ompA* promoter (data not shown).

Figure 5.2 Representative anti HIS tag western blot of total Cc5 *ΔgpdC* bacteria expressing HIS-tagged GpdC under different promoters

ompA.Cc: *ompA* promoter (*Cc5)*
ermF: *ermF* promoter
gpdC.Cc: *gpdC* promoter (*Cc5*)
ompA.Fj: *ompA* promoter (*F. johnsoniae)*



6. Additional unpublished data

6.1. Additional genomes sequencing

Comparative genomic analysis of the *Capnocytophaga* genus has become possible since several sequencing projects achieved complete or advanced draft genome assemblies four human hosted strains: *Capnocytophaga ochracea* strains F0287 (61 contigs) and DSM 7271 (complete), *Capnocytophaga gingivalis* ATCC 33624 (37 contigs), *Capnocytophaga sputigena* ATCC 33612 (65 contigs). In order to characterize the molecular bases of *Cc5* host interactions, additional *C. canimorsus* strains have been targeted for sequencing. Three additional strains isolated from patients' blood (*i.e.* capable of pathogenesis) were selected. Study of genes conservation among the whole *Capnocytophaga* genus and among the *C. canimorsus* genomes could help us to identify genes important for the incidental pathogenesis of C. canimorsus. Strains selected for genome sequencing were *C. canimorsus 2* (*Cc2*), *C. canimorsus 11* (*Cc11*) and *C. canimorsus 12* (*Cc12*). In addition, a *Cc5* transposon mutant derivative - X2E4 - that could not to be mapped with standard arbitrarily primed PCR, has also been sequenced for both insert localization and improvement of the *Cc5* wild type genome read depth.

6.2. Genomes sequencing and Assembly

The so called second generation deep sequencing methods (*e.g.* Solexa/Illumina, 454, ABI SOLiD) generate very high read coverages at the expense of read size (for example 36 bp for Solexa). At the time this work has been performed, Solexa represented the most efficient alternative in terms of sequence coverage and allowed pair ends recovering. This latter feature consisting in generating length homogeneous fragments and keeping track of the relationship shared by two reads coming from both ends of a same fragment.

The three sets of microreads generated have been tested on a series of recent assembler software devoted to microreads or hybrid assemblies (*i.e.* using different sequencing chemistries) (**Table 6.2.1**). However, best assemblies did not go below 1000 contigs when assemblers were used independently in the case of *Cc2*, *Cc11*, *Cc12* and even for the *Cc5* isogenic strain *X2E4* (**Table 6.2.2 and Table 6.2.3**).

Table 6.2.1 Assembler programs tested in the present study

	Assembler type	reference sequence	hybrid	PE[1]	Time & CPU	Other
EDENA	*de novo*	No	No	Yes	+ + +	
VELVET	*de novo*	Yes as tutors	No	Yes	+ + +	Merging limitations
MAQ	Resequencing	Required	Yes	Yes	+ +	Align only
SSAKE	*de novo*	Yes as Elongation Seeds	No	Yes	- - -	Calling quality ignored

[1] Possibility to use Pair End reads data sets

Table 6.2.2 Performances of three *de novo* assemblers on the X2E4 reads set

PROG	Contigs	Reads used (Mb)
VELVET	**4488**	NA
EDENA	**1606**	2.73 (71%)
SSAKE	**NA (>>)**	NA

Table 6.2.3 Performances of the mapping software MAQ on different reads sets

MAQ Run	Contigs	NON COVERED[1]	Reads used (Mb)
X2E4	57	264	3.07 (79%)
Cc2	**1.0461**	339.948	2.96 (61%)
Cc11	**14.800**	621.534	1.40 (52%)
Cc12	**14.238**	677.896	2.30 (50%)

[1] Number of bases from Cc5 that failed to map reads.

6.2.1. Development of a microreads assembly pipe using different assemblers' features.

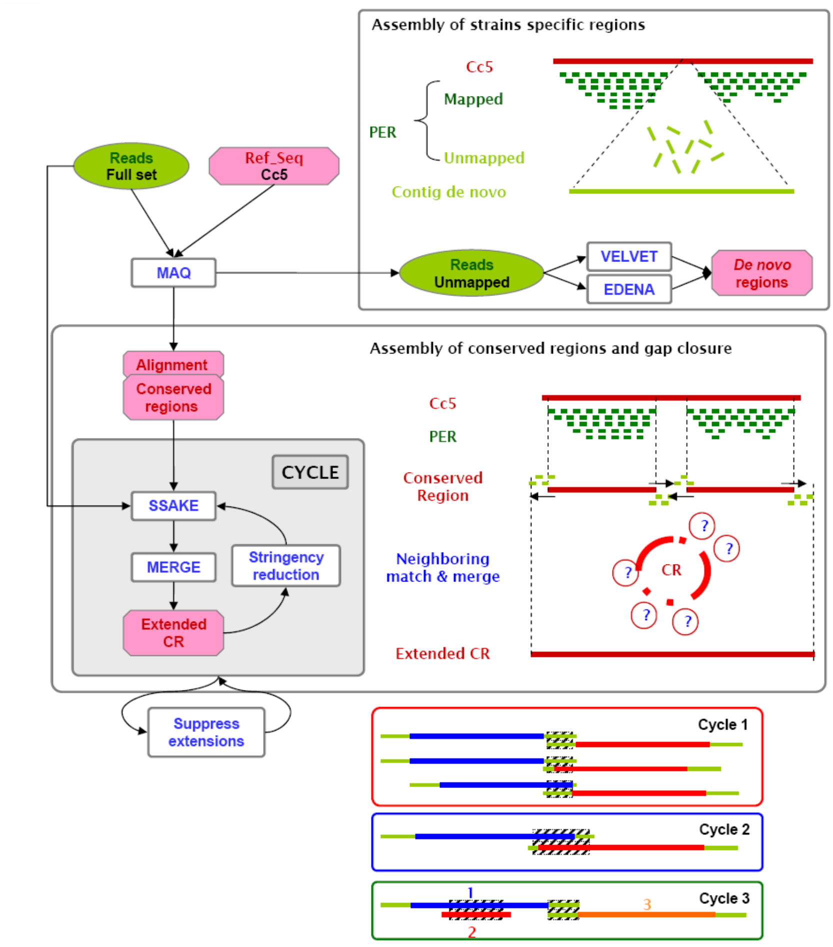
With the aim to optimize assembly process, a hybrid assembly pipe exploiting best features of the currently available software has been developed (**Figure 6.2.1 & Supplementary data, Chapter_6.2_assembling_methods**). In a first step, the complete genome of *Cc5* has been used as scaffold and microreads from the three newly sequenced strains showing 100% identity values with the *Cc5* sequence have been mapped on it by using MAQ [86]. Well covered chromosomal regions were referred as conserved regions (CR) and unmapped reads (UMR) were outputted (*i.e.* recovered) and stored. UMRs were then independently assembled using the *de novo* assemblers Velvet [120] and Edena [121]. Separation of the assembly process between CRs and strain specific regions (SSRs) has been thought to prevent misassembling interferences from the mapped reads during *de novo* assembly of SSRs. In parallel, the complete sets of reads were also employed to extend CRs with the SSAKE [122] software. SSAKE takes CRs, referred as seeds, and the whole set of microreads as inputs. It then only considers seeds extremities for an extension process using overlapping microreads.

Once this process has run over all seeds (*i.e.* CRs), neighboring regions (according to the Cc5 chromosome topology) are pairwise aligned at their contiguous boundaries by Xmatch (http://www.phrap.org/). In case the overlap satisfies the arbitrary assembling constraints (*e.g.* match length, identity values, coverage…), sequences are merged with Merge [89]. This way merging control parameters can be further relaxed while maintaining a high assembly accuracy level. Indeed, we expect that the chances that two contiguous CRs in *Cc5* will be contiguous in another strain are substantially higher than the chances to independently build two overleaping misassembled sequences. Inversely, note that in case of a classical assembly process, the “all against all” alignment step highly increases the chance to find false positive overlaps. In addition, to avoid uncontrolled CRs boundaries extensions that could lead to misassembled edges and prevent two contiguous CRs from merging, SSAKE is integrated in a stepwise stringency reduction loop. Each round SSAKE is seeded with extended and/or merged CRs from the previous round and assembling constraints are incrementally relaxed. This allows potential CRs' merge to occur before abusive extension may occur.

Complete cycling through the extension-merging process is achieved three times with increasing leniency of the merging restriction rules. During the first cycle, merging of two contiguous CRs is allowed if matching parts only span over CRs' extensions. This has been meant to prevent premature gap closure between two repeated regions close in the chromosome of Cc5 but potentially separated by a SSR in another strain. The second cycle allows contiguous regions to be merged over their CRs. This step considers InDel (Insertion / Deletion) events in the evolutionary course separating each strain to Cc5. The last cycles allows merging in case of complete embedment of one region by another. This latter rule allows clearance of false positive CRs that would prevent actual neighboring contigs to merge (most likely in case of duplication events specific to *Cc5*). In addition, because of the decreased assembly stringency at latest steps of each cycle, contigs (including orphan CRs) are cleared for non-joining extensions before considered for the next cycle.

Ultimately, contigs formed of jointed CRs and those resulting from the *de novo* assembled UMRs were assimilated to pseudoreads and inputted into the Phrap assembler (http://www.phrap.org/) for final assembly and visualization. Additionally, primer walking for final gap closure has been performed on *Cc2.*

Figure 6.2.1 Solexa assembly pipeline



The microreads assembly strategy presented here separates the assembling process over strain-scaffold conserved regions from the one over strain specific regions. Gap closure between contiguous conserved regions on the scaffold is performed through edges extension and pairwise merge assessment. Assembly stringency is quantitatively and qualitatively reduced stepwise to maximize assembly and minimize effects of possible misassembly. Microreads datasets are represented as green ovals, processed sequence data as pink hexagons, and programs and scripts as white boxes. Dark and light green short bars respectively represent mapped and unmapped microreads. Red, blue and orange long bars are conserved regions while green long bars are de novo assembled contigs. Hatched boxes represent matching regions. Question marks indicate decisional point for contig joining.

The method described here substantially increased the assembling performances compare to the different assemblers when used independently (**tables 6.2.2, 6.2.3 and 6.2.4**). However, despite an orientated assembly strategy, the process accuracy may have been decreased at several points of the pipeline, in particular during last steps of CRs extension cycles. Integration of a quality score tracking back local assembly accuracy (as the one used by MAQ) would be a necessary step further to achieve better data processing. A series of feed back tests could be done by mapping whole datasets against the new strains assemblies and compare it to previous mappings against the *Cc5* genome. Total amount of mapped reads would be informative of the level of assembling achieved while coverage deviation would indicate presence or absence of sequence redundancy.

Table 6.2.4 Performances of three microreads assembly

Strain	Contigs prior Phrap	Final Contigs[1]	Cumulative Size (Mb)	largest contig	N95[2]	N50[2]
Cc5	-	-	2.571	-	-	-
Cc2	185	22(3)	2.524	1368379	101525 (3)	1368379 (1)
Cc11	516	152	2.508	91762	3413 (109)	36452 (22)
Cc12	266	63	2.531	341916	11115 (39)	94748 (9)

[1] numbers in between brackets correspond the contig number after primer walking

[2] numbers in between brackets correspond to the number of contigs at least as long as the

6.2.2. Preliminary hybrid assembly of the *Cc2*, *Cc11* and *Cc12* genomes using Solexa, 454 and Sanger sequencing chemistries.

In order to enclose complete genome sequencing of *Cc2*, *Cc11* and *Cc12* an additional run of 454 pyrosequencing has been performed at Microsynth, Balgach (CH). Approximately 10X read coverage per strain were generated (**table 6.2.5**).

Table 6.2.5 Lifescience 454 sequencing data

Strain	reads	Avrg. length	Total
Cc2	79 417	309	24.5 Mb
Cc11	80 168	333	26.7 Mb
Cc12	74 882	326	24.5 Mb

Very recently technical progress allowed best usage of current assembly methods: First, a dramatic improvement of the available hardware at the BC2 Basel university framework particularly concerning nods memory. Indeed, second generation sequencing methods generate very large data sets requiring high memory nods. And second, a clear improvement of assembler software that can now perform complex tasks and integrate several sequencing technologies (*e.g.* MIRA.3, http://www.chevreux.org/projects_mira.html). Here, *Cc5* complete genome has been used with MIRA.3 as a scaffold for short (454) and micro (Solexa) reads mapping. Well covered regions (taking into account read coverage and base calling qualities) were then turned to constant low quality Sanger pseudoreads with mktrace (Phred / Phrap / Consed package) and the *BC2_MIRA_output_TCS_file_Parser.pl* (**Supplementary data, Chapter_6.2_assembling_methods**) in-house script. Single Nucleotide Polymorphisms (SNPs) and small Insertions/deletions events (Indels) were tolerated during the mapping phase and therefore appeared within the corresponding pseudoreads (virtually reconstituted chromatograms). Such pseudoreads represent conserved regions between *Cc5* and the assembled strain and were then of great value to orientate reassembling of the whole data set. The generated pseudoreads were then added to 454 and Solexa reads as Sanger reads in a hybrid *de novo* assembly with Mira.3. Since redundancy in assembly was still high (**Table 6.2.6**), contigs were further assembled with Phrap. Assembly statistics Positions exhibiting degenerated base calling were turned to deoxycytidines (C) in order to minimize false negatives during Open Reading Frame determination as stop codons lack deoxycytidines. CDS prediction on newly assembled draft genomes has then been performed as previously described in chapter 4.1. After CDSs translation the three newly predicted proteomes were integrated to further ortholog analysis.

Table 6.2.6 Lifescience 454 sequencing data

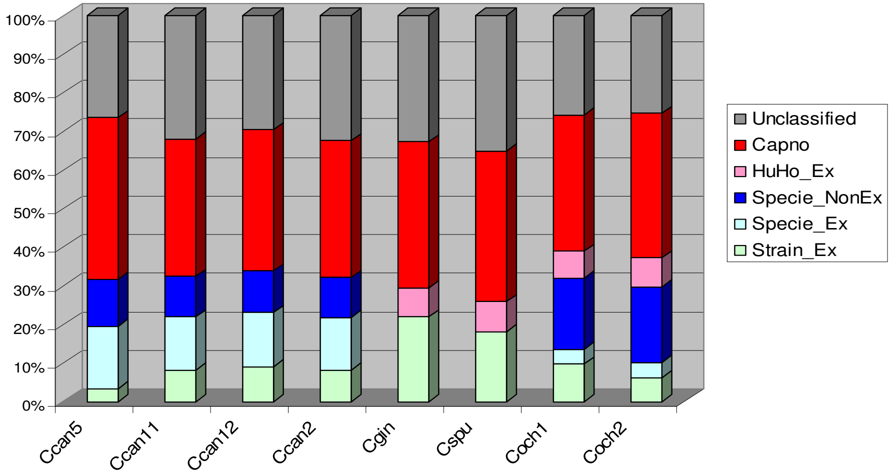
MIRA >500 bp	contigs	Cumulative Size (bp)	Largest contig (bp)	N95[1]	N50[1]
Cc2	262	2573684	75772	3379	21021
Cc11	359	2538073	96082	2062	15391
Cc12	176	2437242	136522	4984	38583
MIRA					
Cc2	3655	3129525	75772	228	16147
Cc11	3080	3082587	96082	276	11088
Cc12	3145	2988930	136522	264	27259
MIRA + Phrap					
Cc2	289	2510543	75826	4870 (107)	28366 (29)
Cc11	267	2446272	117363	3268 (159)	18129 (37)
Cc12	81	2383627	160249	10215 (51)	64490 (12)

[1] numbers in between brackets correspond to the number of contigs at least as long as the corresponding length of the N95 or the N50 contig.

6.3. Genomics of *Capnocytophaga*

With the four *C. canimorsus* genomes presented here, it is now possible to determine gene conservation among different strains or species and isolate set of genes potentially involved in human or dog commensalism but also in pathogenesis of *C. canimorsus*. Clustering of orthologs defined several group of interest: 1) Genes conserved among all *Capnocytophaga* genomes defined the genus core genome and represented 39% of the genome size in average (1009 genes). 2) Genes conserved among *canimorsus* isolates but not conserved or absent from the three HCSs were respectively named inclusive (678 genes) and exclusive (421 genes) *canimorsus* corer genomes. 3) Inversely, genes conserved among *C. gingivalis*, *C. ochracea* and *C. sputigena* but not conserved or absent from the four *C. canimorsus* strains were respectively named inclusive (not counted) and exclusive (202 genes) human-hosted *Capnocytophaga* core genomes.

Figure 4.2.2 Relative taxonomic distribution of orthologs among *Capnocytophaga*



Orthologous groups are classified according to the taxonomy of the concerned genes. Color code is red for genus core, pink for human hosted *Capnocytophaga* species exclusive core genome, blue and teal respectively for inclusive and exclusive species core genomes, green for the strain specific genes and grey for unclassified groups. *Ccan* stands for *C. canimorsus*, *Cgin* for *C. gingivalis ATCC33624*, *Cspu* for *C.sputigena ATCC33612* and *Coch1* for *C. ochracea F0287* and *Coch2* for *C. ochracea DSM7271*.

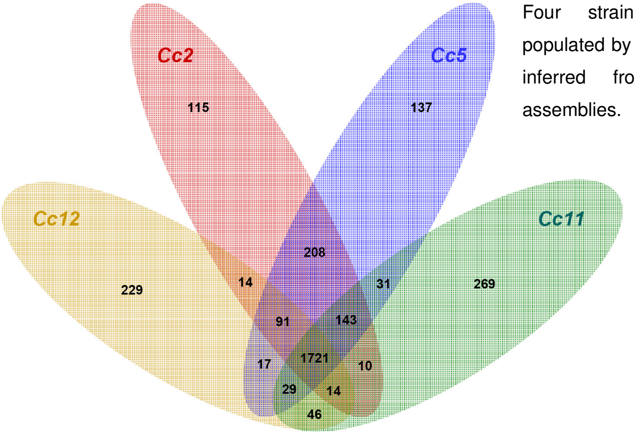
6.3.1. Mapping of the X2E4 transposon mutant

X2E4 is a *Cc5* derivative transposon mutant exhibiting a strong growth defect in presence of cells (around 100 fold less cfu than wild-type after 24h incubation) (M. Mally, doctoral thesis). After mapping reads from X2E4 onto the *Cc5* chromosome, transposon insertion has been successfully mapped in a gene encoding a putative cytosolic dihydroorotase (DHOase) conserved among *Bacteroidetes* (Ccan_03130). DHOase catalyze the reversible interconversion of carbamoyl aspartate to dihydroorotate, a key reaction in pyrimidine biosynthesis. X2E4 display a moderated growth defect in presence of cells (below 10 fold, data not shown) which would then be consistent with the conserved function of Ccan_03130. Besides possible metabolism redundancies, the presence of a second highly conserved DHOase encoded by Ccan_10340 might explain why insertion occurring in X2E4 is not lethal.

6.3.2. Genomics of *C. canimorsus*

Orthologous clustering has been performed among the four *C. canimorsus* strains. Gene populations are mapped on a Venn diagram according to their taxonomic profiles (**Figure 6.3.2**). The species core genome accounted for 1721 genes and represents up to 71% of the *Cc5* genome. Strain specific genes accounted in average for 7.2% of the genomes. It is noteworthy that the more strains from the same species are integrated to the analysis, the lower the gene content will be for the core or the strain specific groups.

Figure 6.3.2 Strain distribution of the *C. canimorsus* orthologous groups



Four strain Venn diagram populated by orthologous groups inferred from Solexa draft assemblies. Colored areas

6.3.3. What makes *C. canimorsus* a dog commensal and a potentially lethal opportunistic pathogen?

Two set of genes are of high interest for commensalism and pathogenesis understanding : i) Coding sequences conserved among the four clinically isolated *C. canimorsus*, with tow flavors, exclusively or not and ii) Coding sequences conserved among the four human associated *Capnocytophaga* species, with again tow flavors, exclusively or not.

Among the 421 genes exclusively conserved in the four *C. canimorsus* strains among *Capnocytophaga* (**Supplementary data, Chapter_6.3_Genomcis_of_Capnocytophaga, Table S6.3_Capnocytophaga_genomics**), most (216) were of unknown function but nine emerging functional categories accounted for 146 proteins : "Protein and amino acids metabolism" represented by 14 CDSs including four peripheral proteins (SPI, SPII, TM) involved in dipeptide binding, transport and degradation; "Phospholipids metabolism" (9); "Polysacharide utilisazion loci" (16 genes from PUL2, -4, -6, -7, -9 and -13); "Other Cazymes" (14 genes including 5 N-acetylosaminidases); "DNA binding and transcriptional regulation" (21 including a putative one-component Histidine kinase sensor protein); "Mobile Genetic Elements" (15); "General Metabolism" (5); "Transporters" (13); "Oxidative stress" (36). 59 CDSs were left unclassified but included several potential candidates for a role in pathogenesis or commensalism of *C. canimorsus* like the two partner secretion protein Ccan_13910 detected by MS at the *Cc5* bacterial surface, a putative vesicle-fusing ATPase (Ccan_05240) that might explain presence of integral outer membrane proteins in *Cc5* culture supernatants (data not shown), an operon including four putative cytolysine (two of them being detected at the OM), a methylglyoxal synthase potentially involved in protein glycation and possibly responsible for difficulties during heterologous expression of *Cc5* proteins in *E.coli,* a putative calcium binding protein (Ccan_07510) and its hypothetical outer membrane partner (Ccan_07520) both detected by MS in the OM and genetically located at the immediate vicinity of the conserved sec secretion regulator "Trigger factor" (Ccan_07530) and two eukaryotic-like proteins (the

Ccan_08450 intimin like protein and Ccan_20350, an ankyrin repeat-containing protein).

Interestingly, the predominant functional category was the "Oxidative stress" group. This set included genes directly or indirectly involved in O_2 utilization and oxidative stress resistance. It is known that protein participating to the respiratory electron transport contribute both to O_2 consumption and oxidative resistance. All genes encoding the Mrp complex but one, all genes from the Cytochrome C oxidase complex 1 (CcO 1) except one and the majority of genes involved in sodium cotransport were therefore assigned to this group. Genes encoding the CcO 2 were all found conserved among *C. canimorsus* strains but also partially present in *C. gingivalis* ATCC 33624. In addition to the conserved phosphoenolpyruvate carboxykinase (Ccan_15480) *C.canimorsus* acquired / maintained a phosphoenolpyruvate carboxylase (Ccan_10960) that is unable to produce ATP while it might increase fumarate production rates and consequently boost respiratory chain transfers (**Figure 6.3.3**).

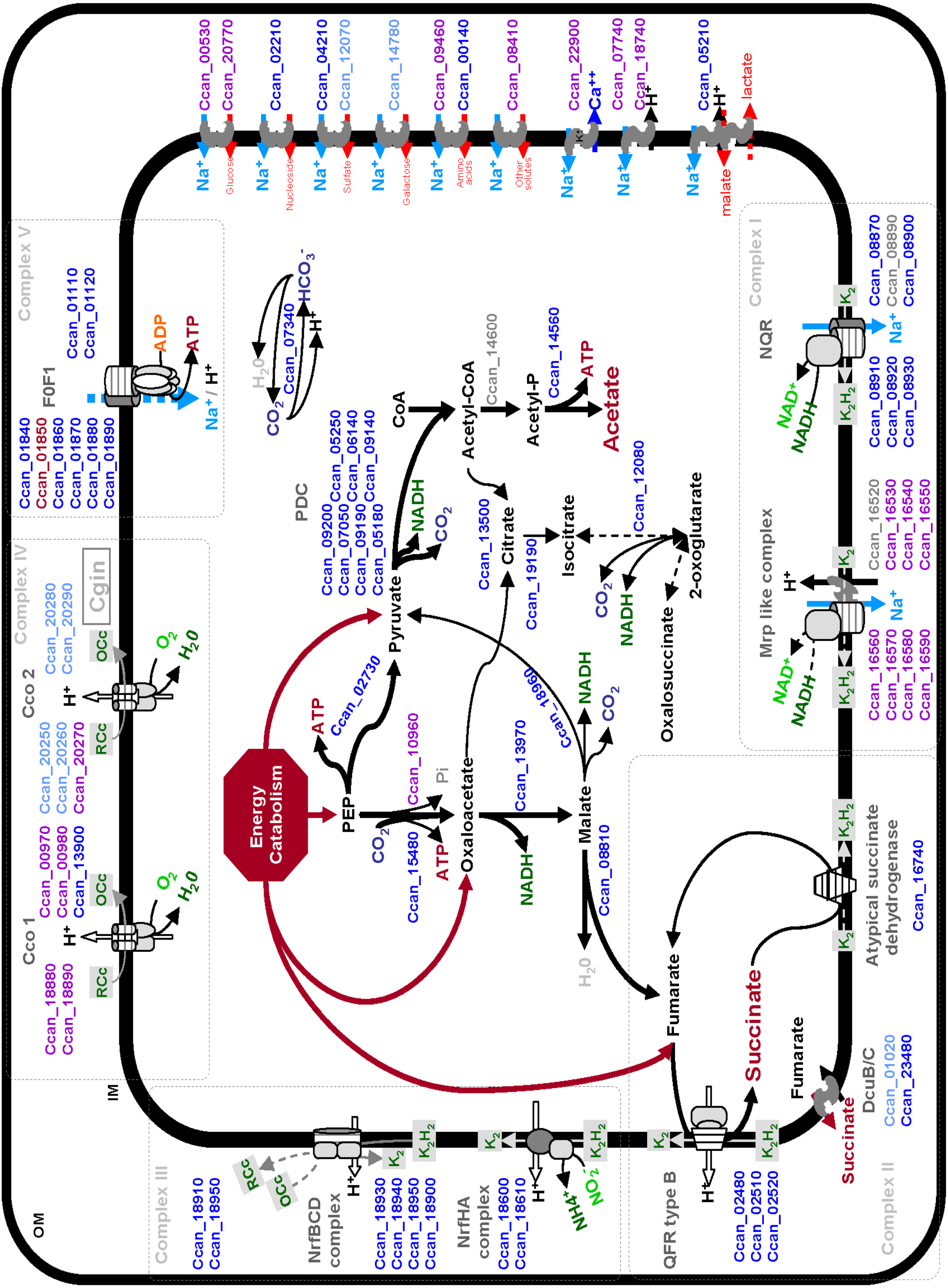

OM
IM
Complex V
F0F1
ADP
ATP
Na+ / H+
Ccan_01840
Ccan_01850
Ccan_01860
Ccan_01870
Ccan_01880
Ccan_01890
Ccan_01110
Ccan_01120
Complex IV
Cgin
Cco 1
Cco 2
RCc
OCc
H+
O2
H2O
Ccan_18880
Ccan_18890
Ccan_00970
Ccan_00980
Ccan_13900
Ccan_20250
Ccan_20260
Ccan_20270
Ccan_20280
Ccan_20290
Complex III
Ccan_18910
Ccan_18950
NrfBCD complex
Ccan_18930
Ccan_18940
Ccan_18950
Ccan_18900
NrfHA complex
Ccan_18600
Ccan_18610
NH4+
NO2-
K2
K2H2
QFR type B
Ccan_02480
Ccan_02510
Ccan_02520
Complex II
DcuB/C
Ccan_01020
Ccan_23480
Succinate
Fumarate
Atypical succinate dehydrogenase
Ccan_16740
Mrp like complex
Ccan_16520
Ccan_16530
Ccan_16540
Ccan_16550
Ccan_16560
Ccan_16570
Ccan_16580
Ccan_16590
NAD+
NADH
Na+
Complex I
NQR
Ccan_08870
Ccan_08890
Ccan_08900
Ccan_08910
Ccan_08920
Ccan_08930
Energy Catabolism
PEP
Pyruvate
Oxaloacetate
Malate
Fumarate
Citrate
Isocitrate
2-oxoglutarate
Oxalosuccinate
CoA
Acetyl-CoA
Acetyl-P
Acetate
PDC
Ccan_09200
Ccan_07050
Ccan_09190
Ccan_05180
Ccan_05250
Ccan_06140
Ccan_09140
Ccan_02730
Ccan_10960
Ccan_15480
Ccan_13970
Ccan_08810
Ccan_18960
Ccan_13500
Ccan_19190
Ccan_12080
Ccan_14600
Ccan_14560
Ccan_07340
CO2
HCO3-
H2O
Pi
ATP
NADH
Ccan_00530
Ccan_20770
Ccan_02210
Ccan_04210
Ccan_12070
Ccan_14780
Ccan_09460
Ccan_00140
Ccan_08410
Ccan_22900
Ccan_07740
Ccan_18740
Ccan_05210
Glucose
Nucleoside
Sulfate
Galactose
Amino acids
Other solutes
K+
Ca++
malate
lactate

Figure 6.3.3

Conservation among *Capnocytophaga* species of the *C. canimorsus* fermentative and respiratory pathways.

The figure has been adapted from chapter 4.1. Gene references have been color coded according to their taxonomic distribution among *Capnocytophaga*: Blue stands for membership of the genus core, purple for the exclusive *canimorsus* core (*i.e.* genes that are exclusively found in all *C. canimorsus* strains), light blue for the inclusive *canimorsus* core (*i.e.* genes that are found in all *C. canimorsus* strains and non exhaustively present in other *Capnocytophaga* species), red for Cc5 strain specific genes and grey for unclassified taxonomic membership. The grey box labeled *Cgin* indicates the four genes of the inclusive *canimorsus* core encoding the Cco2 complex are also found in *C. gingivalis.*

6.3.4. *C. canimorsus* and O_2 utilization

Human hosted *Capnocytophaga* species (HCSs) have been reported to grow in air only if CO_2 supplementation was provided. In addition they are devoid of catalase and oxidase activities and O_2 consumption, as analyzed by oxygen electrodes, has never been detected according to Leadbetter *et al.* [22]. In contrast, Brenner *et al.* reported characteristic catalase and oxidase activities for both *C. canimorsus* and *C. cynodegmi* [7]. In addition to that, we observed slightly delayed but consistent growth of *C. canimorsus* when kept in air at 37℃ without CO_2 supplementation in cell cultures but also on blood agar plates (**table 6.3.4**). Together with the presence of a several *C. canimorsus* specific genes increasing both generation (Mrp, Cco1, Cco2, phosphoenolpyruvate carboxylase) and utilizations (Na^+-cotransporters) of Na^+/H^+ ionic gradients, all these data suggest the occurrence of a metabolic switch from a typical *Capnocytophaga* fermentative metabolism to a more respiratory one.

C. cynodegmi also exhibited slightly delayed growth on blood agar plates without any addition of carbon dioxide (**table 6.3.4**). Thus, it is likely that these features are not responsible of the pathogenic tendencies of *C. canimorsus* in the human host. However, they could be a perquisite to resist oxidative stress in human blood and certainly have a role in maintenance of the bacterium in the canine oral cavity.

Table 6.3.4 Growth of *Capnocytophaga* species under different O_2 and CO_2 concentrations

Blood Agar	Anaerobiose	Candle jar	Air + 5% CO_2	Aerobiose
Cc5	0	+	++	+
HCSs	+?	+?	+?	0
C.cynodegmi	ND	ND	++	+

0 indicates no growth was observed after 4 days incubation; +, growth after 3 days incubation; ++, growth after 2 days incubation; ? means reported from E.R Leadbetter *et al.*, 1979; ND stands for not done.

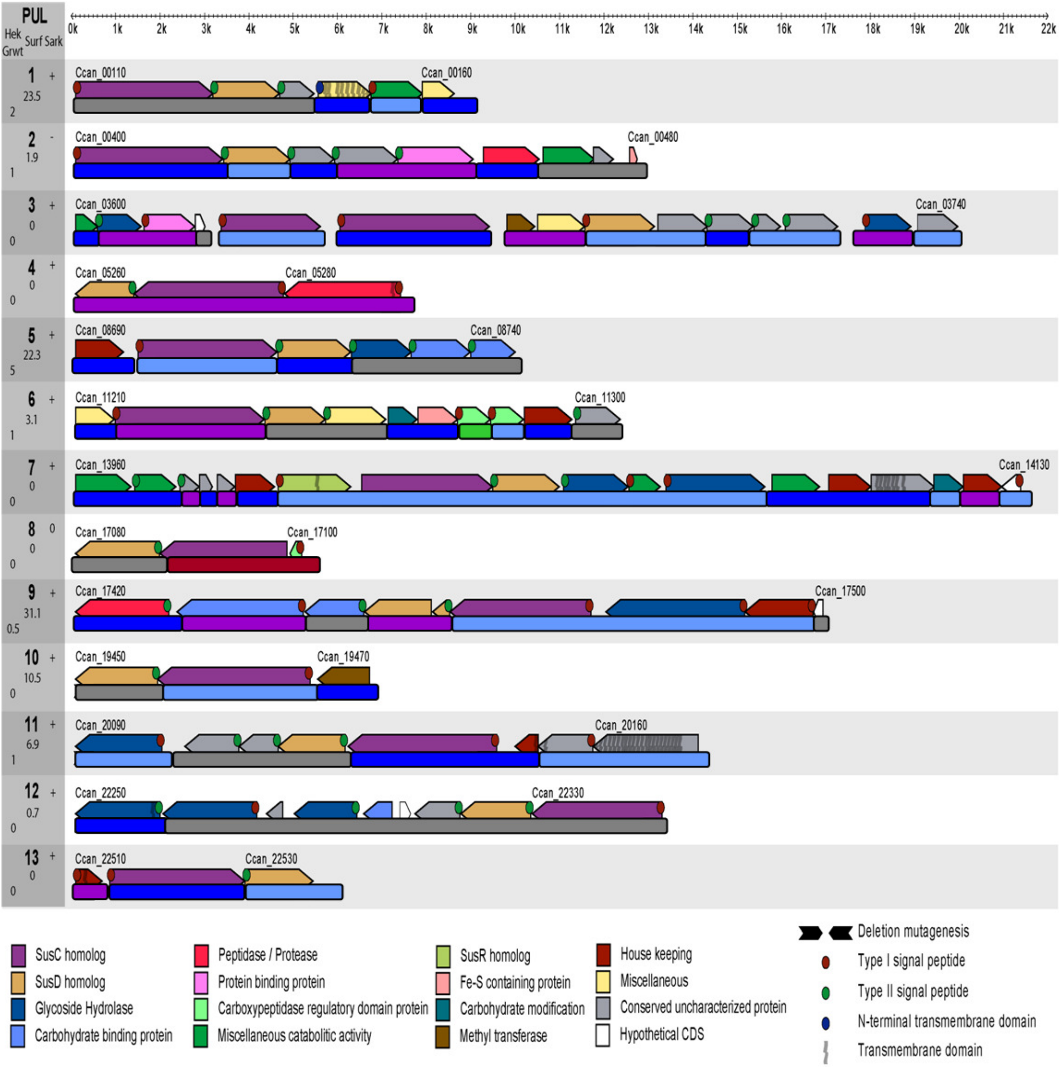
6.3.5. Taxonomic conservation of polysaccharide utilization loci

Expectedly, PULs conservation among *Capnocytophaga* is globally high and among the 102 genes assigned to PULs in the *Cc5* genome, 27 genes were found conserved among all *Capnocytophaga* and 72 genes were conserved among *C. canimorsus* strains (**Figure 6.3.5**). Only 18 genes were exclusively conserved among *C. canimorsus* species and two were only found in *Cc5*. The two latter (Ccan_1690 & Ccan_17100, PUL8) share similarities with two consecutive domains of a single SusC homologue suggesting a pseudogenic event. In addition, no protein encoded by PUL8 has been detected by MS.

Synteny conservation was observed for most genes among the *C. canimorsus* genomes (data not shown). However, as it has been reported for other *Bacteroidetes* [78], PULs recombination was frequent when comparison was extended at the genus level (indirectly in **Figure 6.3.5**).

Surprisingly, in the case of PUL5, the most conserved protein was the SusD homolog GpdD involved in glycan binding. GpdD was present in all *Capnocytophaga* genomes and human hosted *Capnocytophaga* species even exhibited multiple paralogs. Concerning GpdC, all human hosted strains with the exception of *C. gingivalis*, that lacks a GpdC ortholog, presented multiple GpdC paralogs. In the case of *C. canimorsus,* a single couple of GpdCD genes was identified in each genome. The apparent importance of these two genes contrasts with the scarce distribution of the GpdG β-endo-glycosidase that only showed-up in *Cc5*, *Cc2*, *C. sputigena* and C. *ochracea F0287.* Presence of the two putative lectins orthologs GpdE and GpdF strictly correlated to the GpdG occurrence suggesting a collaborative functionality. It is thus tempting to speculate a key role for GpdC and GpdD in polysaccharide binding and selection while the cleavage mechanism involving GpdGEF is accessory and can be replaced by diverse other enzymatic processes (*e.g.* different cleveage sites). Whether these paralogous PULs target the same substrates or if they are still involved in carbon source scavenging remains difficult to predict from current data.

Figure 6.3.5 PULs conservation among *Capnocytophaga* species



The figure has been adapted from chapter 4.1. Gene have been underlined with a color code according to their taxonomic distribution among *Capnocytophaga*: As previously, blue stands for membership of the genus core, purple for the exclusive *canimorsus* core, light blue for the inclusive *canimorsus* core, red for Cc5 strain specific genes and grey for unclassified taxonomic membership. Numbers and symbols on the left correspond to the growth impairment factor in presence of Hek cells compare to wild type (left bottom corner), the percentage of surface abundance among PULS (middle) and the presence (+) or the absence (-) from MS analysis of the outer membrane fraction.

Material and Methods

Solexa run

Genomic DNA samples have been obtained as previously described (*c.f.* chapter 4.1) and sent for sequencing at FASTERIS SA, Geneva. Sequencing method consisted in a single run of Solexa/Illumina GAII *EAS269* on 100 tiles during 36 cycles. Picture acquisition and analysis have been processed through the *GAPipeline1.0rc4* pipeline. 5' sequence extremities were screened for the presence of artificial bar codes segregating for biologic samples (**Table 6.4**). Same bar codes should be present on both reads coming from the same sequenced polony (*i.e.* the PCR amplification product of the targeted insert fixed on a solid phase). Biological sample consistency validation has been carried out between the set of 1st read and 2nd read for each polony (same bar code for both reads). Whenever one read did not display the tag, or two reads displayed different tags, both reads were discarded. After Quality streaming reads were 33 nucleotides length.

Table 6.4 Statistics of bar coded samples used together during Solexa run

Sample	Barr code	PE Reads	% Total reads
CC2	GT	2'394'734	27.4%
CC11	CT	1'344'302	15.4%
CC12	AT	2'294'977	26.2%
X2E4	TT	1'927'946	22.0%
Total	-	7'961'959	91.0%
Read length			33 bps
Average fragment length			250 +/-50 bps

MIRA command line for the Genome mapping

```
/import/bc2/home/guest/cornelis/manfpa00/FASTERIS/MIRA.3/mira_3.2.1rc2_prod_linux-
gnu_x86_64_static/bin/mira
--project=Cc2
--projectin=
/import/bc2/home/guest/cornelis/manfpa00/FASTERIS/MIRA.3/Cc2_mira/Cc2
--job=mapping,genome,accurate,454,solexa
        -SB:abnc=1
        -LR:ssiqf=yes
        SOLEXA_SETTINGS
                -LR:ft=fasta
                -CO:msr=no
                -GE:uti=no:tismin=200:tismax=400
```

MIRA command line for the de novo assembling

```
/import/bc2/home/guest/cornelis/manfpa00/FASTERIS/MIRA.3/mira_3.2.1rc2_prod_linux-
gnu_x86_64_static/bin/mira
--project=Cc2_Reass
--projectin=
/import/bc2/home/guest/cornelis/manfpa00/FASTERIS/MIRA.3/Cc2_mira/Cc2
--job=denovo,genome,accurate,454,solexa,sanger
        -LR:ssiqf=yes
        SOLEXA_SETTINGS
           -LR:ft=fasta
        SANGER_SETTINGS
           -LR:ft=phd
```

Growth of *Capnocytophaga* species under different O_2 and CO_2 concentrations

C. canimorsus 5 and a *C. cynodegmi* strain recently isolated in our lab from dog oral flora were grown on plates routinely (*c.f.* chapter 4.1) with the exception of two the varying conditions O_2 and CO_2 concentrations. Anaerobiosis has been reached by using a GasPak™ EZ Anaerobe Pouch System (Catalog #260683, BD) according to manufacturer recomendations. Microaerophilic conditions were achieved by using a candle extinction jar. Normal or CO_2 complemented aerobiosis were tested in a humidified 37 incubator with or without a 5% CO_2complementation.

7. Conclusions and perspectives

Conclusions and perspectives

The 2,571,405-bp genome sequence of *Cc5* shows close relationships with environmental *flavobacteria* as *Flavobacterium johnsoniae* and *Gramella forsetii*. Among *Capnocytophaga* species, it occupies a taxonomically median position since a phylogeny tree computed on conserved proteins positioned *C. canimorsus* in between three human associated *Capnocytophaga* species. It is thus tempting to think that host specialization occurred after adaptation to the oral environment.

C.canimorsus 5 has undergone large-scale horizontal gene transfers compensated by gene losses thus maintaining a reduced genome size. Consistently, metabolic modelling shows a reduced global pleiotropy and a high degree of specialization to the oral environment. Indeed, we postulate that *Cc5* couples a CO_2-dependent fumarate respiration to a Na^+ based respiratory chain adapted to oral fluids rich in HCO_3^- and Na^+ ions. Further understanding of the metabolic requirements of *C. canimorsus* would significantly reduce complexity of the currently used rich broth (serum or blood complemented). It would allow us to investigate cell cultures supernatant contents for protein or secondary metabolites potentially involved in Cc5's anti-inflammatory features.

The genome of *Cc5* did not encode any classical complex virulence functions as T3SSs or T4SSs. However, it exhibits a very high relative number of surface-exposed lipoproteins that account for 76% of the total surfome and many of which are encoded within 13 different PULs. At least 12 PULs were expressed under our growth conditions and corresponded to more than 54% of total MS-flying peptides detected at the surface. A systematic knockout analysis of the 13 PULs revealed that 6 PULs are involved in growth during cell culture infections with most dramatic effect observed for ΔPUL5.

The PUL5 encoded Gpd surface-complex turned out to be devoted to foraging glycans from N-linked glycoproteins as fetuin but also IgG. It also plays a role in survival in mice and in fresh human serum and therefore represents a new type of virulence factor. In order to further test this hypothesis fresh human blood infection assays [123] would enclose conditions encountered by *C. canimorsus* during systemic infections and eventually help to identify PULs and substrates involved in bacterial survival in the human host. In parallel, a PCR screen for the presence of all 13 PUL among members of our *C. canimorsus* library might reveal correlations between the occurrence of certain PUL genes and the pathogenicity of the strains isolated from patients.

GpdCDEF contribute to the binding of glycoproteins at the bacterial surface while GpdG cleaves N-linked oligosaccharide after the first GlcNAc residue and possible terminal sialic acid residues of the oligosaccharide are removed by SiaC in the periplasm. Finally, degradation of the imported oligosaccharide proceeds sequentially from the desialylated non reducing end by the action of periplasmic exoglycosidases. Identification of others PULs specific substrates has been recently addressed in G.R. Cornelis' lab (L. Sauteur, Master thesis). Despite significant gene conservation and observed expression of most PULs, only few hint of a possible role of PUL6 and PUL9 in mucin O-glycan chains degradation have been found so far. Identification of additional salivary O-glycosylated proteins is currently ongoing.

Two assembling approaches were developed in order to enhance assembly capacities of pre-existing tools. Draft assemblies of the three pathogenic human blood isolates *Cc2*, *Cc11* and *Cc12* together with four available human hosted *Capnocytophaga* species were included to a comparative genomics analysis. The set of genes exclusively present and conserved among *C. canimorsus* strains was enriched in genes involved in respiration, oxidative respiration and oxidative stress resistance. Specific PULs members were also found within the differential gene set.

It is likely that *C. canimorsus* has evolved its human aggressiveness through adaptation to the carnivores' oral environment. However, *C. canimorsus* is often co-isolated with *C. cynodegmi* from canine oral swaps. In fact *C. cynodegmi* has been reported with a higher prevalence in dog's mouth [13]. In contrast to *C. canimorsus* that is mostly associated with systemic infections, *C. cynodegmi* is only known to scarcely trigger local wound infection on individuals with no reported immunosuppression (mostly animals). Such differences in pathogenesis contrast with the nucleic acid similarity levels shared by *C. canimorsus* and *C. cynodegmi* (Closest known species)

[50]. Genome comparison of the more frequent but non systemic *C. cynodegmi* versus the one of the less prevalent but clinically relevant (isolated from human blood) *C. canimorsus* would be a step forward in the identification of genes potentially involved in oral canine adaptation and in those that may have a predominant role in pathogenesis. Particular care could also be given to the group of genes conserved with the human hosted species but absent from *C. cynodegmi*.

8. References

References

1. Lion, C., F. Escande, and J.C. Burdin, *Capnocytophaga canimorsus infections in human: review of the literature and cases report.* European journal of epidemiology, 1996. **12**(5): p. 521-33.
2. de Boer, M.G., et al., *Meningitis caused by Capnocytophaga canimorsus: when to expect the unexpected.* Clinical neurology and neurosurgery, 2007. **109**(5): p. 393-8.
3. Pers, C., B. Gahrn-Hansen, and W. Frederiksen, *Capnocytophaga canimorsus septicemia in Denmark, 1982-1995: review of 39 cases.* Clinical infectious diseases, 1996. **23**(1): p. 71-5.
4. Tierney, D.M., L.P. Strauss, and J.L. Sanchez, *Capnocytophaga canimorsus mycotic abdominal aortic aneurysm: why the mailman is afraid of dogs.* Journal of clinical microbiology, 2006. **44**(2): p. 649-51.
5. Le Moal, G., et al., *Meningitis due to Capnocytophaga canimorsus after receipt of a dog bite: case report and review of the literature.* Clinical infectious diseases, 2003. **36**(3): p. e42-6.
6. Bobo, R.A. and E.J. Newton, *A previously undescribed gram-negative bacillus causing septicemia and meningitis.* American journal of clinical pathology, 1976. **65**(4): p. 564-9.
7. Brenner, D.J., et al., *Capnocytophaga canimorsus sp. nov. (formerly CDC group DF-2), a cause of septicemia following dog bite, and C. cynodegmi sp. nov., a cause of localized wound infection following dog bite.* Journal of clinical microbiology, 1989. **27**(2): p. 231-5.
8. Gaastra, W. and L.J. Lipman, *Capnocytophaga canimorsus.* Veterinary microbiology, 2010. **140**(3-4): p. 339-46.
9. Macrea, M.M., M. McNamee, and T.J. Martin, *Acute onset of fever, chills, and lethargy in a 36-year-old woman.* Chest, 2008. **133**(6): p. 1505-7.
10. Dendle, C. and D. Looke, *Review article: Animal bites: an update for management with a focus on infections.* Emergency medicine Australasia, 2008. **20**(6): p. 458-67.
11. Verghese, A., et al., *Antimicrobial susceptibility of Capnocytophaga spp.* Antimicrobial agents and chemotherapy, 1993. **37**(5): p. 1206.
12. Janda, J.M., et al., *Diagnosing Capnocytophaga canimorsus infections.* Emerging infectious diseases, 2006. **12**(2): p. 340-2.
13. Suzuki, M., et al., *Prevalence of Capnocytophaga canimorsus and Capnocytophaga cynodegmi in dogs and cats determined by using a newly established species-specific PCR.* Veterinary microbiology, 2010. **144**(1-2): p. 172-6.
14. Frandsen, E.V., et al., *Diversity of Capnocytophaga species in children and description of Capnocytophaga leadbetteri sp. nov. and Capnocytophaga genospecies AHN8471.* International journal of systematic and evolutionary microbiology, 2008. **58**(Pt 2): p. 324-36.
15. Buu-Hoi, A.Y., S. Joundy, and J.F. Acar, *Endocarditis caused by Capnocytophaga ochracea.* Journal of clinical microbiology, 1988. **26**(5): p. 1061-2.

16. Parenti, D.M. and D.R. Snydman, *Capnocytophaga species: infections in nonimmunocompromised and immunocompromised hosts.* The Journal of infectious diseases, 1985. **151**(1): p. 140-7.
17. Rubsamen, P.E., et al., *Capnocytophaga endophthalmitis.* Ophthalmology, 1993. **100**(4): p. 456-9.
18. Esteban, J., et al., *Peritonitis involving a Capnocytophaga sp. in a patient undergoing continuous ambulatory peritoneal dialysis.* Journal of clinical microbiology, 1995. **33**(9): p. 2471-2.
19. Campbell, J.R. and M.S. Edwards, *Capnocytophaga species infections in children.* The Pediatric infectious disease journal, 1991. **10**(12): p. 944-8.
20. Font, R.L., et al., *Capnocytophaga keratitis. A clinicopathologic study of three patients, including electron microscopic observations.* Ophthalmology, 1994. **101**(12): p. 1929-34.
21. Desai, S.S., R.A. Harrison, and M.D. Murphy, *Capnocytophaga ochracea causing severe sepsis and purpura fulminans in an immunocompetent patient.* The Journal of infection, 2007. **54**(2): p. e107-9.
22. Leadbetter, E.R., S.C. Holt, and S.S. Socransky, *Capnocytophaga: new genus of gram-negative gliding bacteria. I. General characteristics, taxonomic considerations and significance.* Archives of microbiology, 1979. **122**(1): p. 9-16.
23. Holt, S.C., E.R. Leadbetter, and S.S. Socransky, *Capnocytophaga: new genus of gram-negative gliding bacteria. II. Morphology and ultrastructure.* Archives of microbiology, 1979. **122**(1): p. 17-27.
24. Socransky, S.S., et al., *Capnocytophaga: new genus of gram-negative gliding bacteria. III. Physiological characterization.* Archives of microbiology, 1979. **122**(1): p. 29-33.
25. Williams, B.L. and B.F. Hammond, *Capnocytophaga: new genus of gram-negative gliding bacteria. IV. DNA base composition and sequence homology.* Archives of microbiology, 1979. **122**(1): p. 35-9.
26. Nolan M., T.B.J., Pomrenke H., Lapidus A., Copeland A., Glavina del Rio T., Lucas S., Chen F., Tice H., Cheng J.F., Saunders E., Han C., Bruce D., Goodwin L., Chain P., Pitluck S., Ovchinnikova G., Pati A., Ivanova N., Mavromatis K., Chen A., Palaniappan K., Land M., Hauser L., Chang Y.J., Jeffries C.D., Brettin T., Goker M., Bristow J., Eisen J.A., Markowitz V., Hugenholtz P., Kyrpides N.C., Klenk H.P., Detter J.C., *Complete genome sequence of Rhodothermus marinus type strain (R-10T).* Standards in Genomic Sciences, 2009. **1**(3): p. 283-291.
27. Raymond, J.A., B.C. Christner, and S.C. Schuster, *A bacterial ice-binding protein from the Vostok ice core.* Extremophiles, 2008. **12**(5): p. 713-7.
28. Mongodin, E.F., et al., *The genome of Salinibacter ruber: convergence and gene exchange among hyperhalophilic bacteria and archaea.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(50): p. 18147-52.
29. Agarwal, S., D.W. Hunnicutt, and M.J. McBride, *Cloning and characterization of the Flavobacterium johnsoniae (Cytophaga johnsonae) gliding motility gene, gldA.* Proceedings of the National

Academy of Sciences of the United States of America, 1997. **94**(22): p. 12139-44.

30. Vancanneyt, M., et al., *Larkinella insperata gen. nov., sp. nov., a bacterium of the phylum 'Bacteroidetes' isolated from water of a steam generator.* International journal of systematic and evolutionary microbiology, 2006. **56**(Pt 1): p. 237-41.
31. Lopez-Sanchez, M.J., et al., *Evolutionary convergence and nitrogen metabolism in Blattabacterium strain Bge, primary endosymbiont of the cockroach Blattella germanica.* PLoS genetics, 2009. **5**(11): p. e1000721.
32. Sabree, Z.L., S. Kambhampati, and N.A. Moran, *Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(46): p. 19521-6.
33. Hongoh, Y., et al., *Genome of an endosymbiont coupling N2 fixation to cellulolysis within protist cells in termite gut.* Science (New York, N.Y, 2008. **322**(5904): p. 1108-9.
34. Schmitz-Esser, S., et al., *The genome of the amoeba symbiont "Candidatus Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria.* Journal of bacteriology. **192**(4): p. 1045-57.
35. Kuwahara, T., et al., *Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(41): p. 14919-24.
36. Xu, J., et al., *A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.* Science (New York, N.Y, 2003. **299**(5615): p. 2074-6.
37. Xu, J., et al., *Evolution of symbiotic bacteria in the distal human intestine.* PLoS biology, 2007. **5**(7): p. e156.
38. Nelson, K.E., et al., *Complete genome sequence of the oral pathogenic Bacterium porphyromonas gingivalis strain W83.* Journal of bacteriology, 2003. **185**(18): p. 5591-601.
39. Schuijffel, D.F., et al., *Successful selection of cross-protective vaccine candidates for Ornithobacterium rhinotracheale infection.* Infection and immunity, 2005. **73**(10): p. 6812-21.
40. Crasta, K.C., et al., *Identification and characterization of CAMP cohemolysin as a potential virulence factor of Riemerella anatipestifer.* Journal of bacteriology, 2002. **184**(7): p. 1932-9.
41. Duchaud, E., et al., *Complete genome sequence of the fish pathogen Flavobacterium psychrophilum.* Nature biotechnology, 2007. **25**(7): p. 763-9.
42. Zaura, E., et al., *Defining the healthy "core microbiome" of oral microbial communities.* BMC microbiology, 2009. **9**: p. 259.
43. Socransky, S.S. and A.D. Haffajee, *Periodontal microbial ecology.* Periodontology 2000, 2005. **38**: p. 135-87.
44. Kolenbrander, P.E., et al., *Bacterial interactions and successions during plaque development.* Periodontology 2000, 2006. **42**: p. 47-79.

45. Dilegge, S.K., V.P. Edgcomb, and E.R. Leadbetter, *Presence of the oral bacterium Capnocytophaga canimorsus in the tooth plaque of canines.* Veterinary microbiology, 2010.
46. Riep, B., et al., *Are putative periodontal pathogens reliable diagnostic markers?* Journal of clinical microbiology, 2009. **47**(6): p. 1705-11.
47. Jiang, W., et al., *Investigation of Supragingival Plaque Microbiota in Different Caries Status of Chinese Preschool Children by Denaturing Gradient Gel Electrophoresis.* Microbial ecology.
48. Colombo, A.P., et al., *Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray.* Journal of periodontology, 2009. **80**(9): p. 1421-32.
49. Krasse, B., *Serendipity or luck: stumbling on gingival crevicular fluid.* Journal of dental research, 1996. **75**(9): p. 1627-30.
50. Mally, M., et al., *Prevalence of Capnocytophaga canimorsus in dogs and occurrence of potential virulence factors.* Microbes and infection / Institut Pasteur, 2009. **11**(4): p. 509-14.
51. Blanche, P., E. Bloch, and D. Sicard, *Capnocytophaga canimorsus in the oral flora of dogs and cats.* The Journal of infection, 1998. **36**(1): p. 134.
52. Chauncey, H.H., B.L. Henrigues, and J.M. Tanzer, *Comparative Enzyme Activity of Saliva from the Sheep, Hog, Dog, Rabbit, Rat, and Human.* Archives of oral biology, 1963. **8**: p. 615-27.
53. Lipner, H.J., *Physiological considerations of the relative specificity of dental caries in man.* Journal of dental research, 1947. **26**(4): p. 319-25.
54. Westwell, A.J., et al., *DF-2 infection.* BMJ (Clinical research ed, 1989. **298**(6666): p. 116-7.
55. Shin, H., et al., *Resistance of Capnocytophaga canimorsus to killing by human complement and polymorphonuclear leukocytes.* Infection and immunity, 2009. **77**(6): p. 2262-71.
56. Shin, H., et al., *Escape from immune surveillance by Capnocytophaga canimorsus.* The Journal of infectious diseases, 2007. **195**(3): p. 375-86.
57. Meyer, S., H. Shin, and G.R. Cornelis, *Capnocytophaga canimorsus resists phagocytosis by macrophages and blocks the ability of macrophages to kill other bacteria.* Immunobiology, 2008. **213**(9-10): p. 805-14.
58. Mally, M., et al., *Capnocytophaga canimorsus: a human pathogen feeding at the surface of epithelial cells and phagocytes.* PLoS pathogens, 2008. **4**(9): p. e1000164.
59. Mavromatis, K., Sabine Gronow, Elizabeth Saunders, Miriam Land, Alla, et al.,, *Complete genome sequence of Capnocytophaga ochracea type strain (VPI 2845T).* Lawrence Berkeley National Laboratory, 2010.
60. McBride, M.J., et al., *Novel features of the polysaccharide-digesting gliding bacterium Flavobacterium johnsoniae as revealed by genome sequence analysis.* Applied and environmental microbiology, 2009. **75**(21): p. 6864-75.
61. Song, H., et al., *The early stage of bacterial genome-reductive evolution in the host.* PLoS pathogens, 2010. **6**(5): p. e1000922.

62. Kapke, P.A., A.T. Brown, and T.T. Lillich, *Carbon dioxide metabolism by Capnocytophaga ochracea: identification, characterization, and regulation of a phosphoenolpyruvate carboxykinase.* Infection and immunity, 1980. **27**(3): p. 756-66.
63. Unden, G., H. Hackenberg, and A. Kroger, *Isolation and functional aspects of the fumarate reductase involved in the phosphorylative electron transport of Vibrio succinogenes.* Biochimica et biophysica acta, 1980. **591**(2): p. 275-88.
64. Madej, M.G., et al., *Limited reversibility of transmembrane proton transfer assisting transmembrane electron transfer in a dihaem-containing succinate:quinone oxidoreductase.* Biochimica et biophysica acta, 2009. **1787**(6): p. 593-600.
65. Calmes, R., et al., *Energy metabolism in Capnocytophaga ochracea.* Infection and immunity, 1980. **29**(2): p. 551-60.
66. Macfarlane, S. and G.T. Macfarlane, *Regulation of short-chain fatty acid production.* The Proceedings of the Nutrition Society, 2003. **62**(1): p. 67-72.
67. Kolenbrander, P.E., et al., *Communication among oral bacteria.* Microbiology and molecular biology reviews, 2002. **66**(3): p. 486-505, table of contents.
68. McBride, M.J., *Cytophaga-flavobacterium gliding motility.* Journal of molecular microbiology and biotechnology, 2004. **7**(1-2): p. 63-71.
69. Sato, K., et al., *A protein secretion system linked to bacteroidete gliding motility and pathogenesis.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(1): p. 276-81.
70. Shoji, M., et al., *The major structural components of two cell surface filaments of Porphyromonas gingivalis are matured through lipoprotein precursors.* Molecular microbiology, 2004. **52**(5): p. 1513-25.
71. Shipman, J.A., J.E. Berleman, and A.A. Salyers, *Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron.* Journal of bacteriology, 2000. **182**(19): p. 5365-72.
72. Shipman, J.A., et al., *Physiological characterization of SusG, an outer membrane protein essential for starch utilization by Bacteroides thetaiotaomicron.* Journal of bacteriology, 1999. **181**(23): p. 7206-11.
73. Martens, E.C., H.C. Chiang, and J.I. Gordon, *Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont.* Cell host & microbe, 2008. **4**(5): p. 447-57.
74. Martens, E.C., et al., *Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm.* The Journal of biological chemistry, 2009. **284**(37): p. 24673-7.
75. Bjursell, M.K., E.C. Martens, and J.I. Gordon, *Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period.* The Journal of biological chemistry, 2006. **281**(47): p. 36269-79.
76. Reeves, A.R., et al., *A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch.* Journal of bacteriology, 1996. **178**(3): p. 823-30.
77. Reeves, A.R., G.R. Wang, and A.A. Salyers, *Characterization of four outer membrane proteins that play a role in utilization of starch by*

Bacteroides thetaiotaomicron. Journal of bacteriology, 1997. **179**(3): p. 643-9.

78. Hehemann, J.H., et al., *Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota.* Nature, 2010. **464**(7290): p. 908-12.
79. Bauer, M., et al., *Whole genome analysis of the marine Bacteroidetes'Gramella forsetii' reveals adaptations to degradation of polymeric organic matter.* Environmental microbiology, 2006. **8**(12): p. 2201-13.
80. Cho, K.H. and A.A. Salyers, *Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron.* Journal of bacteriology, 2001. **183**(24): p. 7224-30.
81. Kolenbrander, P.E., et al., *Oral multispecies biofilm development and the key role of cell-cell distance.* Nature reviews, 2010. **8**(7): p. 471-480.
82. Mally, M. and G.R. Cornelis, *Genetic tools for studying Capnocytophaga canimorsus.* Applied and environmental microbiology, 2008. **74**(20): p. 6369-77.
83. Ewing, B., et al., *Base-calling of automated sequencer traces using phred. I. Accuracy assessment.* Genome research, 1998. **8**(3): p. 175-85.
84. Ewing, B. and P. Green, *Base-calling of automated sequencer traces using phred. II. Error probabilities.* Genome research, 1998. **8**(3): p. 186-94.
85. Gordon, D., C. Abajian, and P. Green, *Consed: a graphical tool for sequence finishing.* Genome research, 1998. **8**(3): p. 195-202.
86. Li, H., J. Ruan, and R. Durbin, *Mapping short DNA sequencing reads and calling variants using mapping quality scores.* Genome research, 2008. **18**(11): p. 1851-8.
87. Delcher, A.L., et al., *Identifying bacterial genes and endosymbiont DNA with Glimmer.* Bioinformatics (Oxford, England), 2007. **23**(6): p. 673-9.
88. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic acids research, 1997. **25**(17): p. 3389-402.
89. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite.* Trends in genetics, 2000. **16**(6): p. 276-7.
90. Zdobnov, E.M. and R. Apweiler, *InterProScan--an integration platform for the signature-recognition methods in InterPro.* Bioinformatics (Oxford, England), 2001. **17**(9): p. 847-8.
91. Claudel-Renard, C., et al., *Enzyme-specific profiles for genome annotation: PRIAM.* Nucleic acids research, 2003. **31**(22): p. 6633-9.
92. Jensen, L.J., et al., *STRING 8--a global view on proteins and their functional interactions in 630 organisms.* Nucleic acids research, 2009. **37**(Database issue): p. D412-6.
93. Nawrocki, E.P., D.L. Kolbe, and S.R. Eddy, *Infernal 1.0: inference of RNA alignments.* Bioinformatics (Oxford, England), 2009. **25**(10): p. 1335-7.

94. Lagesen, K., et al., *RNAmmer: consistent and rapid annotation of ribosomal RNA genes.* Nucleic acids research, 2007. **35**(9): p. 3100-8.
95. Lowe, T.M. and S.R. Eddy, *tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.* Nucleic acids research, 1997. **25**(5): p. 955-64.
96. Vernikos, G.S. and J. Parkhill, *Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands.* Bioinformatics (Oxford, England), 2006. **22**(18): p. 2196-203.
97. Chen, F., et al., *Assessing performance of orthology detection strategies applied to eukaryotic genomes.* PloS one, 2007. **2**(4): p. e383.
98. Felsenstein, J., *Mathematics vs. Evolution: Mathematical Evolutionary Theory.* Science (New York, N.Y, 1989. **246**(4932): p. 941-2.
99. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic acids research, 1994. **22**(22): p. 4673-80.
100. Hwang, T.L. and A.J. Shaka, *Multiple-pulse mixing sequences that selectively enhance chemical exchange or cross-relaxation peaks in high-resolution NMR spectra.* Journal of magnetic resonance (San Diego, Calif., 1998. **135**(2): p. 280-7.
101. Doro, F., et al., *Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for Group B Streptococcus hypervirulent strain COH1.* Molecular & cellular proteomics, 2009. **8**(7): p. 1728-37.
102. Walters, M.S. and H.L. Mobley, *Identification of uropathogenic Escherichia coli surface proteins by shotgun proteomics.* Journal of microbiological methods, 2009. **78**(2): p. 131-5.
103. Crump, E.M., et al., *Antigenic characterization of the fish pathogen Flavobacterium psychrophilum.* Applied and environmental microbiology, 2001. **67**(2): p. 750-9.
104. Kristian, S.A., et al., *Alanylation of teichoic acids protects Staphylococcus aureus against Toll-like receptor 2-dependent host defense in a mouse tissue cage infection model.* The Journal of infectious diseases, 2003. **188**(3): p. 414-23.
105. Stothard, P. and D.S. Wishart, *Circular genome visualization and exploration using CGView.* Bioinformatics (Oxford, England), 2005. **21**(4): p. 537-9.
106. Karp, P.D., et al., *Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology.* Briefings in bioinformatics, 2010. **11**(1): p. 40-79.
107. Supek, F. and K. Vlahovicek, *INCA: synonymous codon usage analysis and clustering by means of self-organizing map.* Bioinformatics (Oxford, England), 2004. **20**(14): p. 2329-30.
108. Stepanenko, O.V., et al., *Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes.* Current protein & peptide science, 2008. **9**(4): p. 338-69.

109. Bailie, W.E., E.C. Stowe, and A.M. Schmitt, *Aerobic bacterial flora of oral and nasal fluids of canines with reference to bacteria associated with bites.* Journal of clinical microbiology, 1978. **7**(2): p. 223-31.
110. Koropatkin, N.M., et al., *Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices.* Structure (London, England, 2008. **16**(7): p. 1105-15.
111. King, S.J., K.R. Hippe, and J.N. Weiser, *Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae.* Molecular microbiology, 2006. **59**(3): p. 961-74.
112. Burnaugh, A.M., L.J. Frantz, and S.J. King, *Growth of Streptococcus pneumoniae on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases.* Journal of bacteriology, 2008. **190**(1): p. 221-30.
113. Byers, H.L., et al., *Sequential deglycosylation and utilization of the N-linked, complex-type glycans of human alpha1-acid glycoprotein mediates growth of Streptococcus oralis.* Glycobiology, 1999. **9**(5): p. 469-79.
114. Green, E.D., et al., *The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy.* The Journal of biological chemistry, 1988. **263**(34): p. 18253-68.
115. Collin, M. and A. Olsen, *EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG.* The EMBO journal, 2001. **20**(12): p. 3046-55.
116. Elder, J.H. and S. Alexander, *endo-beta-N-acetylglucosaminidase F: endoglycosidase from Flavobacterium meningosepticum that cleaves both high-mannose and complex glycoproteins.* Proceedings of the National Academy of Sciences of the United States of America, 1982. **79**(15): p. 4540-4.
117. Kovacs-Simon, A., R.W. Titball, and S.L. Michell, *Lipoproteins of bacterial pathogens.* Infection and immunity, 2010. **79**(2): p. 548-61.
118. Chen, S., et al., *Mutational analysis of the ompA promoter from Flavobacterium johnsoniae.* Journal of bacteriology, 2007. **189**(14): p. 5108-18.
119. Chen, S., et al., *Characterization of strong promoters from an environmental Flavobacterium hibernum strain by using a green fluorescent protein-based reporter system.* Applied and environmental microbiology, 2007. **73**(4): p. 1089-100.
120. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome research, 2008. **18**(5): p. 821-9.
121. Hernandez, D., et al., *De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer.* Genome research, 2008. **18**(5): p. 802-9.
122. Warren, R.L., et al., *Assembling millions of short DNA sequences using SSAKE.* Bioinformatics (Oxford, England), 2007. **23**(4): p. 500-1.
123. Carlsson, F., et al., *Evasion of phagocytosis through cooperation between two ligand-binding regions in Streptococcus pyogenes M*

protein. The Journal of experimental medicine, 2003. **198**(7): p. 1057-68.

124. Simon R, P.U., & Puhler A, *A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria.* Nat Biotech., 1983. **1**(9): p. 784-791.

9. Appendix

Strains and plasmids

strains	Description or genotype	Reference	Ch.
E. coli			
S17-1	hsdR17 recA1 RP4-2-*tet::*Mu1*kan::*Tn7; Smr	[124]	4.1&5.1
C. canimorsus			
Cc5ΔPUL1	Site directed mutation of PUL1 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL2	Site directed mutation of PUL2 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL3	Site directed mutation of PUL3 by partial replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL4	Site directed mutation of PUL4 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL6	Site directed mutation of PUL6 by partial replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL7	Site directed mutation of PUL7 by partial replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL8	Site directed mutation of PUL8 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL9	Site directed mutation of PUL9 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL10	Site directed mutation of PUL10 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL11	Site directed mutation of PUL11 by partail replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL12	Site directed mutation of PUL12 by replacement with *ermF;* Emr	This study	4.1
Cc5ΔPUL13	Site directed mutation of PUL13 by replacement with *ermF;* Emr	This study	4.1
Cc5 Δcyt	Site directed mutation of Ccan_04790 by replacement with ermF; Emr	This study	4.1
Cc5	Human fatal septicemia after dog bite 1995	[56]	4.1&5.1
Cc5ΔsiaC	Replacement of *Ccan_00790* by *ermF;* Emr	[58]	4.1&5.1
Cc5ΔPUL5	Replacement of *Ccan_08700, Ccan_08710, Ccan_08720, Ccan_08730* by *ermF* : Emr	This study	4.1&5.1
Cc5ΔgpdC	Replacement of *Ccan_08700* by *ermF* using primers 5073, 5074, 5075, 5083*;* Emr	This study	5.1
Cc5Δ gpdD	Replacement of *Ccan_08710* by *ermF* using primers 4850, 4851,4854, 4855*;* Emr	This study	5.1
Cc5ΔgpdG	Replacement of *Ccan_08720* by *ermF* using primers 5001, 5002, 5005, 5006*;* Emr	This study	5.1
Cc5ΔgpdE	Replacement of *Ccan_08730* by *ermF* using primers 5951, 5952, 5953, 5954*;* Emr	This study	5.1
Cc5ΔgpdF	Replacement of *Ccan_08740* by *ermF* using primers 5955, 5956, 5957, 5958*;* Emr	This study	5.1

Plasmid	Description	Reference	Chapter
pMM47.A	Ori$_{ColE1}$, *ori*$_{pCC7}$, Ap^r ,Cf^r, *E. coli - C. canimorsus* expression shuttle vector.	[82]	5.1
pPM1	pMM47.A where the *ermF* promoter has been replaced by the stronger *gpd* promoter: 117bp upstream of the *gpdC* ORF start codon were amplified with primers 5081 and 5469 and cloned into pMM47.A using *Sal*I and *Nco*I restriction sites.	This study	5.1
pPM2	Full length *gpdC* containing its putative promoter region amplified with primers 5081 and 5082 and cloned into pMM47.A using *Sal*I and *Spe*I restriction sites.	This study	5.1
pPM3	Full length *gpdC* with a C-terminal His-Strep double tag amplified by 2-step overlapping PCR with primers 5081, 5467 and 5530 and cloned into pMM47.A using *Sal*I and *Spe*I restriction sites.	This study	5.1
pFR4	Full length *gpdD* amplified with primers 6133 and 6057 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pFR5	Full length *gpdG* amplified with primers 5008 and 6055 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pFR6	Full length *gpdE* amplified with primers 5959 and 5060 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pFR7	Full length *gpdF* amplified with primers 5062 and 5063 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pFR8	Full length *gpdD* with a C17G point mutation amplified with primers 6056 and 6057 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pFR9	Full length *gpdG* with a C21G point mutation amplified with primers 6054 and 6055 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pFR10	Full length *gpdG* with a E205G point mutation amplified by overlapping PCR using primers 5008/6061 and 6060/6055 and cloned in pPM1 using *Nco*I and *Xba*I restriction sites.	This study	5.1
pMM121.1	Full length *siaC* amplified by inverse PCR using primers 5045 + 5046 on pMM52 as a template to insert to C17Y substitution in *siaC*.	This study	5.1
pMM25	oriColE1 , Kmr , Cfr .Suicide vector for C. canimorsus.	[82]	5.1
pMM52	Full length *siaC* with a C-terminal His tag cloned in pMM47.A using *Nco*I and *Xba*I restriction sites.	[58]	5.1
pMM106	ori$_{ColE1}$, Km^r , Cf^r , Ery^R , Mutator plasmid for the replacement of *siaC* by *ermF*	[82]	5.1

Oligonucleotides

Ref.	Name	Sequence 5'-3'	Restr.	Gene	PCR	Ch.
5508	fwd_PUL9_1.1	CCCTGCAGCGCCTAAAAAGAGCCC	PstI	*PUL9*	A	4.1
5509	rev_PUL9_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGTAAAGACTCAATACAAGCGG		*PUL9*	B	4.1
5510	fwd_PUL9_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTTTCATATCTGATTTTTGG		*PUL9*	C	4.1
5511	rev_PUL9_2.2	GCACTAGTACGCGGATTTCCAACCTG	SpeI	*PUL9*	D	4.1
5512	fwd_PUL10_1.1	CCCTGCAGGGTATCGGCTGTATTAGCC	PstI	*PUL10*	A	4.1
5513	rev_PUL10_1.2	GAAGCTATCGGAGTAGATAAAAGCACTGTTGTAGAGGTTGTTAAATTTGTC		*PUL10*	B	4.1
5514	fwd_PUL10_2.1	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAAATAGAATATAATTTTTTG		*PUL10*	C	4.1
5515	rev_PUL10_2.2	GGACTAGTGGCTAATAAAAAGCCAATAACC	SpeI	*PUL10*	D	4.1
5520	fwd_PUL11_1.1	GGCTGCAGTTCTTTAATGATTTATAGCG	PstI	*PUL11*	A	4.1
5521	rev_PUL11_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGTAAGAAAGCATATGGC		*PUL11*	B	4.1
5522	fwd_PUL11_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTACTTTTTTATTCAATG		*PUL11*	C	4.1
5523	rev_PUL11_2.2	GCACTAGTAAAGTGAGTAAACATTCCCG	SpeI	*PUL11*	D	4.1
5566	fwd_PUL1_1.1	GGCTGCAGGCAATGACTAATAAGTTAGG	PstI	*PUL1*	A	4.1
5567	rev_PUL1_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGCCAAGTTAATTTTAATCTC		*PUL1*	B	4.1
5568	fwd_PUL1_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTTCAATTAAAAATTTCCAACAC		*PUL1*	C	4.1
5569	rev_PUL1_2.2	GCACTAGTTGAAAAAGTGGGATTAGATGC	SpeI	*PUL1*	D	4.1
5570	fwd_PUL2_1.1	GGCTGCAGGCTCTTTTAAAAGCACTATAAAGG	PstI	*PUL2*	A	4.1
5571	rev_PUL2_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAACAACTGGCATCAAGAAGAGC		*PUL2*	B	4.1
5572	fwd_PUL2_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTAAAAACGGAACGTTG		*PUL2*	C	4.1
5573	rev_PUL2_2.2	GCACTAGTATGACCAAAAAGATGCTGG	SpeI	*PUL2*	D	4.1
5574	fwd_00780-820 _1.1	GGCTGCAGGGCAAAAACTTCGGGAAAACC	PstI	*00780-820*	A	4.1
5575	rev_00780-820_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAATTGACAGCAATAATAAC		*00780-820*	B	4.1
5576	fwd_00780-820_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTAGAAATATACTTTTTCATAATC		*00780-820*	C	4.1
5577	rev_00780-820_2.2	GCACTAGTCAGATTCTCCCCATTGCTTTACC	SpeI	*00780-820*	D	4.1
5639	fwd_PUL5_1.1	GGCTGCAGGTATTAGAAGAATATTTTCC	PstI	*PUL5*	A	4.1
5640	rev_PUL5_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGGTTAATAATTATTTCAAAACAAACTAACGCG		*PUL5*	B	4.1
5641	fwd_PUL5_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTTAGTAAGAGTAATAAAGAATGCC		*PUL5*	C	4.1
5642	rev_PUL5_2.2	GCACTAGTTTATCTTCACTCGAAATAGCCTCTCCC	SpeI	*PUL5*	D	4.1
5740	fwd_PUL6_1.1	GGCTGCAGTGTACGCCTATTTGGAACAGGC	PstI	*PUL6*	A	4.1
5741	rev_PUL6_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAGGTAGAAGGTAAAATTTGAATTTATCC		*PUL6*	B	4.1
5742	fwd_PUL6_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTATTTATACGTTTTTTATGAGAAAAATAATTCC		*PUL6*	C	4.1
5743	rev_PUL6_2.2	GCACTAGTTAAGTTATAGATCGCTTTTTCAAAATCGG	SpeI	*PUL6*	D	4.1
5873	fwd_PUL7_1.1	GGCTGCAGATGCGCTATTGCTTCCTGAGG	PstI	*PUL7*	A	4.1
5874	rev_PUL7_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGGTCAATTTAAATGTTTGATAATGAG		*PUL7*	B	4.1
5875	fwd_PUL7_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTAGTAAAAATTTAGACTAATG		*PUL7*	C	4.1
5876	rev_PUL7_2.2	GCACTAGTGTAATTGTAAATCATATCACGAAGCG	SpeI	*PUL7*	D	4.1
5877	fwd_PUL8_1.1	GGCTGCAGGGCAATTGACTATATTTGGG	PstI	*PUL8*	A	4.1
5878	rev_PUL8_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGTTTTTTATCGAGGAGTTAGTTC		*PUL8*	B	4.1
5879	fwd_PUL8_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTAAGTAATGTACAAATTTGC		*PUL8*	C	4.1
5880	rev_PUL8_2.2	GCACTAGTGCGTGTTTGGGCTCTTCTTG	SpeI	*PUL8*	D	4.1
5881	fwd_PUL12_1.1	GGCTGCAGCTGGGTGATGTTTTTCGTGG	PstI	*PUL12*	A	4.1
5882	rev_PUL12_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAGTTCATAAAATTAGTTCATAGC		*PUL12*	B	4.1
5883	fwd_PUL12_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTATAAATATCTTTTAGATTAAAC		*PUL12*	C	4.1
5884	rev_PUL12_2.2	GCACTAGTAAGTCGTGAGCAATTTCTGG	SpeI	*PUL12*	D	4.1
5885	fwd_PUL13_1.1	GGCTGCAGGACAAAAATATGAACTATAAATTTG	PstI	*PUL13*	A	4.1
5886	rev_PUL13_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGGTAAAAAGGATAAAGTAGAAAATG		*PUL13*	B	4.1
5887	fwd_PUL13_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTTCAGGTATAATGGACAAAAATTAGGC		*PUL13*	C	4.1
5888	rev_PUL13_2.2	GCACTAGTTCTAAATGAAAGAACTATTAATCC	SpeI	*PUL13*	D	4.1
5889	fwd_PUL3_1.1	GGCTGCAGCATATTGCTTAAAGTTAATAAATC	PstI	*PUL3*	A	4.1
5890	rev_PUL3_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAAAAACTTCTTACGATTTTTATTTAG		*PUL3*	B	4.1
5891	fwd_PUL3_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTTTTGTAAGGAAGGGACGTGTCC		*PUL3*	C	4.1

5892	rev_PUL3_2.2_SpeI	GCACTAGTCCTTCTCATCGAAATTATTGAC	SpeI	*PUL3*	D	4.1
5893	fwd_PUL4_1.1_PstI	GGCTGCAGGGCTCAACGCTCCGTATTGTAAACC	PstI	*PUL4*	A	4.1
5894	rev_PUL4_1.2	GTTGTCCCTGAAAAATTTCATCCTTCGTAGCCTTAATTGTATCTACTGAGAG		*PUL4*	B	4.1
5895	fwd_PUL4_2.1	GAAGCTATCGGAGTAGATAAAAGCACTGTTAAAAAATGTTAACAAGAATCTTTTCC		*PUL4*	C	4.1
5896	rev_PUL4_2.2_SpeI	GCACTAGTGGCCAAAGTACCTGTTTTATTCCG	SpeI	*PUL4*	D	4.1
5502	ermF-fw_2.1	CTACGAAGGATGAAATTTTTCAGGGACAAC		*ermF*		4.1 & 5.1
5503	ermF-rev_2.2	AACAGTGCTTTTATCTACTCCGATAGCTTC		*ermF*		4.1 & 5.1
4850	gpdDKO-1.1-fw	CCCTGCAGTTAATAAGAAATGAAAAAATAC	PstI	*gpdD*	A	5.1
4851	gpdDKO-1.2-rev	GAGTAGATAAAAGCACTGTTAATACGGTAAGGGACCAAAC		*gpdD*	B	5.1
4854	gpdDKO-2.1-fw	AAAAATTTCATCCTTCGTAGTTCTGAAAATGGGGTAAGCA		*gpdD*	C	5.1
4855	gpdDKO-2.2-rev	CCACTAGTAAGATTATCTTGTATTAGGATTC	SpeI	*gpdD*	D	5.1
5001	gpdGKO-1.1-fw	CGCTGCAGGATTGTAATACCCATCTTTG	PstI	*gpdG*	A	5.1
5002	gpdGKO-1.2-rev	GAGTAGATAAAAGCACTGTTGAGACTTGATAACAAGTAAA		*gpdG*	B	5.1
5005	gpdGKO-2.1-fw	AAAAATTTCATCCTTCGTAGTTACTTTGATAAGTATATTA			C	5.1
5006	gpdGKO-2.2-rev	CCACTAGTCTGACGCCAAATTAGAGTCA	SpeI	*gpdG*	D	5.1
5008	gpdG-fw	CATGCCATGGGAAAAAAAAATATTATAAAATGGGG	NcoI	*gpdG*		5.1
5045	siaCCys-fw	CTTTTGTCGGCTTATGGAAGCCAAAAA		*siaC*		5.1
5046	siaCCys-rev	TTTTTGGCTTCCATAAGCCGACAAAAG		*siaC*		5.1
5073	gpdCKO-1.1-fw	CCCTGCAGActtatagctcttgcgtgcggactttgg	PstI	*gpdC*	A	5.1
5074	gpdCKO-1.2-rev	GAGTAGATAAAAGCACTGTTgcacttcgttgaatgttaatgccagcca		*gpdC*	B	5.1
5075	gpdCKO-2.1-fw	AAAAATTTCATCCTTCGTAGtgaaggcggttcaatgacagcagtg		*gpdC*	C	5.1
5081	PgpdC-fw	CGATGTCGACtgaatatgttgtacatttgtg	SalI			5.1
5082	gpdC-rev	CCACTAGTacctataatgaagctttaattgc	SpeI	*gpdC*		5.1
5083	gpdCKO-2.2-rev	CCACTAGTattcgggatcaaaaggcgctgacaa	SpeI	*gpdC*	D	5.1
5467	gpdC-His-rev	tgACTAGTTAatgatgatgatgatgatgAGCACCACCAGCACCACCtAATGAAGCTTTAATTGCAATACC	SpeI	*gpdC*		5.1
5469	PgpdC-rev	CATACCATGGcaataataaaatgaattag	NcoI			5.1
5530	gpdC-Strep-rev	TgACTAGTTATTTTTCAAATTGAGGATGTGACCAAGCTCCTCCAGCTCCTCCatgatgatgatgatgatgAGC	SpeI	*gpdC*		5.1
5951	gpdEKO-1.1-fw	GGCTGCAGCGGTTACCATCCACAAGAGAAAG	PstI	*gpdE*	A	5.1
5952	gpdEKO-1.2-rev	GTTGTCCCTGAAAAATTTCATCCTTCGTAGAATTTACTATTTTTTAGGTAATCTG		*gpdE*	B	5.1
5953	gpdEKO-2.1-fw	GAAGCTATCGGAGTAGATAAAAGCACTGTTGATTTCCTAATGTTGATTTTAATACC		*gpdE*	C	5.1
5954	gpdEKO-2.2-rev	GCACTAGTGGGTGAGACATCAGATACTTG	SpeI	*gpdE*	D	5.1
5955	gpdFKO-1.1-fw	GGCTGCAGGTTTGAAGCAGCGGGTACTAATCC	PstI	*gpdF*	A	5.1
5956	gpdFKO-1.2-rev	GTTGTCCCTGAAAAATTTCATCCTTCGTAGCCCTACCAGTAATACTGTTGTGAG		*gpdF*	B	5.1
5957	gpdFKO-2.1-fw	GAAGCTATCGGAGTAGATAAAAGCACTGTTGGGAGGAGATCAATATGTTGATATAAATG		*gpdF*	C	5.1
5958	gpdFKO-2.2-rev	GCACTAGTCGGCTTTTTCGAATGAAACGAAC	SpeI	*gpdF*	D	5.1
5959	gpdE-fw	CATACCATGGGAAAGAAATTACATATCTTATTTGTTATCG	NcoI	*gpdE*		5.1
5960	gpdE-rev	GCTCTAGATTAAAATTCTACTTTGGTATTAAAATC	XbaI	*gpdE*		5.1
5962	gpdF-fw	CATACCATGGGAAAAAAACATATAAAAATTTTATTTCTCACAACAG	NcoI	*gpdF*		5.1
5963	gpdF-rev	GCTCTAGACTAATAAAATTCTAATTCATTTATATCAAC	XbaI	*gpdF*		5.1
6054	gpdGCys-fw	CATACCATGGGAAAAAAAAATATTATAAAATGGGGTTTAGCAATACTTATAGGGGTAGCTTCTGTAA	NcoI	*gpdG*		5.1
6055	gpdG-rev	GCTCTAGACTATTTTTTAGGTAATCTGATAATTAATTGCTC	XbaI	*gpdG*		5.1
6056	gpdDCys-fw	CATACCATGGGAAAAAAATACTTTATGATAGGTGCTTTATCTTTAGCTACAATTTCTGGTACGAAAG	NcoI	*gpdD*		5.1
6057	gpdD-rev	GCTCTAGATTATCTTGTATTAGGATTCACATCCCACC	XbaI	*gpdD*		5.1
6060	gpdG-E /G-fw	CCAAAAGATATTGACTGGGGACCTACTGTGGGTAATCATGGAAG		*gpdG*		5.1
6061	gpdG-E /G-rev	CTTCCATGATTACCCACAGTAGGTCCCCAGTCAATATCTTTTGG		*gpdG*		5.1
6133	gpdD-fw	CATACCATGGGAAAAAAATACTTTATGATAGGTGCTTTATCTTTAGC	NcoI	*gpdD*		5.1

10. Acknowledgments

I dedicate this work to my father, Dr. Eduardo Manfredi, and my mother Patricia Manfredi that are, after all, my main funders and my earliest support. My first thanks go to my Wife, Dr. Med. Dominique Cenedese, for being a motor for me in my career and in all other aspects of my life. I also want to thanks our son Eden that reminds me every day that scientist are not just a sum of papers and diploma and that there is a “real” life out of the lab. Many thanks to my two brothers Guido an Martin Manfredi for their support during all these years.

Thanks to Dr. Marco Pagni not only for being referee on this work but also for being my unique support in the field of bioinformatics. He was like a lighthouse for me during the first years. Many thanks to Prof. Dr. Christoph Dehio for reviewing my work and sharing his expertise through comments and advices.

Heartfelt thanks go to Dr. Francesco Renzi for our long term collaboration and its constructive criticism always embedded in humor. I am also deeply grateful to the master students that worked successively on my projects : Loïc Sauteur, Silvia Pietsch and Frédéric Lauber. Special thanks go to Dr. Simon Ittig for its energy and optimism; one gets some of it just by standing next to him.

My sincere appreciation to the rest of the Cc team: Manuella Mally, Hwain Shin, Salome Meyer and Cecile Paroz-Pfaff for their help and support. Thanks to Mathias Schmaler for the animal experiments. Special thanks to Dr. Philip Engel for his support. I think that for a while we were the only two able to write a script in the whole floor. I also would like to acknowledge my closest collaborators from the rest of the Cornelis team - Marlise Amstutz, Andreas Diepold, Isabel Sorg, Caroline Montagner and Steffi Wagner - , the MS people that helped us so much – Dr. Paul Jenö & Suzette Moes -, The whole technical staff, many experiment would have been nightmares without them - Marina Kuhn, Roger Sauder, Roland Gut & Andreas Hefti-, the IT people from the BC2 (particularly Reiner and Konstantin). Thanks to Michaela Hanisch for so many thing that they would not fit in here. Special thanks to Scrapt (Dr. Gaël Sugano) and Dr. Mad (Dr. Amin M. Madoui) for having shared with me their experience in academics.

Last but not least, I am especially thankful to Prof. Dr. Guy R Cornelis and I always will feel in debt with him for the opportunity he gave to me. He always raised the level by considering me more as a collaborator than as a student. This had, and still has, as a consequence an extremely creative and bilateral collaboration.

11. *Curriculum vitae*

Pablo Manfredi

Private address:	**Work address:**
2E rue de Belfort	Klingelbergstrasse 70
68330 Huningue, France	CH-4056 Basel
Phone: +33 (0)6 72 01 56 82	Phone: +41 (0)61 267 21 27

E-mail: pablo.manfredi@unibas.ch

Date of Birth: July 23th 1980
Nationality: French
Birth place: Ithaca, New York state (USA)
Marital Status: married, 2 children

Languages

French	Native tongue
Spanish	Native tongue
English	Fluent (spoken and written)

Education & Experience

Biozentrum, Infectious Diseases, University of Basel, CH 03/11 – 08/12	**Post doctoral position** Group of Prof. Guy R. Cornelis Genome sequencing of clinical and environemental Capnocytophaga strains. Identification of a transferrins specific iron scavenging system in C. canimorsus exclusively present in clinical isolates.
Biozentrum, Infectious Diseases, University of Basel, CH 01/07 – 02/11	**PhD *summa cum laude* in Microbiology** Group of Prof. Guy R. Cornelis "*Capnocytophaga canimorsus*: Genomic characterization of a specialised host-dependent lifestyle and implications in pathogenesis" Analysis of host - pathogen interactions with a focus on the innate immune system using *in silico* genomic approach (genome sequencing, comparative genomics, phylogenetics), *in vitro* and *in vivo* infections, molecular biological, biochemical, and immunological techniques.
INPT – UPS III, Toulouse, France 09/03 – 09/06	**Master's degree** in "Genetics and Molecular Physiology of plants and associated microorganisms" Group of Dr Christian Boucher & Dr Stephane Genin. Majors: microbiology, plant physiology, cell biology, virology, parasitology, gene technology, and enzyme technology. Minors: organic chemistry and biochemistry.
ENSAT, Toulouse, France 09/03 – 09/06	**Master's degree** in food sciences and agricultural engineering, "Diplôme approfondi d'agronomie" specialisation in crop plant sciences
UPS III, Toulouse, France 09/01 – 09/03	**Bachelor's degree** in biology (DEUG SV)

Scientific Publications

2012	Manfredi P, Lauber F, Renzi F, Cornelis GR. **"Transferrin specific Iron acquisition in human serum by surfacer polysacharide Utilisation complexes in the pathogenic *Bacteroidetes Capnocytophaga canimorsus*."** In preparation.
2011	Ittig S, Lindner B, Stenta M, Manfredi P, Zdorovenko E, Knirel YA, Dal peraro M , Cornelis GR, Zähringer U. **"The Lipopolysaccharide from Capnocytophaga canimorsus Reveals an Unexpected Role of the Core-Oligosaccharide in MD-2 Binding."** accepted in PLoS Pathogens.
2011	Malone J, Jaeger T, Manfredi P, Doetsch A, Blanka A, Cornelis GR, Haeussler S, Jenal U. **"The YfiBNR signal transduction mechanism reveals novel targets for the evolution of persistent Pseudomonas aeruginosa in cystic fibrosis airways."** accepted in PLoS Pathogens.
2011	Manfredi P, Pagni M, Cornelis GR. “**Complete genome sequence of the dog commensal and human pathogen Capnocytophaga canimorsus strain 5.”** J Bacteriol. 2011 Oct;193(19).
2010	Manfredi P, Renzi F, Mally M, Sauteur L, Schmaler M, Moes S, Jenö P, Cornelis GR. “**The genome and surface proteome of *Capnocytophaga canimorsus* reveal a key role of glycan foraging systems in host glycoproteins deglycosylation.**” Mol Microbiol. 2011 Aug;81(4)
2010	Renzi F, Manfredi P, Mally M, Moes S, Jenö P, Cornelis GR. **“The N-glycan glycoprotein deglycosylation complex (Gpd) from *Capnocytophaga canimorsus* deglycosylates human IgG.”** PLoS Pathog. 2011 Jun;7(6)
2010	Plener L, Manfredi P, Valls M and Genin S. **“PrhG, a transcriptional regulator responding to growth conditions, is involved in the control of the type III secretion system regulon in Ralstonia solanacearum.”** J Bacteriol. 2010 Feb;192(4)

Skills

Molecular biology	Standard PCR and cloning methods, standard microbiology methods, standard cell culture methods, standard protein techniques (chromatography, electrophoresis (WB & stains), (co-)purifications, MS analysis…), genome sequencing methods, transcriptomics methods (RT or Q-PCR, microarrays, deep sequencing), and basics in optical microscopy.
Software	PC, Mac and linux OS, Microsoft Word/Excel/Power Point, CorelDraw, Adobe Acrobat Reader, Adobe Illustrator CS, EndNote and ImageJ (microscopy analysis programs).
Programming	Trained to Algorithmic and data base designing (Access, MySQL, XML). Sound notions in Perl, BioPerl and PHP. Basic level in VBA, java and C++.
Teaching	Supervision of master's degree students at the Biozentrum, University of Basel, CH. - Silvia Pietsch, (2010), currently Associate Scientist at Novartis - Loïc Sauteur (2010-2011), currently phd student at the Biozentrum, Basel (Werner Siemens fellowship). - Frédéric Lauber (2012), currently phd student at the Facultés universitaires Notre-Dame de la Paix, Belgium. Lecturer and supervisor during the practical "Block-course of microbiology" : bacterial resistance to the complement system. (2007-2011)

References

Prof. Dr. Guy R. Cornelis
Biozentrum, University of Basel
Klingelbergstrasse 70, CH-4056 Basel
Tel. secret. +41 61 267 21 21
guy.cornelis@unibas.ch

Dr. Marco Pagni
Swiss Institute of Bioinformatics, Vital-IT group
Quartier Sorge - Batiment Génopode, CH-1015 Lausanne Switzerland
Marco.Pagni@isb-sib.ch

Dr. DR2 Stéphane Génin
Laboratoire interactions plantes-microorganismes
UMR 2594, CNRS-INRA, 31326 Castanet Tolosan Cedex, France
Tel. +33 (0)5 61 28 5416 and +33 (0)5 61 28 5045
Stephane.Genin@toulouse.inra.fr

Basel, July 2012

Pablo Manfredi

12. Supplementary data

(see on the CD support)

Capnocytophaga canimorsus : Genomic characterization of a specialised host-dependent lifestyle and implications in pathogenesis

- **The complete genome of *Capnocytophaga canimorsus* 5 (*Cc5*), a bacterium causing fatal septicaemia in humans, draw the picture of an organism with a high degree of specialization to its natural environment : the canine oral cavity.**
- **Unexpectedly, *Cc5* does not encode any classical virulence complex. However it exhibits a very high number of surface-exposed lipoproteins mostly encoded within 13 putative polysaccharide utilization loci (PULs). Analysis of the *Cc5* surfome identified 73 surface exposed proteins among which lipoproteins accounted for 76% of the total quantification. Interestingly, 54% of total peptides detected were encoded in PULs. A systematic knockout analysis of the 13 PULs revealed that 6 PULs are involved in growth during cell culture infections with most dramatic effect observed for ΔPUL5.**
- **PUL5 turned out to be devoted to foraging glycans from N-linked glycoproteins as fetuin or IgG. It was not only essential for growth on cells but also for survival in mice and in human serum therefore representing a new type of virulence factor.**
- **Further characterization of the deglycosylation mechanism revealed that it involves a large surface complex spanning the outer membrane and consisting of 5 Gpd proteins. GpdDEF are surface-exposed outer membrane lipoproteins that contribute to the binding of glycoproteins at the bacterial surface while GpdG is a β-endo-glycosidase cleaving the N-linked oligosaccharide. In addition, GpdC resembles a TonB-dependent OM transporter that imports oligosaccharides into the periplasm. Finally, degradation of the oligosaccharide proceeds by the action of periplasmic exoglycosidases.**
- **Genome sequencing of additional human blood isolates of *canimorsus* have been performed with the only use of microreads methods. Two assembling approaches were developed in order to enhance assembly capacities of pre-existing tools. In addition, comparative genome analysis stressed features exclusively conserved among clinical isolates like oxidative stress resistance, the presence of an oxidative respiratory chain, or the conservation of a specific pattern of PUL genes. Therefore we propose these features as potential factors involved in the pathogenesis of *C. canimorsus*.**

BIOZENTRUM

Universität Basel
The Center for Molecular Life Sciences

© 2012 Pablo Manfredi