FUNCTIONAL GENOMICS of MOTILE COMMENSAL INTESTINAL BACTERIA

Candidate: B. Anne Neville (BSc)

A thesis presented in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy.

Department of Microbiology,

National University of Ireland,

Cork,

Ireland.

January 2013

Supervisors: Prof. Paul W. O'Toole & Prof. R. Paul Ross

Examiners: Dr. Anne McCartney (External) & Prof. Douwe van Sinderen (Internal)

Head of Dept.: Prof. Gerald Fitzgerald

Declaration

I hereby declare that the content of this thesis is the result of my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

_____

B. Anne Neville

This thesis is dedicated to my family.

General Table of Contents

Table of Figures

Table of Tables

Abbreviations

AA = Amino Acid

ACT = Artemis Comparison Tool

ACTB = β-actin

ANI = Average Nucleotide Identity

BCA = Bicinchoninic Acid

BLAST = Basic Local Alignment Search Tool

BSH = Bile Salt Hydrolase

CAG = Chronic Atrophic Gastritis

CAMP = Cationic Antimicrobial Peptide

CARD = Caspase Activation and Recruitment Domain

cDNA = Complementary DNA

CDS = Coding DNA Sequence

COG = Clusters of Orthologous Groups

CFU = Colony Forming Units

CM = Conditioned Media/Medium

DC = Dendritic Cells

DDH = DNA-DNA Hybridisation

DGGE = Denaturing Gel Gradient Electrophoresis

DMEM = Dulbecco's Modified Eagle Medium

drTLR5 = *Danio rerio* TLR5

DSS = Dextran Sulphate Sodium

ECF = Extracytoplasmic Function

EDTA = Ethylene-diamine-tetraacetic acid

EFSA = European Food Safety Authority

EHEC = Enterohaemorrhagic *E. coli*

ELISA = Enzyme Linked Immunosorbent Assay

EM = Electron Microscopy

EPS = Exopolysaccharide

EST = Expressed Sequence Tag

FAO/WHO = Food and Agriculture Organisation of the United Nations and the World Health Organisation

FBS = Foetal Bovine Serum

FOP = Flagellar Operon Protein

GABA = Gamma Amino Butyric Acid

GALT = Gut Associated Lymphoid Tissue

GAPDH = Glyceraldehyde 3 Phosphate Dehydrogenase

GI = Gastrointestinal

GRAS = Generally Regarded As Safe

GST = Glutathione S Transferase

GTP = Guanosine Nucleoside Triphosphate

HAP = Hook Associated Protein

HGT = Horizontal Gene Transfer

HMP = Human Microbiome Project

hTLR5 = Human Toll Like Receptor 5

IEC = Intestinal Epithelial Cells

IFN-γ = Interferon-γ

IL = Interleukin

IL10-KO = Interleukin-10 Knock-out

IκB = Inhibitor of NF-κB

iNOS = Inducible Nitric Oxide Synthase

IPAF = ICE-Protease Activating Factor

IPTG = Isopropylthio-β-Galactoside

IRAK = Interleukin 1 Receptor Associated Kinase

JNK = c-Jun Kinase

KEGG = Kyoto Encyclopedia of Genes and Genomes

LAB = Lactic Acid Bacteria

LB = Lysogeny Broth

LPS = Lipopolysaccharide

LRR = Leucine Rich Repeat

MAMP = Microbe Associated Molecular Pattern

MAPK = Mitogen Activated Protein Kinase

MBL = MreB Like

MCP = Methyl accepting chemotaxis protein

MKK = Mitogen-activated Protein Kinase Kinase

MLN = Mesenteric Lymph Nodes

MLNC = Mesenteric Lymph Node Cells

MOI = Multiplicity of Infection

MRS = de Mann, Rogosa, Sharpe

MSD = Meso Scale Discovery

mTLR5 = Murine Toll Like Receptor 5

NAG = N-acetyl glucosamine

NAIP = NLR family, apoptosis inhibitory proteins

NAM = N-acetyl muramic acid

NCBI = National Center for Biotechnology Information

NF- κB = Nuclear Factor κ B

NGS = Next-generation Sequencing

NLR = Nucleotide-binding domain and Leucine-rich repeat containing Receptors

NLRC = NLR family, CARD domain containing

NR = Non-redundant

NT = Nucleotide

ORF = Open Reading Frame

ORI = Origin

OTU = Operational Taxonomic Unit

PAMP = Pathogen Associated Molecular Pattern

PAMP-pm = PAMP-post-mortem

PAMP-pv = PAMP-pro-vita

PB = Peripheral Blood

PBMC = Peripheral Blood Mononuclear Cells

PBS = Phosphate Buffered Saline

PCR = Polymerase Chain Reaction

PGAAP = Prokaryotic Genomes Automatic Annotation Pipeline

ppGpp = Guanosine tetraphosphate

PRR = Pattern Recognition Receptor

PVDF = Polyvinylidene fluoride

QPS = Qualified Presumption of Safety

qRT-PCR = Quantitative Real-Time PCR

QTL = Quantitative Trait Locus

RACE = Rapid Amplification of cDNA Ends

RBS = Ribosome Binding Site

RFLP = Restriction Fragment Length Polymorphism

rRNA = Ribosomal RNA

SCFA = Short Chain Fatty Acid

SDP = Sortase Dependent Surface Proteins

SDS-PAGE = Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis

siRNA = Short Interfering Ribonucleic Acid

S. typhimurium = *Salmonella enterica* serovar Typhimurium

T3SS = Type 3 Secretion System

TAB = TAK-1 Binding Protein

TAK = TGF-β Associated Kinase

TBS = Tris Buffered Saline

TCDA = Taurochenodeoxycholic Acid

TEM = Transmission Electron Microscopy

TER = Terminus

TGF-β = Transforming Growth Factor-β

T_{regs} = Regulatory T cells

TNF-α =Tumour Necrosis Factor -α

TIR = Toll/Interleukin-1

TLR = Toll Like Receptor

TRAF = Tumour Necrosis Factor Receptor Associated Factor

T-RFLP = Terminal Restriction Fragment Length Polymorphism

TRIF = TIR domain containing adaptor-inducing interferon-β

VISA = Vancomycin Intermediate Resistant *Staphylococcus aureus*

Abstract

Flagella confer upon bacteria the ability to move and are therefore organelles of significant bacteriological importance. The innate immune system has evolved to recognise flagellin, (the major protein component of the bacterial flagellar filament). Flagellate microbes can potentially stimulate the immune systems of mammals, and thus have significant immunomodulatory potential.

The flagellum-biogenesis genotype and phenotype of *Lactobacillus ruminis*, an autochthonous intestinal commensal, was studied. The flagellum-biogenesis genotypes of motile enteric *Eubacterium* and *Roseburia* species were also investigated. Flagellin proteins were recovered from these commensal species, their amino-termini were sequenced and the proteins were found to be pro-inflammatory, as assessed by measurement of interleukin-8 (IL-8) secretion from human intestinal epithelial cell lines. For *L. ruminis*, this IL-8 secretion required signalling through Toll Like Receptor 5.

A model for the regulation of flagellum-biogenesis in *L. ruminis* was inferred from transcriptomics data and bioinformatics analyses. Motility gene expression in this species may be under the control of a novel regulator, LRC_15730. Potential promoters for genes encoding flagellin proteins in the *Eubacterium* and *Roseburia* genomes analysed were inferred *in silico*.

Relative abundances of the target *Eubacterium* and *Roseburia* species in the intestinal microbiota of 25 elderly individuals were determined. These species were found to be variably abundant in these individuals. Motility genes from these species were variably detected in the shotgun metagenome databases generated by the ELDERMET project. This suggested that a greater depth of sequencing, or improved evenness of sequencing, would be required to capture

the full diversity of microbial functions for specific target or low abundance species in microbial communities by metagenomics.

In summary, this thesis used a functional genomics approach to describe flagellum-mediated motility in selected Gram-positive commensal bacteria. The regulation of flagellum biosynthesis in these species, and the consequences of flagella expression from a host-interaction perspective were also considered.

Chapter 1

General Introduction

1.1 The *Lactobacillus* component of the mammalian gastrointestinal tract – composition and defining influences.

1.1.1 General introduction to the gastrointestinal microbiota of humans.

The ubiquitous nature of microorganisms means that they are intimately associated with all multi-cellular life forms. Microbes are naturally present at the environmentally exposed body sites of humans and other animals, such as on the skin and in the respiratory, urogenital and gastrointestinal (GI) tracts (Huttenhower *et al.*, 2012). The human body therefore bears a significant microbiological load which is known as its microbiota.

1.1.2 Diversity, variability and composition of the intestinal microbiota.

Bacteria are the numerically dominant microorganisms present in the human intestine (Arumugam *et al.*, 2011) and a large proportion of these are considered as novel species (Eckburg *et al.*, 2005). An *in silico* analysis of gut-derived 16S rRNA gene sequences estimated that the bacterial component of the human gut microbiota represents approximately 800 different species or more than 7000 different bacterial strains (Backhed *et al.*, 2005). Accordingly, significant biodiversity and inter-individual variation in the composition of the gut microbiota is perceived at the level of species or strain.

The intestinal microbiotas of healthy vertebrates are dominated by species of the phyla *Firmicutes* and *Bacteroidetes* (Arumugam *et al.*, 2011; Claesson *et al.*, 2011; Ley *et al.*, 2008). The phylum *Firmicutes* is subdivided into several classes and many genera, and includes the genus *Lactobacillus*. Most species of the *Firmicutes* lineage that are found in the intestinal microbiota belong to the class *Clostridia* (Eckburg *et al.*, 2005), which may be divided into 19 clusters on the basis of phenotypic and molecular analyses, and phylogenetic relatedness (Collins *et al.*,

1994). Species of *Clostridium* clusters IV and XIVa, are well represented in the adult human gut (Claesson *et al.,* 2011). Although lactobacilli have been frequently isolated from mammalian intestines, they are usually minority or subdominant species in this niche (Rinttila *et al.*, 2004; Tannock *et al.*, 2000). The phylum *Bacteroidetes* is also dominant in the intestinal microbiota of humans (Arumugam *et al.,* 2011; Claesson *et al.,* 2011). The less abundant phyla of the intestinal microbiota include the *Proteobacteria, Actinobacteria, Fusobacteria,* and *Verrucomicrobia* (Eckburg *et al*., 2005).

1.1.2.1 *The* Lactobacillus *component of the commensal human GI microbiota.*

The *Lactobacillus* component of the human GI tract microbiota forms in early life and may temporarily dominate the overall microbiota of vaginally delivered neonates (Dominguez-Bello *et al.*, 2010). This is a direct reflection of the dominance of *Lactobacillus* species in the vaginal microbiota of healthy females, which is partially transferred to offspring during birth (Dominguez-Bello *et al.,* 2010). Although the composition of the GI microbiota is in a state of flux for the first few months of an infant's life, the lactobacilli are typically in the minority (Adlerberth & Wold, 2009; Ahrne *et al.,* 2005), a condition that is preserved through to adulthood (Rinttila *et al.*, 2004).

1.1.2.1.1 *Lactobacilli in the oral cavity.*

Lactobacillus species occur in low numbers throughout the human GI tract (Reuter, 2001). In the mouth, lactobacilli are found on the tongue, the teeth, the cheeks, the hard palate and in the saliva (Zaura *et al.,* 2009). Although several *Lactobacillus* species may be found in the oral cavity (Dewhirst *et al.,* 2010; Maukonen *et al.,* 2008), the *Lactobacillus* community is not usually dominant in this niche (Zaura *et al.*, 2009).

The salivary *Lactobacillus* load has been estimated at 10^4 - 10^5 colony forming units (CFU) / ml according to culture dependent techniques (Maukonen *et al.*, 2008; van Houte & Green, 1974) and at ~ 1.45 % of the oral microbiota by culture independent approaches (Dewhirst *et al.*, 2010). *Lactobacillus gasseri*, *Lactobacillus fermentum*, *Lactobacillus paracasei*, *Lactobacillus rhamnosus* and *Lactobacillus vaginalis* are the most frequently identified and most abundant *Lactobacillus* species in the mouth (Badet & Thebaud, 2008; Byun *et al.*, 2004; Dal Bello & Hertel, 2006; Maukonen *et al.*, 2008), and lactobacilli are believed to be associated with the progression of cavities (Byun *et al.*, 2004; Gross *et al.*, 2010).

The oral *Lactobacillus* community may include a number of species that do not persist in this niche but rather, were introduced with food, rendering them "opportunistic and exogenous" bacteria (Caufield *et al.*, 2007). Indeed, the salivary and faecal *Lactobacillus* communities of healthy adult humans have many strains in common (Dal Bello & Hertel, 2006; Maukonen *et al.*, 2008), suggesting that many "intestinal" lactobacilli are allochthonous or transient residents of the intestines (Dal Bello & Hertel, 2006; Walter, 2008).

1.1.2.1.2 *Lactobacilli in the stomach.*

The bacterial population of the human stomach has been estimated at ~ 10^3 CFU / ml of gastric juice (FAO/WHO, 2001). The harsh, acidic conditions of the stomach largely prevent its colonisation by unspecialised microbes. Nevertheless, on the basis of 16S rRNA gene analysis of gastric biopsies, the human stomach was found to harbour a diverse bacterial population in which eight different phyla were represented (Bik *et al.*, 2006). *L. fermentum* and *L. delbruckeii* were the only *Lactobacillus* species identified, and were present in low numbers in only two of twenty-three subjects (Bik *et al.*, 2006). However, the potential *Lactobacillus*

population of the human stomach is considerably more diverse. Four novel *Lactobacillus* species, *Lactobacillus gastricus, Lactobacillus antri, Lactobacillus kalixensis* and *Lactobacillus ultunensis* were first isolated from the human stomach (Roos *et al.*, 2005) and several more species, including *Lactobacillus casei, L. fermentum, L. gasseri, L. vaginalis, L. reuteri* and *L. salivarius* have also been recovered from this niche (Hakalehto *et al.*, 2011; Reuter, 2001; Ryan *et al.*, 2008b) (Figure 1.1).

1.1.2.1.3 *Lactobacilli in the small intestines.*

A meta-analysis by Reuter (Reuter, 2001) described a sizable and diverse *Lactobacillus* community in the human jejunum and ileum. Samples were taken either by biopsy post-mortem or by inviting adults and children to swallow a capsule that would automatically open and close during GI transit to trap the intestinal contents. According to microbiological, culture-based analyses, sampled cadavers harboured a sizable (median = $\log_{10}$ 7.7 CFU / g) *Lactobacillus* community, and *L. gasseri* and *L. reuteri* were the prevalent *Lactobacillus* species present. *L. salivarius* occurred less frequently, being identified in only three of the seven cadavers. The *L. salivarius* population was however, similar in size to the *L. gasseri* and *L. reuteri* populations of these individuals (Reuter, 2001).

Lactobacilli were the most frequently detected microorganisms in the ilea of 10 children that were repeatedly sampled using the capsule method (n = 38). *L. ruminis* was regularly present, being detected in 21 of the 38 samples. Lactobacilli were also recovered from the small intestines of adults who swallowed the capsule. *L. gasseri, L. reuteri, L. salivarius* and *L. ruminis* were the *Lactobacillus* species most frequently isolated and were believed to be autochthonous to this niche. *L. casei, L.*

buchneri, *L. fermentum* and *L. plantarum* were recovered from the jejuna and ilea of these adults, but were considered allochthonous to the GI tract (Reuter, 2001).

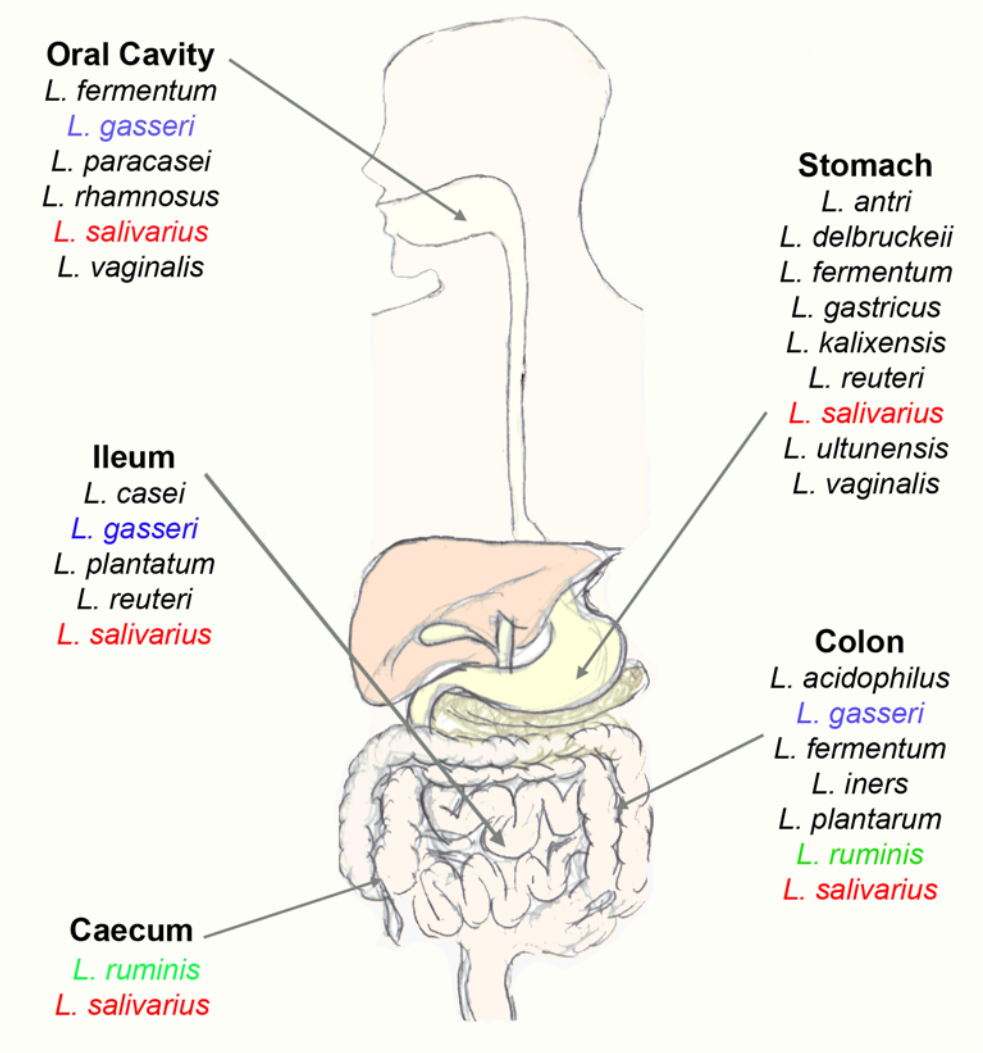


Figure 1.1: Anatomy and *Lactobacillus* communities of the human gastrointestinal tract. The names of species believed to be autochthonous to this niche are coloured blue, green and red. Compiled from (Badet & Thebaud, 2008; Dal Bello & Hertel, 2006; Hakalehto *et al.*, 2011; Maukonen *et al.*, 2008; Roos *et al.*, 2005; Ryan *et al.*, 2008b).

.

However, separate culture-independent studies of the small intestinal microbiota of elderly cadavers and healthy adult humans described a low diversity *Lactobacillus* community in the human jejunum and ileum (Booijink *et al.*, 2010; Hayashi *et al.*, 2005). *Lactobacillus* operational taxonomic units (OTUs) were not detected in the jejunal and ileal clone libraries for two of the three cadavers (Hayashi *et al.*, 2005). One *L. reuteri* subgroup OTU and one *L. mali* (*L. salivarius* clade) subgroup OTU were found in the third cadaver, the latter being the dominant *Lactobacillus* OTU. The prevalence of *L. mali*-like sequences in the jejunum and ileum of this individual was confirmed by Terminal Restriction Fragment Length Polymorphism (T-RFLP) (Hayashi *et al.*, 2005).

A study of ileostomy effluent from asymptomatic human subjects and from individuals with Crohn's disease, found that lactobacilli can achieve reasonably high numbers in the ileum, ranging from $< 10^4$ to $\sim 10^8$ copies of the 16S rRNA gene per gramme of ileal effluent (Booijink *et al.*, 2010). Indeed, the lactobacilli typically represented 0.16 - 0.9 % and 0.04 – 0.14 % of the microbiota of asymptomatic individuals with an ileostomy stoma and ileostomy patients taking medication, respectively. One asymptomatic male ileostomy patient harboured a *Lactobacillus* population that fluctuated significantly from day to day over a ten day period, representing on average 10.9 % (quantitative real-time PCR) or ~ 13 % (HITChip) of the microbiota. On one sampling occasion, *Lactobacillus* sequences constituted 43 % (quantitative real-time PCR) or 47 % (HITChip) of the total bacterial microbiota of this individual. This study therefore showed that individuals harbour a personal ileal microbiota whose community composition is quite unstable from day to day (Booijink *et al.*, 2010).

The lactobacilli represented a high proportion of the effluent and tissue-adherent bacterial populations of 17 patients with an ileostomy stoma (Hartman *et al.*, 2009). The dominance of the lactobacilli was usurped by populations of strict anaerobes, specifically the *Clostridia* and *Bacteroides* following surgical closure of these ileostomy stomata (Hartman *et al.*, 2009), which implied that oxygen exposure may have been responsible for the high *Lactobacillus* numbers.

1.1.2.1.4 *Lactobacilli in the caecum.*

The caecum is found between the terminal ileum and the ascending colon (Gray, 1918). It is separated from the ileum by the ileo-caecal valve (Gray, 1918) and functions as a repository for the liquid chyme before it proceeds through the colon.

Marteau and co-workers found that the caecal microbiota was quantitatively and qualitatively different from the faecal microbiota in human subjects (Marteau *et al.*, 2001). Specifically, the anaerobic bacterial population was smaller (mean = 10^8 CFU / ml) in the caecum than in the faeces (mean = $10^{10.4}$ CFU / ml) of eight healthy people, but no significant difference was recorded for the numbers of facultative anaerobes found in the caecum (mean = $10^{7.4}$ CFU / ml) or faeces (mean = $10^{7.8}$ CFU / ml). Therefore, facultative anaerobic bacteria, which included the lactobacilli, represented a larger proportion of the total anaerobic microbiota of the caecum (25 %) than of the faeces (1 %) in these individuals (Marteau *et al.*, 2001). Molecular hybridisation with a *Lactobacillus-Enterococcus* 16S rRNA gene specific probe showed that the lactobacilli and enterococci together represented 22.8 % of bacterial RNA in the caecum versus 6.6 % in the faeces of these people (Marteau *et al.*, 2001).

Hayashi and colleagues similarly reported that the composition of the caecal microbiota was markedly different from the microbiota associated with the human jejunum, ileum and recto-sigmoidal colon (Hayashi *et al.*, 2005). Lactobacilli of the

L. mali subgroup were detected as one of the major components of the cecal microbiota of one individual by T-RFLP analysis (Hayashi *et al.*, 2005), even though this subgroup was represented by only 7 clones (1 OTU) in a 16S rRNA gene library containing 92 sequences (22 OTUs) from the same site. Heilig and co-workers also found that species of the *L. salivarius* clade dominated the cecal *Lactobacillus* population, with *L. ruminis*- like amplicons being especially abundant in caecal chyme (Heilig *et al.*, 2002).

1.1.2.1.5 *Lactobacilli in the colon.*

The *Lactobacillus* population of the colon represents less than 1 % of the colonic microbiota (Lebeer *et al.*, 2008), but its composition varies from site to site. The *Lactobacillus* population of the ascending colon was significantly smaller (mean = $10^{4.5}$ copies 16S rRNA gene / mg mucosal tissue) than that of the terminal ileum (mean = $10^{5.2}$ copies 16S rRNA gene / mg mucosal tissue) or the descending colon (mean = 10^{6} copies 16S rRNA gene / mg mucosal tissue) (Ahmed *et al.*, 2007).

Biopsies from various sites in the colons of ten adult humans were used to evaluate the diversity of the mucosa-adherent microbial community (Zoetendal *et al.*, 2002). The bacterial communities associated with biopsies from the ascending colon tended to be numerically smaller than the bacterial communities associated with biopsies from other areas, but this difference was not statistically significant. This observation, that there is no significant difference between the total bacterial numbers from various sites in the colon, is in accordance with the findings of Ahmed and colleagues (Ahmed *et al.*, 2007).

According to denaturing gel gradient electrophoresis (DGGE) and 16S rRNA gene analyses, the mucosa-adherent microbial community was host-specific and relatively uniform throughout the colon (Zoetendal *et al.*, 2002), but significantly

different from the faecal microbiota of the same individual. The *Lactobacillus* component of the adherent bacterial community was similar in biopsy and faecal samples for six of the ten individuals tested. *L. gasseri* was the most common *Lactobacillus* species detected and was present in nine of the ten people. Some minor intra-individual variation in the mucosa-adherent *Lactobacillus* populations of biopsies from different colonic locations was recorded for three of the ten individuals.

In another study, biopsies from the ascending, transverse and descending colon of four healthy individuals with a family history of polyps, revealed a complex bacterial community that was host specific and which did not vary with sampling site in the colon (Nielsen *et al.*, 2003). However, similar to Zoetendal and colleagues (Zoetendal *et al.*, 2002), the adherent *Lactobacillus*-like community varied between individuals and also between sampling sites (Nielsen *et al.*, 2003). Sequencing revealed that *L. fermentum*-, *L. iners*- and *L. ruminis*-like sequences were present in the colonic biopsies. Traditional plating methods also recovered *L. ruminis* and species of the *L. casei* clade from the biopsy samples (Nielsen *et al.*, 2003). However, *Leuconostoc* and *Weissella* species were also significant components of the mucosa-associated lactic acid bacterial population in the sampled individuals (Nielsen *et al.*, 2003).

A later study of healthy colonic tissue from 26 humans undergoing emergency colonic resection surgery also found that the microbiota of the colon was complex and that intra-individual variation was low, indicating broad conservation of the species present at each of the sampling sites (Ahmed *et al.*, 2007). Inter-individual differences in the microbiota of the ascending colon was also low in this cohort. Lactobacilli were more prominent in distal than in the proximal colon, although they were detected throughout the colon. *L. plantarum*-like and *L. rhamnosus* sequences

were recovered from ascending and transverse colon samples respectively, while *L. acidophilus*, *L. fermentum* and *L. salivarius* were identified in the descending colon (Ahmed *et al.*, 2007).

Biopsies derived from the major regions of the colons of three adults revealed that intra-individual mucosa-derived diversity profiles were similar (Eckburg *et al.*, 2005). Significant differences were found between the mucosa-adherent and the faecal microbiotas of two of the three individuals examined (Eckburg *et al.*, 2005). Much inter-individual difference in the mucosal and faecal microbiota was also recorded.

1.1.2.1.6 *The faecal microbiota of healthy humans contains autochthonous and allochthonous lactobacilli.*

The faecal microbiota is representative of the microbial community of the colon and as such, its composition is influenced by the same factors that define and shape the intestinal microbial community. Human diets often include foods fermented by or supplemented with *Lactobacillus* species and strains. Thus, dietary *Lactobacillus* consumption impacts on the size and structure of the *Lactobacillus* component of the intestines, and also of the faecal microbiota. Autochthonous species are those that colonise and replicate in the GI tract, while allochthonous species in contrast, are merely passing through (Tannock *et al.*, 2000).

The importance of the distinction between the allochthonous and autochthonous *Lactobacillus* species of the human GI tract became apparent in a study that aimed to describe the effects of prolonged probiotic consumption on the composition of the faecal microbiota of ten humans (Tannock *et al.*, 2000). Participating individuals provided faecal samples once monthly and the species-level microbial component was investigated. The study included a six month pre-test

control period, a six month test period during which a milk product containing *L. rhamnosus* DR20 was consumed daily, followed by a three month post-test recovery and monitoring phase. Lactobacilli were recovered from the faeces of each person during each of the test periods, but the diversity of the faecal *Lactobacillus* population present differed from person to person. *L. rhamnosus* DR20 was the dominant *Lactobacillus* species in the faeces of only six of the ten subjects during the test period. However, this strain was no longer detected in large numbers in their faeces once probiotic consumption had ceased. A different set of lactobacilli dominated during the pre-test and post-test faecal samples of these six individuals. For example, previously absent species such as *L. acidophilus,* were detected among the predominant bacteria in their faeces during the post-test period (Tannock *et al.*, 2000).

The faecal *Lactobacillus* profiles of the four subjects who did not harbour a large *L. rhamnosus* community during the test-period remained unchanged throughout the study. These individuals harboured stable *Lactobacillus* populations that were relatively simple, being dominated by only one or two species, which often included *L. ruminis* and *L. salivarius*. These species were thus considered autochthonous to the GI tract of at least four of these subjects. The failure of the administered *L. rhamnosus* strain to dominate the faecal *Lactobacillus* communities of these individuals suggested that the autochthonous *Lactobacillus* species prevented the establishment of the allochthonous species in the intestinal niche (Tannock *et al.*, 2000).

Long term probiotic consumption therefore influenced the composition of the *Lactobacillus* component of the faecal microbiota in individuals that lacked a stable community of autochthonous lactobacilli. Probiotic consumption did however

influence the size of the faecal *Lactobacillus* community of all individuals, which was larger during the test period (Tannock *et al.*, 2000).

In addition to *L. ruminis* and *L. salivarius*, several other species were also present in the faecal *Lactobacillus* communities of healthy individuals (Figure 1.1). These included *Lactobacillus brevis, Lactobacillus crispatus, L. acidophilus, L. gasseri* and species of the *L. casei* and *L. plantarum* groups (Tannock *et al.*, 2000). Some of these lactobacilli are food-associated microorganisms and their presence in the faecal microbiota may indicate that the host had consumed foods containing such species. If these species are allochthonous or transient members of the GI tract, it would be expected that they would become undetectable in faeces once consumption of the bacterium stops. For example, *L. acidophilus* NCFM failed to colonise the GI tract of healthy adults following a two week feeding trial, despite becoming the predominant species in faeces during the test period (Sui *et al.*, 2002). In another human feeding trial, a probiotic *L. casei* strain survived well in the GI tract and was detected in faeces of healthy adults up to one week after its last consumption, suggesting that this particular strain may multiply in the intestinal niche (Tuohy *et al.*, 2006). Oral administration of *L. paracasei* to young children twice a day for three weeks did not result in prolonged detection of the probiotic strain in all individuals after the trial (Heilig *et al.*, 2002).

1.1.3 The physiology of the GI tract presents a survival challenge that shapes the composition of the microbiota.

The normal physiological functions of the host GI tract impose significant selective pressures on the intestinal microbiota and, from a microbial perspective, the intestinal niche poses many survival challenges.

1.1.3.1 *Intrinsic host factors that influence the composition of the* **Lactobacillus** *GI microbiota.*

Lactobacilli that occupy the oral cavity must adapt to certain environmental stresses that intestinal *Lactobacillus* species are not usually subjected to, such as oxygen exposure, sharp fluctuations in local pH and temperature during meal-times and recurring exposures to anti-microbials as part of routine oral hygiene practices.

To colonise the intestines, a bacterium must first withstand challenges to its structural integrity presented by digestive enzymes and acid in the mouth and the stomach. The bacteria that enter the intestines may therefore be sublethally damaged. Mechanical and chemical forces, such as the peristaltic contractions of the intestines and the secretion of anti-microbial substances by the host and the microbiota alike, may further hinder the ability of a bacterium to colonise this niche.

1.1.3.1.1 *Influence of pH.*

The pH in the mouth is in the range of 6.75 - 7.25 (Marsh, 2003), and the lactobacilli are a subdominant component of the oral microbiota (Dewhirst *et al.*, 2010). Dental plaque is a multi-species biofilm that forms on teeth. However, the species composition of dental plaque may be influenced by the prevailing conditions (Marsh, 2003), such as such as pH, oxygen and nutrient availability. This is summarised in the "ecological plaque hypothesis" which is founded on the assumption that microbial homeostasis is established in the oral cavity and that any sustained alteration to the prevailing conditions could favour a cariogenic oral environment. Changes to the oral environment would select for species that are more tolerant of the new conditions (Marsh, 1994). A major ecological pressure such as an enrichment of fermentable sugars in the host diet, may favour an acidic oral

environment and a dental plaque that is dominated by aciduric and acidogenic *Lactobacillus* and *Streptococcus* species (Marsh, 1994).

Gastric acid (pH = 1 - 2) is a major colonisation barrier that restricts outgrowth of microorganisms in both the upper and lower GI tracts. Lactobacilli were more numerous in the faecal microbiotas of persons with hypochlorhydria, (a condition defined by reduced hydrochloric acid secretion in the stomach), when compared to asymptomatic individuals without chronic atrophic gastritis (CAG) (Kanno *et al.*, 2009).

Importantly, *Lactobacillus* strains intended for use as probiotics or as drug and vaccine delivery agents, must be able to withstand gastric transit (FAO/WHO, 2001). During colonisation of the murine stomach, *L. reuteri* induces expression of acid-tolerance genes, including glutamate decarboxylase and urease (Wilson, 2011; Wilson *et al.*, 2012). Presumably, the aciduric *Lactobacillus* species that colonise the human stomach have a similar complement of acid-tolerance genes.

The pH of the intestinal lumen is influenced by gastric acid, pancreatic secretions, microbial fermentations and host diet. The small intestine and colon have mildly acidic pHs (pH ~ 5 - 7) which increase distally (Bown *et al.*, 1974; Ridlon *et al.*, 2006). The unequal availability of fermentable carbohydrate throughout the colon is probably partly responsible for the proximal to distal pH gradient (Macfarlane *et al.*, 1992). Specifically, carbohydrate and protein catabolism by enteric microbes tends to reduce the luminal pH by the production of acidic metabolic by-products such as short-chain fatty acids (SCFAs), phenols, indoles, thiols, hydrogen sulphide, carbon dioxide and hydrogen gas (Macfarlane *et al.*, 1992; Macfarlane & Macfarlane, 2012). Protein fermentation additionally yields molecules which have basic properties, such as ammonia (Macfarlane *et al.*, 1992). A combination of pH levels

and peptide availability influence the composition of the intestinal microbiota and also the intestinal SCFA profile produced (Walker *et al.*, 2005). Experiments performed in a continuous flow fermentor showed that more butyrate was present at a maintained fermentor pH of 5.5 than at pH 6.5, under conditions of both high and low peptide availability. This coincided with expansion of the butyrogenic bacterial population at the lower pH. Propionate formation was greatest under high peptide (0.6 %) conditions at pH 6.5 (Walker *et al.*, 2005). Thus, it may be inferred that in the colon, the local pH, peptide availability and utilisation of bacterial-derived metabolites cumulatively influence the composition of the microbiota.

1.1.3.1.2 *Influence of digestive enzymes and bile.*

The duodenum receives enzymatic secretions from the pancreas and biliary secretions from the gall bladder as part of the normal digestive process. Bile acids have a dual purpose; they are important for the emulsification of dietary lipids, but also as anti-microbials (Begley *et al.*, 2005). Thus, bile acids have a significant influence on the composition of the intestinal microbiota and also play a role in enteroprotection by stimulating the host's anti-microbial defences (Inagaki *et al.*, 2006).

When bile acids are synthesised from cholesterol in the liver they are usually conjugated to either taurine or glycine, which increases their solubility (Hofmann, 1994). These conjugated bile acids are released into the duodenum from the gall bladder following a meal to assist the digestion and absorption of dietary fats and fat soluble vitamins. Bile acids are modified by the actions of the resident microbiota in the intestines, yielding unconjugated and secondary bile acids (Hofmann, 1994; Ridlon *et al.*, 2006). Bile acids and their derivatives have significant anti-microbial properties (Begley *et al.*, 2005). Membrane damage caused by the affinity of the

inherently hydrophobic bile acids for the phospholipid bilayers of bacteria is fundamental to the bactericidal effect of bile (Kurdi *et al.*, 2006). Bile acids are efficiently reabsorbed in the ileum, so bile concentrations are greatest in the proximal small intestine and decrease distally (Hofmann, 1994).

The *Lactobacillus* bile-resistance phenotype is conferred by a number of genes and proteins which respond to different aspects of bile challenge (Hamon *et al.*, 2011; Hamon *et al.*, 2012). Those with a role in modifying the cell surface or in maintaining the structural integrity of the cell wall and membrane were found to contribute to the bile tolerance phenotype of *L. casei* and *L. plantarum* strains (Hamon *et al.*, 2011; Hamon *et al.*, 2012). In *L. delbrueckii* subsp. *lactis* and *L. casei*, some central aspects of cellular function such as translation and carbohydrate metabolism were altered in response to bile (Burns *et al.*, 2010; Hamon *et al.*, 2012). Bile induced *L. plantarum* strains to produce glutathione reductases, presumably to cope with bile-induced oxidative damage (Hamon *et al.*, 2011). In *L. casei* and *L. delbrueckii* subsp. *lactis*, proteins with a role in cell protection, such as the molecular chaperone-protease ClpL, contributed to the early stages of the bile-resistance phenotype (Burns *et al.*, 2010; Hamon *et al.*, 2012). *L. plantarum* expressed an Opu ABC transporter to overcome the osmotic stresses imposed on the bacterium by bile (Hamon *et al.*, 2011), and expression of an F_0F_1-ATPase (AtpH) may have contributed to the resistance phenotype by maintaining intracellular pH (Hamon *et al.*, 2011). Bile salt hydrolases are enzymes with a role in the detoxification of bile acids (Begley *et al.*, 2006). While lactobacilli that are intended for use as probiotics are routinely screened for the presence of bile salt hydrolases (BSH) (FAO/WHO, 2001), these enzymes may not play a central role in the bile-resistance phenotype of

lactobacilli (Begley *et al.*, 2005; Fang *et al.*, 2009; Hamon *et al.*, 2011; Hamon *et al.*, 2012; Moser & Savage, 2001).

1.1.3.1.3 *Influence of oxygen availability.*

The concentration of oxygen in the gut decreases distally (He *et al.*, 1999), and oxygen tension is particularly low in the rectal lumen (Lind Due *et al.*, 2003). The intestine is essentially an anaerobic environment (Hartman *et al.*, 2009) and any swallowed oxygen that arrives to the intestine is rapidly consumed by facultative anaerobes. Lactobacilli can be detected in the microbiota of vaginally delivered neonates 48 hours after birth (Karlsson *et al.*, 2011) and are continually represented as part of the intestinal microbiota of infants through weaning and during the first 18 months of life (Ahrne *et al.*, 2005). This suggests that the enteric *Lactobacillus* community is tolerant of aerobic and anaerobic conditions.

1.1.3.1.4 *Influence of mucus and cationic anti-microbial compounds.*

Mucus is a viscous fluid that lines the GI tract and provides a physical barrier between the luminal contents and the underlying epithelium and also acts as a non-specific immune defence against invading bacteria. The mucus lining that covers the GI epithelium consists of a thick outer layer which is in direct contact with the intestinal microbiota, and a thinner inner layer that is effectively devoid of microbes (McGuckin *et al.*, 2011; Swidsinski *et al.*, 2007). The outer mucus layer is thickest at sites of potential injury, such as in the stomach and colon (Atuma *et al.*, 2001). However, the outer mucus layer is constantly sloughed off by the movement of undigested food boluses and faecal material through the GI tract, and also by the digestive activities of the ~ 1 % of the resident microbiota that can use mucus components as a substrate (Derrien *et al.*, 2004; Hoskins *et al.*, 1985; Ruas-Madiedo *et al.*, 2008).

Lactobacilli do not usually degrade mucins (Ruseler-van Embden *et al.*, 1995; Turpin *et al.*, 2012; Zhou *et al.*, 2001). Rather, lactobacilli such as *L. rhamnosus* (Kankainen *et al.*, 2009) and *L. fermentum* (Macias-Rodriguez *et al.*, 2009) have evolved mechanisms to adhere to components of colonic mucus. This may represent a colonisation factor that helps these bacteria to persist in the intestines.

Cationic anti-microbial compounds (CAMPs) are peptides (Peschel & Sahl, 2006) found in mucus that influence the composition of the intestinal microbiota. The lactobacilli may either stimulate CAMP secretion or they may be sensitive to CAMP molecules. For example, *Lactobacillus rhamnosus* GG is sensitive to human beta-defensin-2, but not to beta-defensin-1 (De Keersmaecker *et al.*, 2006a). A number of *Lactobacillus* species stimulate beta-defensin-2 secretion *in vitro* (Schlee *et al.*, 2008) and are presumably resistant to the anti-microbial effects of these peptides. *L. reuteri* 100-23 is also resistant to human beta-defensins 1 - 4 (Walter *et al.*, 2007). *L. delbrueckii* subsp *lactis* CIDCA 133 but not *L. delbrueckii* subsp. *bulgaricus* CIDCA 331 was resistant to human defensins (Hugo *et al.*, 2010).

Thus, the fate of the microorganisms that reach the duodenum alive is dependant on their ability to further withstand the many barriers to colonisation imposed by the host. However, the resident microbiota may also manipulate the microbiota locally, which facilitates the colonisation or adherence of some species at the expense of others.

1.1.3.2 *Microbial influences on the composition of intestinal microbiota.*

Bacteriocins are short ribosomally synthesised anti-microbial peptides that are produced by some bacteria to the detriment of other microbes (Cotter *et al.*, 2005). Bacteriocins may exhibit either a broad or a narrow spectrum of activity. Bacteriocin-producing bacteria harbour an immunity gene which protects them from the actions of

their own bacteriocin. Lactobacilli encode bacteriocins that are expressed *in vivo* (Corr *et al.*, 2007; Riboulet-Bisson *et al.*, 2012).

Lactobacilli also produce a number of other indiscriminate anti-microbials. Reuterin is a broad spectrum anti-microbial compound produced by *L. reuteri* during fermentation of glycerol (Talarico *et al.*, 1988) and is expressed *in vivo* (Morita *et al.*, 2008). Reuterin expression is stimulated by close contact with other bacteria and its anti-microbial effect may be partly due to its ability to subject target cells to oxidative stress (Schaefer *et al.*, 2010). Generic products of *Lactobacillus* metabolism, such as lactic-acid and hydrogen peroxide also have an anti-microbial role (see Appendix A).

Syntrophy in the GI tract may also influence the composition of the resident microbiota (Fischbach & Sonnenberg, 2010). This phenomenon, whereby one microorganism utilises digestive substrates and metabolic products released by the activities of another microbe (Duncan *et al.*, 2004) or through cellular decay (Rozen *et al.*, 2009), may be universally applied to ecoystems inhabited by complex microbial communities. *Lactobacillus* fermentation of dietary sugars releases lactic acid which can be used by other species in the gut (Duncan *et al.*, 2004).

1.1.3.3 *Influence of host genetics on the composition of intestinal microbiota.*

Animals engineered to be deficient in a particular Pattern Recognition Receptor (PRR) fail to stimulate typical signalling pathways, leading to alternations in the intestinal microbiotas of these animals (Petnicki-Ocwieja *et al.*, 2009; Vijay-Kumar *et al.*, 2010). Under normal circumstances the host immune system imposes selective pressures on the microbiota that influence and maintain its composition. Microbiota alterations resulting from atypical or compromised immune function may manifest in diseases such as Crohn's disease (Petnicki-Ocwieja *et al.*, 2009) or metabolic syndrome (Vijay-Kumar *et al.*, 2010). The composition of the murine

intestinal *Lactobacillus* microbiota is influenced by host genetics at the level of species (Benson *et al.*, 2010). Quantitative trait locus (QTL) analysis was used to determine if specific taxa cosegregated with linked genomic markers as quantitative traits. While QTL were not identified for the lactobacilli at genus level, the *L. johnsonii/L. gasseri* group of lactobacilli were associated with two QTL in mice, suggesting that host-genetics exerts its influence at lower taxonomic ranks, and that the intimate association between a host and some of its commensal species is influenced by heritable factors (Benson *et al.*, 2010). A separate study based on 16S rRNA gene sequencing found that host genetics had a greater influence than maternal effect on the *L. johnsonii* population in the murine gut (Buhnik-Rosenblau *et al.*, 2011).

1.1.3.4 *Extrinsic host factors that influence the* Lactobacillus *population of GI tract.*

The age of the human host has a considerable influence on the composition of its resident microbiota. In infancy, the microbiota is in a state of flux (Palmer *et al.*, 2007). The founding microbiota is established during or immediately after birth (Dominguez-Bello *et al.*, 2010) and an infant's microbiota is subsequently influenced by contact with adults, especially the mother while breast-feeding. Post-weaning, the infantile microbiota is influenced by food-associated microbes, such as lactobacilli (Adlerberth & Wold, 2009).

At about one year of age, a stable and "adult-like" personal microbiota has become established (Palmer *et al.*, 2007). This is mediated in part, by the transition to a solid food diet (Palmer *et al.*, 2007). While there may be some temporal variation in the intestinal microbiota over the first year of life, less intra- than inter-individual variation has been observed (Palmer *et al.*, 2007).

As individuals enter old age, the composition of their intestinal microbiota is again subject to change (Biagi *et al.*, 2010; Claesson *et al.*, 2011; Makivuokko *et al.*, 2010). Different reports have described the *Lactobacillus* microbiota as being larger and more diverse (Stsepetova *et al.*, 2011) or smaller and less diverse (Woodmansey *et al.*, 2004) than the *Lactobacillus* microbiota of younger adults.

Age related variation in the composition of the intestinal microbiota may reflect the status of the immune system. At birth, the immune system of human children have a limited capacity to respond to infections, and infants rely on maternally acquired antibodies to bolster their immune defences in early life (Holt & Jones, 2000). Animal studies have demonstrated that exposure to microbes assists development and maturation of the immune system (Chung *et al.*, 2012; Mazmanian *et al.*, 2005; Round & Mazmanian, 2010). The fully developed immune system is thought to distinguish commensal from pathogenic microbes (Srinivasan, 2010) and it likely influences the composition of the microbiota residing on and in the human body. Immune function and responsiveness gradually decrease with age in a process known as immunosenescence. "Inflamm-aging", a concept that describes a generalised, age-dependent reduction in the ability of the immune system to cope with stresses and antigens leading to an increased pro-inflammatory status, may affect elderly individuals (Franceschi *et al.*, 2000). Atypical immune responses may increase the likelihood of altered microbiota composition in these individuals.

Living environment and lifestyle choices also affect the composition of the human intestinal microbiota. Diet has a major influence on the composition of the GI microbiota (Claesson *et al.*, 2012; Wu *et al.*, 2011; Zimmer *et al.*, 2012) although probiotic consumption has only a subtle overall impact (Tannock *et al.*, 2000). Prebiotics are intended to selectively modify the intestinal microbiota by increasing

the population size and activity of certain target species (Roberfroid *et al.*, 2010). Antibiotics significantly alter the structure of the enteric microbial community (Jakobsson *et al.*, 2010), while smoking (Benjamin *et al.*, 2012), alcoholism (Mutlu *et al.*, 2012) and consumption of non-steroidal anti-inflammatory drugs (Makivuokko *et al.*, 2010) may also effect changes in the composition of the GI microbiota.

The establishment of the intestinal microbiota is a dynamic process and the microorganisms that transit and/or colonise the GI tract have many intrinsic and extrinsic selective pressures to overcome. Nonetheless, the GI tract is a nutrient rich and relatively stable environment. Accordingly, many species, including the autochthonous lactobacilli, have evolved mechanisms to overcome the environmental stressors imposed by the host so that they may colonise the GI tract.

1.1.4 The interkingdom relationship between a host and its commensal microbiota.

"Commensalism" is a form of symbiosis in which one member of the partnership derives a benefit from the interaction, while the other member remains unaffected (Casadevall & Pirofski, 2000). In the intestines, the commensal bacteria are typically non-invasive and are found in the lumen and in the surface mucus layers. The intestinal microbiota do however exert significant health benefits on the host, leading some authors to question if "commensalism" adequately describes this form of symbiosis (Walter *et al.*, 2011).

1.1.4.1 *The influence of the commensal microbiota on host physiology.*

The commensal microbiotas of mammals have a considerable impact on host physiology and immune defences. Microbial colonisation of the murine GI tract stimulates angiogenesis and structural development of the mucosal capillary vasculature that is effected via Paneth cells in the ileum (Stappenbeck *et al.*, 2002).

Repair, maintenance and protection of the intestinal epithelium are also mediated by the commensal microbiota (Rakoff-Nahoum *et al.*, 2004; Swanson *et al.*, 2011). For example, the generation of reactive oxygen species by commensals such as *L. rhamnosus* strain GG contributed to enhanced cell migration and faster wound healing in mice (Swanson *et al.*, 2011). Furthermore, the host may be protected from intestinal injury by recognition of microbe associated molecular patterns (MAMPs) from the commensal microbiota (Rakoff-Nahoum *et al.*, 2004).

The microbiota plays a multi-faceted role in immune maturation and tolerance. The structural proteins and polysaccharides of commensal microbes may direct cellular maturation of the host's immune system (Mazmanian *et al.*, 2005; Round & Mazmanian, 2010). Certain combinations of commensal species can stimulate the development of gut associated lymphoid tissue (GALT) in rabbits (Rhee *et al.*, 2004) while specific commensal species induce mucosal T-cell responses in mice (Gaboriau-Routhiau *et al.*, 2009). Ingestion of *L. plantarum* in different growth phases stimulated differential nuclear factor kappa B (NF-κB) mediated immune signalling that was consistent with the development of immune tolerance to the bacterium in the duodena of healthy human adults (van Baarlen *et al.*, 2009). *Bacteroides fragilis* actively suppresses immune responses to promote tolerance and permit colonisation (Round *et al.*, 2011).

The microbiota also has an anatomical impact on the host. The caeca of germ-free mice are characteristically, but reversibly, enlarged (Reikvam *et al.*, 2011) and the swelling is caused by an accumulation of mucus (Lindstedt *et al.*, 1965) in the absence of microbes. Axenic animals lack a microbiota and suffer a range of physiological, metabolic and immune system deficits, some of which can be overcome by recolonisation (Smith *et al.*, 2007).

Furthermore, the metabolic activities of the commensal microorganisms may additionally provide the host with beneficial metabolites such as SCFAs (Macfarlane & Macfarlane, 2012) and vitamins (LeBlanc *et al.*, 2012). The resident microbiota also confers colonisation resistance which prevents the establishment of pathogens in this niche (Kamada *et al.*, 2012).

1.1.4.2 *Host PRRs detect MAMPs and allow the host to sense and respond to its commensal microbiota.*

The activities of the commensal microbiota result in microbial and immune homeostasis in the GI tract. PRRs allow the host to sense and respond to its resident microbiota by recognising MAMPs. These PRRs occur either on the surface or in the cytosol of epithelial, endocrine and immune cells or may be associated with specific intracellular compartments (Table 1.1). PRRs may also be secreted (Agrawal *et al.*, 2009; Gomez *et al.*, 2009). Several PRR families contribute to innate immunity in humans and other animals (Takeuchi & Akira, 2010). The best characterised of these are the Toll-like receptors (TLRs) and the nucleotide-binding domain and leucine-rich repeat-containing receptors (NLRs).

1.1.4.2.1 *Toll-like receptor family.*

The toll-like receptors are a family of transmembrane receptors that recognise and mount an immune response to bacterial and viral components. Ten TLRs have been described for humans (Table 1.1). TLRs 1, 2, 4, 5, 6 and 10 are cell-surface receptors with extracellular and cytoplasmic domains. In contrast, TLRs 3, 7, 8, 9 are located at intracellular organelles (endoplasmic reticulum, endosome, lysozome) and respond to nucleic acids from bacteria and viruses (Table 1.1).

TLR signalling ultimately requires specific and regulated communication between the stimulated receptor and the transcription factors in the cell's nucleus to

Table 1.1: Cellular location and ligands of Toll like receptors.

Receptors	In Humans?	Location	Ligand	PMID*
TLR1	Yes	Cell surface	Lipopeptides	12077222
TLR2	Yes	Cell surface	Peptidoglycan, Lipoproteins, Haemagglutinin	10364168, 10426996, 12163593
TLR3	Yes	Intracellular compartment	dsRNA	11607032
TLR4	Yes	Cell surface	Lipopolysaccharides, Viral structural proteins	10201887, 11062499
TLR5	Yes	Cell surface	Flagellin	11323673
TLR6	Yes	Cell surface	Lipoproteins	11431423
TLR7	Yes	Intracellular compartment	ssRNA	15034168
TLR8	Yes	Intracellular compartment	ssRNA	14976262
TLR9	Yes	Cell surface; Intracellular compartment	DNA, Haemozonin	11130078, 20114028
TLR10	Yes	Cell surface	Lipopeptides (with TLR2)	20348427
TLR11	No, mouse	Cell surface	Uropathogneic bacteria, Profillin	15860593, 15001781
TLR12	No (pseudogene), mouse	Unknown	Unknown	19077284, 1172252
TLR13	No, mouse	Intracellular	23S rRNA sequence, Vesicular stomatitis virus	22821982, 21131352

*PMID: PubMed accession number

initiate pro-inflammatory gene transcription. The immunostimulatory signal is relayed to the nucleus via adapter molecules and protein kinases (O'Neill & Bowie, 2007). A conserved Toll/Interleukin-1 (TIR) domain that is common to the cytoplasmic portion of the TLRs and to their adaptor molecules facilitates initial signal propagation between these receptor and adaptor molecules (O'Neill & Bowie, 2007).

Structurally, the TLRs are transmembrane proteins. Their amino termini are responsible for specific ligand recognition. Their horseshoe-shaped ecto-domains are typically glycoproteins that include a variable number of leucine rich repeats (LRRs) in their structure (Botos *et al.*, 2011). The ecto-domains are found either on the cell surface or on the inner-membrane side of intracellular organelles and compartments, and are responsible for ligand recognition and binding. The cytoplasmic carboxy-terminus of each TLR contains the TIR domain and a transmembrane helix links the inner and outer domains (Botos *et al.*, 2011).

TLRs may function as either homo- (Zhou *et al.*, 2012) or hetero- dimers (Kang *et al.*, 2009). Engagement of the target ligand usually initiates conformational changes in the receptor dimers and prompts adaptor molecule recruitment (O'Neill & Bowie, 2007) and activation of downstream signalling pathways. The exact combination of receptor molecules in each dimer, and the adaptor molecule(s) recruited, contributes to the specificity of ligand-recognition and the outcome of signalling (Farhat *et al.*, 2008; O'Neill *et al.*, 2009).

1.1.4.2.2 *The nucleotide-binding domain and leucine-rich repeat-containing receptor family.*

The NLR family of innate immune receptors are structurally and functionally conserved across the animal and plant kingdoms, and like the TLRs, they have a tripartite structure (Proell *et al.*, 2008). The carboxy- terminus of each NLR also contains leucine rich repeats, and it is this carboxy- terminus that is believed to be responsible for sensing MAMPs (Proell *et al.*, 2008). Four types of effector domain distinguish the subfamilies of NLR (Ting *et al.*, 2008). Signal transduction is initiated by recruitment of effector molecules leading to pro-inflammatory gene expression or cell death (Lavelle *et al.*, 2010; Proell *et al.*, 2008). Given their cytoplasmic location, the NLRs respond to intracellular, cytosolic MAMPs.

1.1.4.3 ***MAMPs and the discrimination of pathogens and commensals.***

The term "pattern recognition receptor" was originally proposed by Janeway in 1989 to describe receptors that would recognise pathogen-associated molecular molecules (PAMPs) (Janeway, 1989). These PAMPs are microbial molecules bearing non-host signatures that are capable of triggering innate immune responses via PRRs. However, these molecular patterns are common to pathogenic and commensal microbes alike. Therefore in recent times, the more inclusive acronym "MAMP" (microbe associated molecular pattern) has been used to refer to the molecules that trigger these innate immune responses.

MAMPs may be broadly described as molecules with microbial signatures that are directly recognised as foreign or "non-self" by the host's immune system. MAMPs are often the ligands of one or more of the PRRs described above. These soluble and insoluble MAMPs may be secreted, or may be found within or on microbial surfaces. The universal occurrence of MAMPs among microorganisms

means that the immune system must rely on additional cues to distinguish pathogens from commensals.

Blander and Sander (2012) suggest that the severity of the immune response is correlated to the perceived magnitude of a microbial threat. They outline using five non-hierarchical checkpoints, how the scale of the immune response may be heightened in response to viable microorganisms and pathogens, while soluble MAMPs and dead microbes induce milder inflammatory responses (Blander & Sander, 2012).

One such checkpoint is centred on the concept of vita-PAMPs (Sander *et al.*, 2011) or PAMPs-pv (*pro-vita*) (Vance *et al.*, 2009), which are molecules that are indicative of viable, growing microorganisms. Examples of PAMPs-pv include quorum-regulating autoinducers, cyclic-di-GMP, tracheal cytotoxin and components of DNA replication (Vance *et al.*, 2009). These molecules are synthesised by active processes and therefore represent viable cells.

A complementary concept is that of PAMPs-pm (*post-mortem*) (Herskovits *et al.*, 2007; Vance *et al.*, 2009). These molecules may signify that an invading microorganism has already been killed and therefore poses a low-risk threat so that dampened or alternative immune responses, aimed at clearing the debris from the cells, may ensue (Herskovits *et al.*, 2007; Sander *et al.*, 2011). Examples of PAMPs-pm have included muramyl dipeptide, macromolecular DNA and RNA and other bacterial degradation products which represent lysed and therefore non-viable cells (Herskovits *et al.*, 2007; Vance *et al.*, 2009).

In practice, a synergism may exist between the detection of PAMPs-pv or PAMPS-pm and regular (M/P)AMPs to shape immune responses. The outcome of signalling is also influenced by the compartmentalisation of MAMPs.

1.1.4.4 *Compartmentalisation of PRRs and MAMPs.*

Within the human body, different organs and anatomical sites are variably permissive of colonisation by microorganisms. While the lungs cannot function if colonised with microorganisms, the enteric microbiota is required for intestinal development and maturation as described earlier. However, even non-pathogenic, avirulent, commensal organisms can trigger severe immune responses when they occur in normally sterile tissues (Sander *et al.*, 2011).

As previously mentioned, the microbiota in the gut is largely confined to the lumen and surface mucus layer. The inner mucus layer is usually free of bacteria and tight junctions exist between the epithelial cells of the mucosa, denying the microbiota access to the tissues below. Some transmembrane PRRs are strategically located to recognise when these compartmentalisation barriers have been breached. For example, TLR5 is expressed on the basolateral sides of colonic epithelial cells (Gewirtz *et al.*, 2001). Therefore, TLR5 may trigger an immune response to flagellin proteins only when the integrity of the intestinal epithelial barrier has been compromised. In the lung, however, TLR5 may be mobilised to the apical surface of epithelial cells during infections with flagellate microorganisms so that it can contribute to the immune responses aimed at clearing the infection (Adamo *et al.*, 2004). AsialoGM1, a TLR5 co-receptor for flagellin is also expressed on the apical surface of lung epithelia (Adamo *et al.*, 2004), suggesting that infections with flagellate microbes are extremely undesirable in the lung, unlike in the intestinal epithelium which endures chronic apical exposure to flagellate species. These tissue-specific differences in PRR localisation highlight how the relationship between host and microbe varies according to the physiological function of the anatomical site.

Just as cellular polarisation and localised receptor expression contribute to compartmentalisation in the gut, the existence of surface and cytosolic PRRs contribute to compartmentalisation at a cellular level. Cytoplasmic PRRs may cooperate with cell surface receptors to mount an immune response to a pathogen (Rad *et al.*, 2009). Even within a cell, the immune responses triggered by bacteria in phagosomes differ from those that would be triggered by the same bacterium if present in the cytoplasm (Herskovits *et al.*, 2007; McCaffrey *et al.*, 2004). Thus, compartmentalisation of PRRs helps the host to distinguish "colonisers" from "invaders" and virulent from avirulent species, but always in the context of the anatomical site and its physiological function.

1.1.4.5 *Microbiota-host dialogue in the intestine.*

Interkingdom signalling in the intestine may result in microbiota-derived molecules impacting on the host, or vice versa. The ability of the microbiota and the host to sense and respond to chemical signals such as hormones, quorum sensing molecules, secreted immunomodulators and metabolites contributes to this interspecies and interkingdom dialogue.

1.1.4.5.1 *Microbial endocrinology and the effects of neurochemicals on the microbiota-gut-brain axis.*

The concept of "microbial endocrinology" proposes that the microbiota can sense and respond to the neuroendocrine signals of the host, and similarly, that the host can respond to the hormone-like signals produced by microorganisms (Lyte, 2004). A significant neural network exists in the intestines and signalling through the microbiota-gut-brain axis may prompt immediate physiological (Hagbom *et al.*, 2011; Hu *et al.*, 2007) or longer-term behavioural responses (Bravo *et al.*, 2011; Heijtz *et al.*, 2011) in the host.

Some microorganisms produce neurotransmitters (Roshchina, 2010) which may impact on signalling throught the microbiota-gut-brain axis. Certain *Lactobacillus* strains produce the major inhibitory neurotransmitter γ-amino-butyric acid (GABA) (Barrett *et al.*, 2012; Li *et al.*, 2010). Bacteria use glutamate decarboxylase enzymes to produce GABA by the α-decarboxylation of L-glutamic acid (Roberts & Frankel, 1950). Furthermore, comensal bacteria may modulate the expression of GABA receptors in the brain (Bravo *et al.*, 2011).

1.1.4.5.2 *Lactobacillus-host communication via other microbiota-derived signalling molecules and metabolites.*

Lactobacillus-derived molecules and metabolites also influence the host immune system. *L. reuteri* 6475 harbours a histidine decarboxylase gene cluster which is involved in the conversion of L-histidine to histamine (Thomas *et al.*, 2012). The histamine produced acts as an immunoregulatory signal that suppresses tumour necrosis factor - α (TNF-α) production in human cell lines (Thomas *et al.*, 2012). Secreted proteins such as p40 and p75 are autolysins that are found in several lactobacilli including *L. rhamnosus* GG and *L. casei* BL23. These proteins were shown to have immunomodulatory properties and could also stimulate cell proliferation and demonstrate anti-apoptotic effects in human and mouse intestinal epithelial cells (Bauerl *et al.*, 2010; Yan *et al.*, 2007). *L. paracasei* and *L. casei* strains may secrete lactocepin, a protease that can degrade pro-inflammatory chemokines, leading to "physiologically relevant anti-inflammatory effects" (von Schillde *et al.*, 2012). *Lactobacillus*-derived metabolites such as lactic acid, may also contribute to enteroprotection by interfering with the lipopolysaccharide-TLR4 signalling pathway, and could contribute to the protection from indomethacin-induced small intestinal injury in rats (Watanabe *et al.*, 2009).

1.1.4.6 *Suppression and manipulation of the host's immune responses to facilitate microbial colonisation of the intestines.*

Although the host is capable of discriminating pathogens from commensals, recent research has shown that certain species of the microbiota actively manipulate, suppress and/or exploit host immune function to facilitate their colonisation of this niche.

L. reuteri strain 100-23 is an autochthonous species of the murine forestomach. It produces an exopolysaccharide (EPS) with immunomodulatory properties that also impacts on its colonisation of the GI niche (Sims *et al.*, 2011). The fructosyl transferase (*ftf)* gene was found to be responsible for EPS production in *L. reuteri* 100-23 and insertional mutagenesis targeting this gene abolished EPS production by this strain. While the EPS negative mutant strain successfully colonised mono-associated mice, it had reduced colonisation ability when it was administered in competition with the wild-type strain, to *Lactobacillus*-free mice. Furthermore, the mutant strain had an impaired immunomodulatory capacity. More T_{reg} cells were recorded in the spleens of mice colonised with *L. reuteri* 100-23 than in those colonised by the mutant strain or in *Lactobacillus*-free animals. Thus, *L. reuteri* EPS production played a role in both the colonisation of the GI tract and in the induction of T_{regs} in the systemic immune system (Sims *et al.*, 2011).

Similarly, the polysaccharide component of the *L. casei* strain Shirota cell wall can suppress innate immune responses, such as secretion of pro-inflammatory cytokines in murine macrophage derived cell lines (Yasuda *et al.*, 2008). Polysaccharide negative mutants induced more severe pro-inflammatory immune responses than wild-type bacteria (Yasuda *et al.*, 2008).

NF-κB also serves as a target for bacteria that modulate host immune responses for colonisation purposes. Signalling that initiates from a number of different receptors may eventually converge on the NF-κB family of eukaryotic transcription factors, which effects pro-inflammatory gene transcription. Thus, by targeting NF-κB, microbes can effectively manipulate the immune responses of the host (Kelly *et al.*, 2004; Sokol *et al.*, 2008). Several *Lactobacillus* species either activate or repress NF-κB. For example, *L. ruminis* may stimulate transcription of NF-κB (Taweechotipatr *et al.*, 2009) while *L. reuteri* downregulates it (Iyer *et al.*, 2008). This type of immunomodulation could contribute to the probiotic effects of certain *Lactobacillus* strains.

1.2 Bacterial motility and motile commensals of the GI tract.

1.2.1 General introduction to motility organelles and the different modes of bacterial motility.

The bacterial flagellum is a sophisticated motility organelle and secretory apparatus that consists of a whip-like, extracellular structure that is rotated by its wall anchored motor component to generate thrust for bacterial propulsion over solid and through liquid environments. The number and surface distribution of flagella vary among bacterial species, ranging from single polar flagella to multiple peritrichous flagella (Table 1.2). Flagellum rotation is usually highly coordinated and reversible, which gives rise to the running and tumbling phenotypes of swimming bacteria (Macnab, 1996). Proton and sodium motive forces drive flagellum rotation (Hirota *et al.*, 1981; Manson *et al.*, 1977).

Type IV pili are cell-surface structures that are assembled from polymers of a protein called pilin. Like flagella, pili may cover the surface of a bacterium or alternatively, may be found localised at the poles of the cell. Pili contribute to niche

colonisation, protein secretion and DNA transfer, retraction of type IV pili that facilities specific forms of bacterial movement (Merz *et al.*, 2000). A number of different motility mechanisms, some of which are motility-organelle independent have been described (Table 1.3).

1.2.1.1 ***Modes of bacterial motility.***

1.2.1.1.1 *Swimming motility.*

Swimming motility is the best characterised of all the bacterial motility types and it relies on the rotation of flagella. Swimming bacteria "run" and "tumble". Runs are characterised by smooth, directional movement along a concentration gradient, while "tumbles" describe the end-over-end rotational movement of the bacterium that allow it to change course. The direction of flagella rotation determines whether a cell "runs" or "tumbles". Counterclockwise flagella rotation pushes the bacterium forward, while clockwise rotation causes the cell to tumble (Macnab, 1999). When engaged in a "run", the flagella of multi-flagellate bacteria associate to form a bundle in which the rotation of flagella is synchronised (Macnab, 1977). Disruption of the rotating flagella bundle causes the cell to "tumble.

1.2.1.1.2 *General mechanism of chemotaxis.*

Chemotaxis is the process by which bacteria sense and respond with movement, to chemical concentration gradients in their environment. Bacteria use chemotaxis to seek out more favourable surroundings by moving towards chemo-attractants (often nutrients) or away from chemo-repellents (toxins). Chemoreceptors in the bacterial cell membrane sense concentrations of chemo- attractants and repellents. The bacterial chemotaxis system "adapts" to concentrations of chemo-attractants. Adaptation is akin to determining "background" levels of a particular chemo- attractant or repellent, so that a change in the concentration of the target

Table 1.2: Flagella distribution patterns.

Flagellation pattern	Description	Examples	PMID
Peritrichous	Many flagella emerge from sites all over the bacterial cell surface.	*Bacillus* spp, *Clostridium* spp, *Escherichia coli*	Bergey's Manual, 20156455
Lophotrichous	Two or more unipolar flagella.	*Bartonella bacilliformis*, *Helicobacter pylori*	(Minnick & Anderson, 2006), 2252384
Amphitrichous	One or more flagella present at each pole of the bacterial cell. May also be described as "bipolar" flagella.	*Campylobacter jejuni*, *Gallaecimonas xiamenensis*	22144902, 22659502
Monotrichous	A single flagellum occurs at only one pole of the bacterial cell. May also be described as a "single, polar" flagellum.	*Pseudomonas aeruginosa*, *Vibrio parahaemolyticus*	10758151, 6159520

molecule can be perceived. This means that the stimulated chemotaxis system of the bacterium can be "reset" to pre-stimulus levels, even in the continued presence of chemo-attractants (Macnab & Koshland, 1972; Rao *et al.*, 2008). Thus, chemotaxis involves a significant temporal component. In practice, bacteria respond to the rate of change of receptor-bound molecules rather than to the absolute number of chemo-attractant or repellent molecules in their environment (Rao *et al.*, 2008; Szurmant & Ordal, 2004).

In *B. subtilis* and *E. coli*, chemotaxis relies on a two-component signal-transduction pathway involving CheA and CheY to relay environmental signals to the flagellar motor. The presence or absence of receptor-bound chemo-attractants is sensed by either methylation or demethylation of the chemoreceptors and this either

Table 1.3: Modes of bacterial motility.

Type of motility	Appendages required	Description of the movement	Examples	References (PMID)
Swimming	Flagella	Movement through liquid, propelled by flagella.	*Bacillus subtilis*, *Escherichia coli*, *Lactobacillus ruminis*, *Salmonella typhimurium*	PMC216775, PMC315881, 4723944, 7530362
Swarming	Flagella	Multicellular movement across a solid surface, propelled by flagella.	*Bacillus subtilis*, *Lactobacillus ruminis*, *Proteus mirabilis*	20694026, 20383990, 8932309, 16030230
Gliding	(Type IV pili)	A smooth directional movement of bacteria over surfaces, usually unaided by flagella or pili.	*Bdellovibrio bacteriovorus*, *Lactobacillus koreensis*, *Myxococcus xanthus*	20435754, 21515772, 10996798
Twitching	Type IV pilli	A jerky surface-associated, flagellum-independent mode of motility in which movement is achieved by the forceful extension and retraction of type IV pili.	*Neisseria gonorrhoeae*, *Pseudomonas aeruginosa*	10993081, 11381130, 10537208
Sliding	None	A passive form of movement that is achieved by the expansive forces of a growing colony under conditions of low surface tension.	*Mycobacterium smegmatis*, *Mycobacterium avium*, *Pseudomonas aeruginosa*	10572138, 18065549
Darting	Uncharacterised	Sudden, jerky movement achieved by the ejection of bacteria from a cellular aggregate. Movement of the bacteria only happens during the ejection event and the bacteria are not otherwise motile.	*Staphylococcus albus* (HV 54)	4631369
Surfing	Flagella	A mucin-dependent form of surface translocation.	*Pseudomonas aeruginosa*	22550036

activates or inhibits CheA kinase activity, which modifies the intracellular concentration of phosphorylated CheY (Szurmant & Ordal, 2004). Motile bacteria can have more than one adaptation system, and in *B. subtilis*, these function synergistically (Rao *et al.*, 2008).

1.2.1.1.3 *Swarming motility.*

Swarming may be defined as the "rapid, multicellular movement of bacteria across a surface, powered by rotating flagella" (Kearns, 2010). Thus, unlike swimming bacteria that are capable of unicellular movement, bacteria at a swarm-front arrange themselves side-by-side to form raft-like structures and move across surfaces in this multicellular arrangement (Darnton *et al.*, 2010; Girgis *et al.*, 2007). Furthermore, when compared to flagellate bacteria of the same species from liquid environments, swarming cells are elongated (Belas & Suvanasuthi, 2005; Darnton *et al.*, 2010) and hyperflagellate (Harshey & Matsuyama, 1994). This may reflect a need for greater propulsion to move across solid surfaces. Swarming bacteria often secrete surfactants and wetting agents to further ease their movement across surfaces (Burch *et al.*, 2012; Chen *et al.*, 2007; Ghelardi *et al.*, 2012).

To transition from swimming to swarming, a bacterium must sense its new, solid environment. Conditions that inhibit the rotation of flagella, such as high viscosity medium or tethering of the flagella by anti-flagellin antibodies, were found to induce swarmer cell differentiation in *Proteus mirabilis* (Belas & Suvanasuthi, 2005). Unlike swimming cells, swarming cells may run, but in general, do not tumble (Darnton *et al.*, 2010). Reversal of flagellar rotation may cause swarming cells to reverse the direction of their motility (Turner *et al.*, 2010). Chemotaxis genes are required for swarming in *S. typhimurium*, but have a mechanical rather than a chemosensory role (Mariconda *et al.*, 2006). *L. ruminis* strains may swim (Sharpe *et al.*, 1973) and swarm (Willing *et al.*, 2009).

Several other types of bacterial motility are described in Table 1.3. So many modes of bacterial movement suggest that motility is an advantageous trait for bacteria.

1.2.1.2 *Roles of bacterial flagella.*

The bacterial flagellum is a multi-protein, multipartite structure that has been best characterised for its roles as a motility organelle and as a molecular motor. Nevertheless, the flagellum has several functions in bacteria, apart from locomotion. For example, some flagellate bacterial pathogens such as *C. jejuni*, exploit the flagellar type III secretion system (T3SS) to secrete non-flagellar virulence factors (Christensen *et al.*, 2009; Konkel *et al.*, 2004). Interspecies communication between the bacterium *Pelotomaculum thermopropionicum* and the archeon *Methanothermobacter thermautotrophicus*, which facilitates syntrophy and symbiosis between these species, depends on the perception of specific protein components of the bacterial flagellum by the archeon (Shimoyama *et al.*, 2009). The transcriptional response induced in *M. thermautotrophicus* following exposure to the flagellar capping protein FliD, prepares it for syntrophy with *P. thermopropionicum* (Shimoyama *et al.*, 2009).

Flagella also act as sensors of environmental wetness (Wang *et al.*, 2005) and as mechanosensors that perceive the prevailing conditions of the colonisation environment (Anderson *et al.*, 2009). The regulation of biofilm formation (Blair *et al.*, 2008; Houry *et al.*, 2010) and niche colonisation (Eaton *et al.*, 1992) may also be influenced by either the presence or absence of flagella.

1.2.1.3 *Architecture of a typical flagellum.*

The three distinct parts of the bacterial flagellum are the basal-body motor complex, the hook and the filament.

1.2.1.3.1 *Structure and function of the basal-body.*

The basal-body and motor complex are found at the base of the flagellum and constitute the major intracellular component of the flagellum. The basal-body is a

symmetrical structure that revolves around its vertical axis to transmit torque from the motor to the hook and filament (Macnab, 2003).

A typical basal body is formed from a central rod and a number of mounted, coaxial rings, but differences exist in the basal-body architectures between different species, particularly between Gram positive and Gram negative bacteria (Figure 1.2). While a hollow central rod structure is common to the basal-bodies of all bacteria, a different number of rings may be found in each structure. More rings are associated with the Gram negative than the Gram positive basal-body. Gram negative bacteria harbour the L, P, M, S and C rings which are located in the outermembrane (L), peptidoglycan layer (P), cell membrane (M, S) and cytoplasm (C) (Chen *et al.*, 2011). The L and P rings associate to form a rigid and stationary structure that serves as a bushing to protect the central rod from sheer forces during its rotation (Chen *et al.*, 2011). The M and S rings associate to form an interface with the rod (Suzuki *et al.*, 2004). The C ring is localised on the cytoplasmic side of the MS ring, and its size varies from species to species (Chen *et al.*, 2011). Differences in cell wall architecture between Gram positive and Gram negative bacteria mean that flagellate Gram-positive bacteria lack the L and P rings (DePamphilis & Adler, 1971) (Figure 1.2).

The basal-body also operates as an export apparatus. Proteins that are required for the construction of the extracellular components of the bacterial flagellum are exported from their site of synthesis in the cytoplasm via the flagellar T3SS which lies within the MS ring (Minamino & Macnab, 1999). The export of proteins for flagellum assembly is carefully ordered and regulated. During flagellum biogenesis, the substrate specificity of the T3SS apparatus varies depending on whether the hook has been fully assembled or not. This substrate specificity reflects

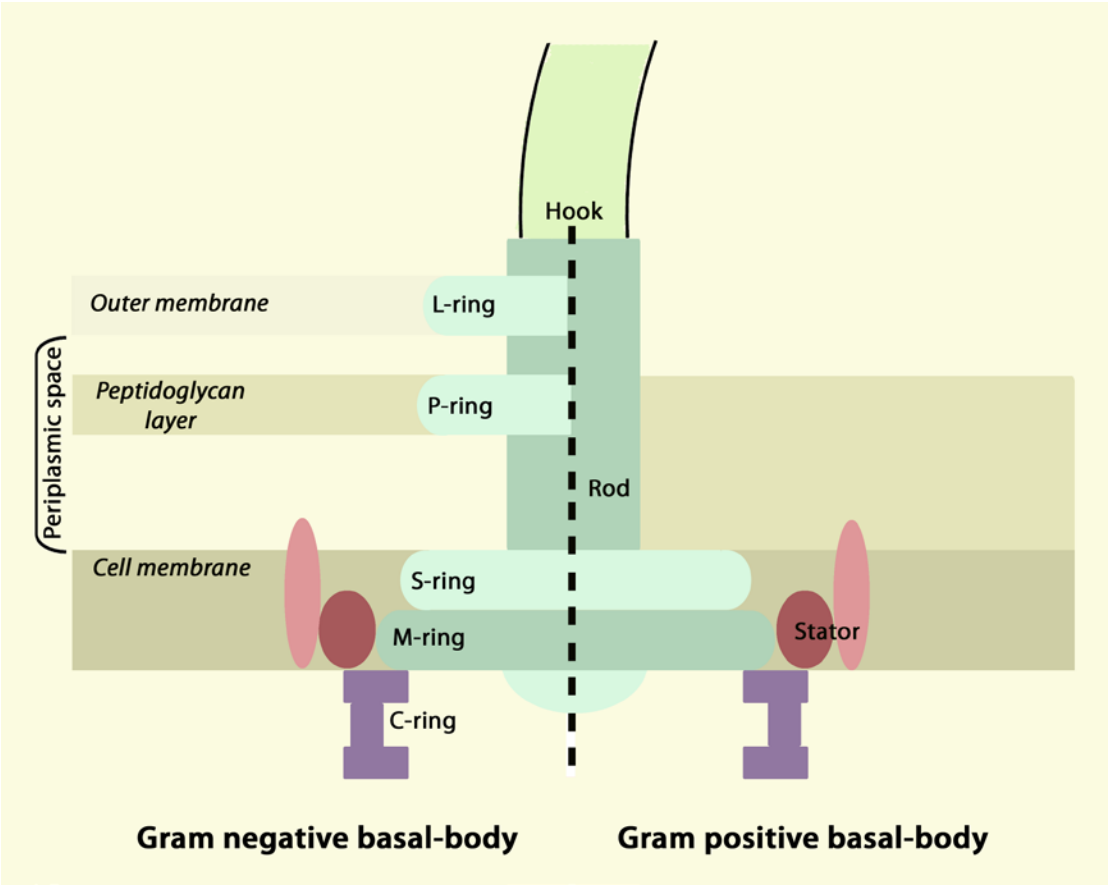


Figure 1.2: Structural differences between basal-bodies of Gram positive and Gram negative bacteria. Due to differences in the structure of the cell wall between Gram negative and Gram positive bacteria, there are fewer rings in the Gram positive than in the Gram negative basal-body structure. A dotted line separates the two structures in this diagram.

the careful regulation of flagellum synthesis that requires completion of the hook before assembly of the filament can begin. Thus until the hook has been fully assembled, the export apparatus preferentially exports hook proteins. Once the hook has been completed, the specificity of the export apparatus changes and FlgM (an inhibitor of σ^{28} regulated genes involved in filament formation) is exported, which permits transcription, translation and export of filament proteins (Brown *et al.*, 2009).

The flagellar motor is composed of two parts, a revolvable rotor and a stator that occupies a stationary point of reference in the cell (Macnab, 2003). The rod and

its movable axial rings together comprise the rotor. The stator is composed of motor proteins, MotA and MotB in H^+ driven motors, or their homologs PomA and PomB, MotX and MotY or MotP and MotS in Na^+ driven motors (Chen *et al.*, 2011; Ito *et al.*, 2005; Terashima *et al.*, 2006). These motor proteins typically follow a 4:2 proteinA:proteinB stoichiometry in a single stator complex (Kojima & Blair, 2004). However, many stator complexes surround the base of each rotor between the MS and C rings (Chen *et al.*, 2011). The actual number of stator complexes present varies from species to species (Chen *et al.*, 2011). For example, 16 and 13 stator complexes form a ring around each flagellar rotor in *Borellia burgdorferi* and *Hylemonelia gracilis* respectively (Chen *et al.*, 2011). These stator complexes interact with the C ring proteins to convert the inward flow of ions down an electrochemical gradient into torque that drives flagellar rotation (Chen *et al.*, 2011). Near zero load, the speed of motor rotation is independent of the number of stator units present and one stator is sufficient to facilitate rotation (Ryu *et al.*, 2000; Yuan & Berg, 2008).

A motor switch complex is required to control the transition from counter-clockwise to clockwise movement and it is influenced by chemotaxis signals. Modification of a particular helix of one FliG protein causes it to change from a conformation that favours counter-clockwise rotation to one that promotes clockwise rotation. It is thought that a conformational change in one FliG molecule impacts on neighbouring molecules until all of the FliG molecules in the ring have adopted the new conformation (Minamino *et al.*, 2011). Interestingly, FliG is the interaction target of EspE, a protein that serves as a molecular clutch which arrests *B. subtilis* flagellar rotation to promote biofilm formation by disengaging the rotor from its power-source (Blair *et al.*, 2008).

1.2.1.3.2 *Structure and function of the hook.*

The hook is a curved, flexible and hollow tubular structure located between the rigid structures of the basal-body and flagellar filament that also functions as a universal joint (Samatey *et al.*, 2004). This flexure generates propulsion from rotary motion as it transmits torque from the motor to the filament.

Structural analysis of the hook revealed similarities and differences between this flagellar substructure and the flagellar filament. A number of protofilaments run along the length of the hook and filament. These protofilaments are arranged in a ring around a narrow central channel through which the proteins for export pass (Figure 1.3) (Galkin *et al.*, 2008; Samatey *et al.*, 2004). To accommodate the curvature of the hook, the protofilaments within it must be either extended or compressed. Indeed, the protofilaments at the inner and more curved side of the hook were found to be more compressed than those at the outer edge (Samatey *et al.*, 2004). Each of the protofilaments will be either compressed or extended as the hook rotates (Samatey *et al.*, 2004). FlgE is the major protein component of the hook protofilaments and the adaptable intra- and inter- molecular interactions of FlgE proteins in these protofilaments contributes to the considerable flexibility of the hook (Samatey *et al.*, 2004).

Hook length and curvature is carefully controlled and differs between species (Waters *et al.*, 2007) and may determine the efficacy of propulsion. For example, the hooks of monotrichously flagellate bacteria may be longer than those of bacteria with peritrichate flagella (Waters *et al.*, 2007). These longer hooks might be necessary so that a single flagellum acting alone can still propel a bacterium forward, or alternatively because a single flagellum may need to withstand a greater torque load (Power *et al.*, 1992). For peritrichously flagellate bacteria that swim, the flexibility of

the hook allows for synchronous rotation of flagella with different foci (Brown *et al.*, 2012).

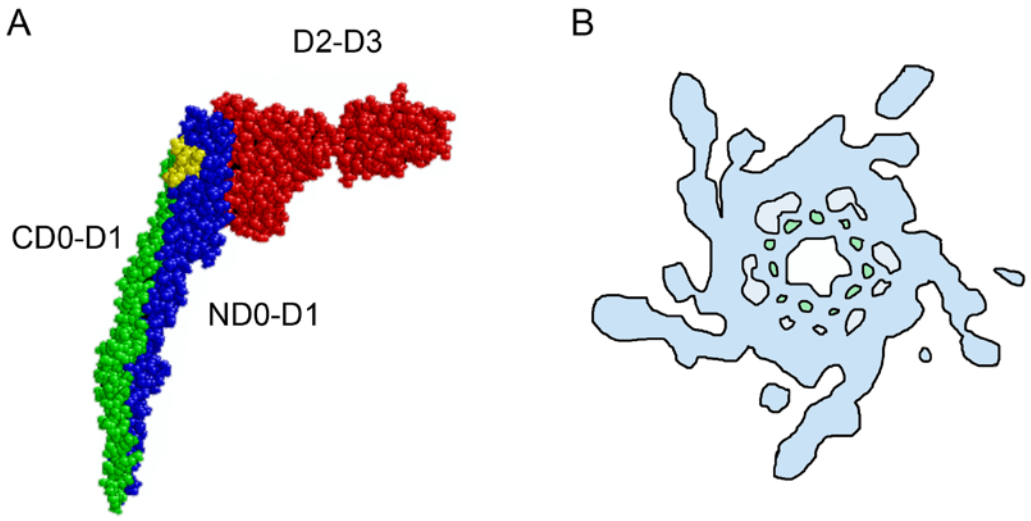


Figure 1.3: Structure of a flagellin monomer and the flagellar filament.
A: Structure of the flagellin protein of *Salmonella typhimurium* with various domains labelled. Image was created from RCSB Protein Data Bank accession 1ucu using RasMol (rasmol.org). B: An outline of a transverse section of a model of the *S. typhimurium* flagellar filament, showing the central lumen, and protofilaments (coloured green). Adapted from (Galkin *et al.*, 2008).

1.2.1.3.3 *Structure and function of the hook filament junction.*

The hook-filament junction is demarcated by the proteins FlgL (Hook Associated Protein (HAP) 3) and FlgK (HAP1) which buffer the transition between the two structures with distinct mechanical properties (Samatey *et al.*, 2004). The flagellar capping protein, FliD (HAP-2), is initially located upon the hook filament junction protein complex, but moves away from this junction as the filament increases in length (Homma & Iino, 1985; Yonekura *et al.*, 2000).

1.2.1.4 *Structure and function of the flagellar filament.*

The bacterial flagellar filament is a long, rigid, helical structure that functions as a propeller for swimming and swarming. Structurally, the filament is composed of a number of protofilaments arranged around a central channel (Galkin *et al.*, 2008). Protofilaments are formed by the polymerisation of one or more different flagellin proteins (Kostrzynska *et al.*, 1991; Samatey *et al.*, 2001). The number of protofilaments per flagellar filament varies among bacterial species. For example, *S. typhimurium* and *C. jejuni* have eleven and seven protofilaments per flagellum, respectively (Galkin *et al.*, 2008). Flagellar protofilaments occur in different supercoiled states. When all the protofilaments are of the same conformation, the flagellar filament is straight, and depending on the handedness of the helical turn of the protofilaments, the filament is described as being in either the L (left-handed) or R (right-handed) form (Maki-Yonekura *et al.*, 2010; Samatey *et al.*, 2001). The various protofilaments within a single flagellum may adopt either the L or R form (Darnton & Berg, 2007). This impacts on the packing of flagellin subunits and on the helical pitch of the protofilaments (Maki-Yonekura *et al.*, 2010). The structural conformations that may be adopted by a flagellar filament range from straight to curly, and are determined by the ratio of L to R protofilaments present (Darnton & Berg, 2007; Maki-Yonekura *et al.*, 2010). Chemotaxis induced motor reversal causes the protofilaments to change from an L (forward swimming, counterclockwise flagellum rotation) to an R (tumbling, clockwise flagellum rotation) conformation in a phenomenon known as polymorphic switching (Maki-Yonekura *et al.*, 2010).

Flagellar filaments increase in length by the addition of flagellin molecules to the distal tip of the filament (Emerson *et al.*, 1970). Flagellin, the major protein component of flagellar filaments has a boomerang or an upper case Greek gamma (Γ)

shaped structure (Samatey *et al.*, 2001; Yonekura *et al.*, 2003) (Figure 1.3). A pentagonal flagellar capping protein, FliD is found at the elongating tip of the filament and directs the addition of each new flagellin subunit to the elongating filament in a helical manner (Yonekura *et al.*, 2000).

The flagellar filament may be modified by post-translational modifications of flagellin proteins or by the incorporation of a sheath over the filament structure. Glycosylation of flagellin proteins has been well described and several examples of glycosylated flagellins are available from both Gram positive and Gram negative bacteria (Logan, 2009). Moreover, *S. typhimurium* flagellins may become methylated at lysine residues through the action of FliB, a flagellin specific N-methylase (Burnens *et al.*, 1997; Tronick & Martinez, 1971). Phosphorylation of *P. aeruginosa* and *Campylobacter* flagellin proteins has also been reported (Kelly-Wintenberg *et al.*, 1993; Logan *et al.*, 1989). *H. pylori* (Geis *et al.*, 1993), *Brucella melitensis* (Ferooz & Letesson, 2010) and *V. cholerae* (Yoon & Mekalanos, 2008) have sheathed flagella. The sheath is a membranous covering that surrounds the flagellar filament. It is thought that the sheath limits flagellin shedding, and in doing so, restricts immune responses to flagellins *in vivo* (Yoon & Mekalanos, 2008).

1.2.1.5 *Regulation of flagellum biogenesis.*

Assembly of the bacterial flagellum is hierarchically regulated at several levels. A series of sigma factors and regulators coordinate and control transcription to ensure that the genes required for the assembly of the basal-body and hook structures are transcribed before those of the filament (Brown *et al.*, 2009; Smith & Hoover, 2009). On the basis of their position in the transcriptional hierarchy, flagellum biogenesis genes have been described as either “early” or “late” and may also have been assigned to a particular “class”. For instance, three and four classes of flagellum

biosynthesis genes have been identified in *S. typhimurium* and *Pseudomonas aeruginosa* respectively (Chilcott & Hughes, 2000; Dasgupta *et al.*, 2003). Each gene class may be under the control of a different sigma factor and/or set of regulators (Brown *et al.*, 2009; Smith & Hoover, 2009). The transition from one class to another is coupled to flagellum assembly (Brown *et al.*, 2009). Specific intermediate stages in flagellum synthesis serve as checkpoints, which once completed, promote the transcription of the next class of genes to further advance flagellum construction (Brown *et al.*, 2009; Chilcott & Hughes, 2000).

Flagellum biosynthesis often requires the expression of a master regulator, of which there are many (Smith & Hoover, 2009), to initiate expression of the subsequent gene classes. This master regulator may respond to environmental and growth-dependent signals to stimulate motility gene expression. However, master regulators and the molecular triggers that initiate flagellum synthesis have not been identified for every flagellate bacterial species. The appreciable diversity among the known master regulators of flagellum biogenesis and the lack of mechanistic molecular data for the control of many more flagellar systems is indicative of the complexity and variety of options for the regulation of flagellum biogenesis.

In several systems, it is the housekeeping sigma factors that regulate either the transcription of the master regulators or of the class I and II genes. In *S. typhimurium* for example, it is a heteromultimeric complex of master regulators FlhD and FlhC ($FlhD_4C_2$) that activates σ^{70} directed transcription from class II promoters (Chilcott & Hughes, 2000); while in *H. pylori* the housekeeping sigma factor, σ^{80}, controls expression of class I genes, which are thought to be constitutively expressed in this species (Niehus *et al.*, 2004). In *S. typhimurium*, σ^{28} (the sigma-factor that is responsible for transcription of the late and filament genes) is inactive while it is

bound to the anti-sigma factor FlgM. This sequestration and inactivation of σ^{28} is enforced until hook assembly is completed and FlgM can be exported. This releases the inhibition on σ^{28} and permits transcription from σ^{28} target promoters (Smith & Hoover, 2009). In *H. pylori*, however, FlgM serves its regulatory functions without being secreted (Rust *et al.*, 2009), emphasising the functional diversity that exists in bacterial flagellar systems.

Some regulators repress rather than activate the early stages of flagellum biogenesis. For example, the extracytoplasmic function (ECF) sigma factor RpoE1 inhibits flagellum biogenesis in *B. melitensis* by acting upstream of the master regulator, *ftcR*, in this species (Ferooz *et al.*, 2011). Moreover, in *Listeria monocytogenes*, MogR represses transcription of the initial flagellar genes to inhibit flagellation and motility during growth at 37 °C (Shen & Higgins, 2006).

1.2.1.6 *Phylogenetic and environmental distribution of motile bacteria.*

The trait of flagellum mediated motility occurs in many bacterial genera, but may be unevenly distributed within these lineages (Snyder *et al.*, 2009). Flagellum assembly places a significant energetic burden on the bacterium (Macnab, 1996), so the benefits of being flagellate and motile must justify the associated energy expenditure. While it may be generally advantageous for a bacterium to be motile, the flagella of the motile bacteria that colonise humans, animals, plants and other organisms may also represent an obstacle to colonisation. Flagellin is universally recognised by host receptors in the animal and plant kingdoms (Gomez-Gomez *et al.*, 2001; Hayashi *et al.*, 2001). Recognition of flagellin by the host triggers pro-inflammatory responses aimed at eliminating the offending organism (Tallant *et al.*, 2004). Accordingly, there is selection against flagellate microbes in these *in vivo* niches.

Characterization of motility genes and the availability of genome sequences for many motile bacteria has resulted in “cell motility” having its own clusters of orthologous groups (COG) category; namely “N” (Tatusov *et al.*, 2003). COGs are collections of individual orthologous proteins or of orthologous sets of paralogous proteins from at least three different phylogenetic lineages (Tatusov *et al.*, 1997). It is assumed that each member sequence of a COG was derived from a common ancestral sequence by speciation and duplication events. Thus, the sequences assigned to a particular COG are assumed to share the same function (Tatusov *et al.*, 1997) and COGs assigned to a particular category assume similar roles. For example, the cell motility COG category includes COGs for bacterial and archael flagellum biogenesis in addition to COGs for the assembly of type IV pili. COGs and also KEGG databases have been routinely used to assess the functions encoded by a genome or in a microbiome (Turnbaugh *et al.*, 2009). Accordingly, an assessment of the capacity of various microbial communities for motility has been indirectly built into most metagenome projects to date.

Compared to the proportional abundance of the other KEGG and COG categories such as “carbohydrate transport and metabolism” (G) and “translation, ribosomal structure and biogenesis” (J) which represent housekeeping or metabolic functions that are encoded in all bacterial genomes, COG category N is usually represented at low abundance in metagenome datasets (Kurokawa *et al.*, 2007). The cell motility KEGG was found to be variably represented in the core microbiome of lean and obese twins (Turnbaugh *et al.*, 2009).

In spite of its generally low representation in intestinal and faecal microbial communities, flagellum-mediated motility nevertheless remains a physiologically relevant bacteriological trait. The biological significance and consequences of

bearing flagella are dependent on the selective pressures imposed on motile bacteria by their environment, which varies from niche to niche.

1.2.1.7 *The interaction of motile bacteria with a mammalian host.*

1.2.1.7.1 *Immune responses of mammals to flagellate bacteria.*

The innate immune system of humans includes several PRRs that recognise and respond to flagellin proteins. The expression and cellular localisation of these receptors varies depending on the cell and tissue type as mentioned previously. The reactivity of the host to flagellin proteins is determined by host and bacterial factors. In particular, the responsiveness of TLR5 to flagellin proteins varies in line with TLR5 receptor sequence differences between different host species, and with respect to the bacterial origin of the prokaryotic flagellin (Andersen-Nissen *et al.*, 2005; Andersen-Nissen *et al.*, 2007). This is an important consideration for TLR5 signalling experiments involving model organisms.

1.2.1.7.2 *Tissue and cell based expression profile of human TLR5.*

An investigation of the tissue-specific expression profiles of the human TLRs by real-time PCR, revealed that TLR5 is expressed in several organs and tissues, with highest relative expression in the trachea, placenta, salivary gland and lung (Nishimura & Naito, 2005). A tissue specific gene expression profile for TLR5 has also been inferred from an analysis of publically available expressed sequence tag information (ESTs) and cDNA libraries (UniGene) (Table 1.4). TLR5 is expressed by peripheral blood monocytes, neutrophils, macrophages, dendritic cells (Muzio *et al.*, 2000; O'Mahony *et al.*, 2008) and also occurs on the surfaces of epithelial cells (Adamo *et al.*, 2004; Gewirtz *et al.*, 2001).

It is generally accepted that in the colon, TLR5 is basolaterally expressed (Gewirtz *et al.*, 2001) and that the pro-inflammatory immune response to flagellin

proteins requires basolateral rather than apical flagellin exposure (Rhee *et al.*, 2005). In apparent contravention of the traditional understanding of the TLR5-flagellin interaction however, apical expression of TLR5 has been reported *in vitro* for polarised human IECs (Bambou *et al.*, 2004; Eaves-Pyles *et al.*, 2011), and also for murine ileal enterocytes *in situ* (Bambou *et al.*, 2004). Furthermore, apically applied flagellin is now known to be internalised by a variety of human IECs in a TLR5 dependent manner, resulting in IL-8 secretion and transepithelial migration of dendritic and polymorphonuclear cells *in vitro* (Eaves-Pyles *et al.*, 2011). In the intestinal environment, it is therefore conceivable that free flagellins on the luminal side of the epithelium may not need to breach epithelial barriers to stimulate IL-8 secretion, implying a potential role for flagellin signalling in the establishment of homeostasis in the intestine. Further *in vivo* research to demonstrate the apical expression of TLR5 in the intestine and flagellin induced signalling through these apically expressed receptors would be required to corroborate this hypothesis.

1.2.1.7.3 *Specific receptor and ligand residues influence the responsiveness of TLR5 to flagellin.*

Typical flagellin proteins are at least 250 residues long with conserved amino- and carboxy- termini and a variable intervening region (Beatson *et al.*, 2006). This means that flagellins generally have conserved D0-D1 domains (Figure 1.3) which are arranged in close proximity to each other. The flagellin D2-D3 domains are non-conserved and may vary in both sequence and length (Beatson *et al.*, 2006).

Pool Name	TPM	Gene EST Total EST	Colour Scale
Adipose Tissue	0	0/12865	
Adrenal Gland	0	0/32935	
Ascites	0	0/39833	
Bladder	0	0/29858	
Blood	0	0/122244	
Bone	0	0/71614	
Bone Marrow	0	0/48724	
Brain	9	10/1092676	
Cervix	0	0/48477	
Connective Tissue	13	2/149066	
Ear	0	0/16100	
Embryonic tissue	0	0/212899	
Esophagus	0	0/20152	
Eye	0	0/208833	
Heart	0	0/89528	
Intestine	0	0/232028	
Kidney	4	1/210773	
Larynx	0	0/23492	
Liver	4	1/205263	
Lung	8	3/334805	
Lymph	0	0/44301	
Lymph Node	0	0/89729	
Mammary Gland	26	4/151237	
Mouth	75	5/66158	
Muscle	0	0/106370	
Nerve	0	0/15536	
Ovary	9	1/101490	
Pancreas	4	1/213426	
Parathyroid	0	0/20588	
Pharynx	0	0/40723	
Pituitary Gland	0	0/16525	
Placenta	17	5/282989	
Prostate	0	0/189536	
Salivary Gland	0	0/20263	
Skin	4	1/210749	
Spleen	0	0/53391	
Stomach	0	0/95793	
Testis	11	5/435194	
Thymus	0	0/79682	
Thyroid	0	0/46582	
Tonsil	0	0/17020	
Trachea	19	1/51780	
Umbilical cord	0	0/13765	
Uterus	21	5/232089	
Vascular	0	0/51649	

Table 1.4: Profile of human TLR5 gene expression in various body sites as inferred from EST and cDNA library sources. TPM = Transcripts per million. The colour scale is based on TPM and reflects the level of gene expression. Image prepared from Unigene, accession Hs.604542. Correct as of 20th Nov 2012.

Specific residues at the convex surface of flagellin monomers are involved in both flagellin polymerisation and in the recognition of flagellin by TLR5 (Andersen-Nissen *et al.*, 2005; Smith *et al.*, 2003). Experimental manipulation which mutates residues at this site, or which alter the tertiary structure of the convex surface, may abrogate either TLR5 recognition of the modified protein and/or the motility of the mutant strain (Smith *et al.*, 2003). Indeed, some species of the α- and ε-proteobacteria have evolved to evade TLR5 mediated immune responses by substituting a number of flagellin residues within this critical region (Andersen-Nissen *et al.*, 2005). These bacteria nevertheless remain motile due to compensatory mutations within the

flagellin structure that facilitate flagellin polymerisation. Thus, for some flagellate bacteria, the evolution of flagellin proteins that preserve motility without triggering TLR5 mediated pro-inflammatory immune responses is the outcome of a series of sequence and structural modifications to their flagellin proteins (Andersen-Nissen *et al.*, 2005).

Variations in the TLR5 sequence may also impact on flagellin recognition. In fact, the catalogue of flagellin proteins recognised by TLR5 varies between eukaryotic host species. Natural variation at TLR5 residue 268 (reference position for murine TLR5) imparts a specific flagellin recognition profile on the host (Andersen-Nissen *et al.*, 2007). Murine TLR5 (mTLR5) is more sensitive to flagellin proteins than the human receptor. For example, mTLR5 responded to purified flagellin proteins from *E. coli, S. typhimurium, P. aeruginosa* and *L. monocytogenes* at lower doses than human TLR5 (hTLR5), while hTLR5 was more sensitive than mTLR5 to *S. marcescens* flagellin (Andersen-Nissen *et al.*, 2007). Mouse and human TLR5 encode proline and alanine residues respectively, at position 268. It was possible to change the specificity of mTLR5 to that of hTLR5 by mutating the proline to alanine at position 268. The reciprocal mutation in the hTLR5 receptor resulted in reduced sensitivity to flagellin. Additionally, mutation of mTLR5 residues D295 and D367 decreased the sensitivity and weakened recognition of mutated flagellins respectively (Andersen-Nissen *et al.*, 2007). In cell lines expressing TLR5 from various animals and birds, *S. typhimurium* flagellin exerted a greater response via chicken TLR5 than via hTLR5, as determined by activation of NF-κB. There was no difference between the responsiveness of chicken TLR5 and mTLR5 to *S. typhimurium* flagellin protein. Interestingly, a species specific difference was not observed among the different receptors following exposure to *S. enteritidis* flagellin (Keestra *et al.*, 2008).

1.2.1.8 *Structural basis for flagellin TLR5 signalling.*

The structures of human and zebrafish (*Danio rerio*) TLR5 were recently elucidated, providing insight into the molecular basis for flagellin-TLR5 interactions (Yoon *et al.*, 2012; Zhou *et al.*, 2012). The hTLR5 structure was determined in the absence of its ligand by homology modelling and electron microscopy image reconstruction (Zhou *et al.*, 2012), while the crystal structure of the zebrafish receptor (drTLR5) was solved in a complex with flagellin (Yoon *et al.*, 2012). In the absence of its ligand, hTLR5 was found to exist as an asymmetric homodimer, and this structure was proposed to be optimised for the cooperative binding of two flagellin molecules (Zhou *et al.*, 2012). Indeed, one flagellin molecule was found to interact with each part of the drTLR5 homodimer, so that two flagellin molecules are bound by each TLR5 complex (Yoon *et al.*, 2012).

The LRR9 loop found in the TLR5 ectodomain provides the major flagellin binding site (Yoon *et al.*, 2012). Flagellin residues that are known to be critical for TLR5 recognition, including Arg^{90} and Glu^{114}, were shown to interact with residues of the TLR5 structure. In particular, drTLR5 residue 271 (equivalent to mouse and human TLR5 residue 268), interacted with flagellin residue Arg^{90} (Yoon *et al.*, 2012). The asparagine residues at mTLR positions 294 and 366 which were shown to influence flagellin recognition (Andersen-Nissen *et al.*, 2007), were found not to be directly involved in flagellin binding, but rather were proposed to impact on TLR5 folding and stability (Yoon *et al.*, 2012).

1.2.1.9 *The TLR5 signalling pathway.*

TLR5 signalling induces pro-inflammatory gene transcription in response to monomeric flagellin proteins (Tallant *et al.*, 2004). This pro-inflammatory signalling is initiated by the recruitment of an adaptor molecule, usually MyD88 to the TIR

domain of the TLR5 receptor (Tallant *et al.*, 2004). It has been shown recently that TIR domain containing adaptor-inducing interferon-β (TRIF) may also serve as an adaptor molecule for TLR5 (Choi *et al.*, 2010). Following the initial recruitment of MyD88 to TLR5, signal propagation proceeds via a network of intermediary molecules and complexes which ultimately lead to the activation of NF-κB and mitogen activated protein kinase (MAPK) signalling pathways and pro-inflammatory gene transcription in the nucleus (Figure 1.4).

1.2.1.9.1 *Consequences of TLR5 receptor polymorphisms for flagellin-induced signalling.*

Some individuals harbour a non-functional TLR5 receptor as a consequence of a single nucleotide polymorphism C1174T, that introduces a premature stop-codon into the TLR5 coding sequence at residue 392 (Hawn *et al.*, 2003). This mutation was predicted to yield a receptor with a truncated ectodomain. Individuals that are heterozygous for this TLR5 allele have an impaired ability to secrete pro-inflammatory cytokines in response to flagellin proteins, demonstrating that this receptor polymorphism acts in a dominant negative fashion (Hawn *et al.*, 2003).

Individuals that harbour the $TLR5^{392STOP}$ receptor variant are more susceptible to infections with flagellate microbes and are less susceptible to inflammatory diseases in which exposure to microbial flagellin proteins has a significant contributory effect. For example, individuals with the truncated receptor are more susceptible to Legionnaire's disease (Hawn *et al.*, 2003) but are less susceptible to systemic lupus erythematosus (Hawn *et al.*, 2005) and Crohn's disease (Gewirtz *et al.*, 2006).

1.2.1.10 *Signalling via intracellular flagellin receptors.*

NLR family, apoptosis inhibitory proteins (NAIPs), are a subset of NLRs that are activated by interaction with MAMPs, such as flagellin. Activation of NAIPs

results in the formation of a multi-protein, immune complex that is known as an inflammasome, and which plays a role in cytoplasmic immunosurveillance. NAIPs may be activated either by a direct interaction with a specific ligand or indirectly, by sensing the stress response of the host cell to infection (Kofoed & Vance, 2012). Inflammasomes activate caspase I which leads to cytokine production and cell death by pyroptosis (Kofoed & Vance, 2012).

NAIP5 is a sensor of cytosolic flagellin that works in concert with NLR family, CARD domain containing (NLRC) protein 4, a molecule formerly known as ICE-protease activating factor (IPAF) that was also recognised as an intracellular flagellin sensor (Kofoed & Vance, 2012). Unlike NAIP5, NLRC4 includes a caspase activation and recruitment domain (CARD) in its structure, which facilitates cleavage and activation of caspase I. Although the flagellin induced inflammasome is known to include flagellin, NAIP5 and NLRC4, the mechanism of its assembly remains speculative (Kofoed & Vance, 2012). Co-immunoprecipitation experiments revealed that flagellin interacts directly with these NLR proteins (Kofoed & Vance, 2011).

MAMPs must gain access to the cytoplasm to activate an intracellular PRR. Pathogens that adopt an intracellular life-style, such as *Legionella pneumophila* (Vinzing *et al.*, 2008), or those that are capable of translocating MAMPs into the cytoplasm of host cells, like *Salmonella* (Sun *et al.*, 2007b), present a microbial threat that requires an urgent immune response. The cytoplasmic PRRs thus serve as immune sentinels which respond to serious microbial threats when the barriers that separate the extracellular from the intracellular environment are compromised.

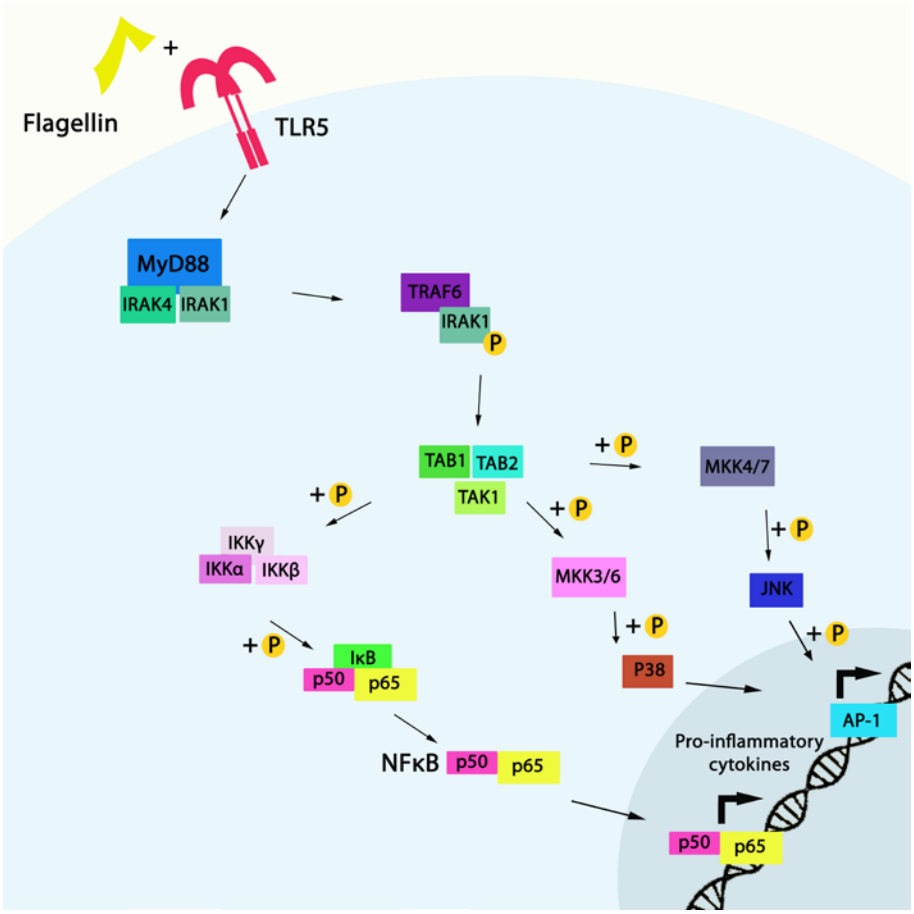


Figure 1.4: Flagellin induced TLR5 signalling pathway.

TLR5 homodimers on the cell surface recognise and respond to flagellin monomers. Upon ligand binding, the MyD88 adaptor molecule interacts with the TIR domain of the receptor leading to the recruitment of Interleukin-1 Receptor-associated kinase (IRAK) molecules 4 and 1. Phosphorylated IRAK1 and tumour necrosis factor receptor associated factor (TRAF) 6 dissociate from the complex with MyD88. TRAF6 associates with TGF-beta activated kinase (TAK) 1 and TAK1 binding protein (TAB) 1 and 2. A series of phosphorylation and ubiquitination steps ensue, leading to activation of either the NF-κB or the mitogen-activated protein kinase (MAPK) signalling pathways. Phosphorylation of IκB (inhibitor of NF-κB) results in its degradation, releasing NF-κB, which is free to translocate to the nucleus where it initiates pro-inflammatory gene transcription. Phosphorylation of mitogen-activated protein kinase kinases (MKKs) leads to phosphorylation of c-Jun kinase (JNK) and P38 and the initiation of pro-inflammatory gene transcription in the nucleus.

Prepared with reference to KEGG pathway map hsa04620 (www.genome.jp); www.sabiosciences.com/pathway/php?sn=Toll_Like_Receptors; (O'Neill *et al.*, 2009; Takeda & Akira, 2004).

1.2.1.11 *Strategies for immune evasion by motile bacteria.*

Some pathogenic bacteria, for which flagella and motility are essential for niche colonisation and virulence, have evolved to evade the immune responses of their host. In general terms, this is achieved by limiting the release and exposure of flagellins to the host PRRs, or by adopting flagellin modifications that avoid triggering severe immune responses.

Listeria monocytogenes strains demonstrate differential motility depending on the temperature of their immediate environment. While *L. monocytogenes* strains are vigorously motile and flagellate at temperatures close to 20 °C, it is generally agreed that motility and flagella expression is repressed at temperatures at or above 37 °C (Peel *et al.*, 1988). Nevertheless, a laboratory strain (1 of 6 tested) and 20 % (5 of 25 tested) of clinical *L. monocytogenes* isolates were found to trigger signalling through TLR5, which corresponded to flagellin expression at 37 °C (Way *et al.*, 2004). A motility gene repressor MogR, regulates the expression of *L. monocytogenes* flagellum biogenesis genes (Grundling *et al.*, 2004). MogR expression affects the virulence of *L. monocytogenes* strains. Mutant strains lacking MogR were 250 times more virulent than the wild-type (Grundling *et al.*, 2004).

The production of sheathed flagella limits the dissemination of flagellin monomers (Yoon & Mekalanos, 2008). This is important because flagellin monomers are recognised by TLR5. The *Helicobacter* and *Vibrio* species that produce sheathed flagella are therefore less likely to shed flagellin molecules that could potentially trigger immune responses (Yoon & Mekalanos, 2008). The flagellar sheath that surrounds the hook and filament of *Helicobacter* and *Vibrio* species is continuous with the outer membrane (Chen *et al.*, 2011) and is composed of lipopolysaccharide and protein (Geis *et al.*, 1993). Thus, sheathed flagella may be more stable and less

likely to trigger TLR5 mediated immune responses *in vivo.* It has also been suggested that flagellin phase variation, such as that which occurs in *Salmonella typhimurium* may allow this bacterium to temporally evade immune recognition (Fierer & Guiney, 2001; Zeng *et al.*, 2003).

Although some bacteria have evolved to successfully evade detection by TLR5, it must be remembered that the host harbours additional PRRs that are capable of recognising and mounting an immune response to other MAMPs, such as LPS, peptidoglycan and prokaryotic nucleic acids. Thus, a flagellate bacterium may trigger signalling through more than one receptor. This functional redundancy in PRR signalling is an integral component of the ability of the host to ward off pathogenic infections. Nevertheless, flagellins often serve as immunodominant antigens that are the major stimulants of immune response to flagellate microbes (Duck *et al.*, 2007; Lodes *et al.*, 2004).

1.2.1.12 *The motile lactobacilli - a historical perspective.*

The genus *Lactobacillus* contains more than one hundred species, ~ 10 % of which are motile or harbour genes for flagellum biogenesis. The biological significance of this trait has been poorly acknowledged in the scientific literature to date. Observations of motility among species described as lactobacilli have been reported since the mid 20th century (Cunningham & Smith, 1939; Deibel & Niven Jr., 1958; Fred *et al.*, 1919; Harrison Jr & Arne Hansen, 1950; Vankova, 1957). The identification of these species as lactobacilli relied on culture techniques and phenotypic assessments; methods which may have led to incorrect species identification. Accordingly, several of these motile lactobacilli have since been reclassified to another genus or as another species (Gemmel & Hodgkiss, 1964). For contempory species assignment, phenotypic tests are supported by molecular data and

often whole genome sequences, which may better resolve differences between species with overlapping phenotypes.

L. salivarius, L. acidophilus, L. fermentum, L. rhamnosus and *L. casei* have all been described as flagellate and motile in the scientific literature (Gemmel & Hodgkiss, 1964; McGroarty, 1994). Although motility could be a strain-specific trait, the genomes available for these species have confirmed that flagellum biogenesis genes are not present. In retrospect, it seems likely that the methods used to assign these motile strains to a particular species were unreliable.

The motile, flagellate lactobacilli of reliable taxonomy belong primarily to the *L. salivarius* phylogenetic clade, and most were first described only recently (Table S2.1, Chapter 2). Nonetheless, motile strains of the species *L. curvatus* and *L. sakei* subsp. *carnosus* have also been described (Koort *et al.*, 2004; Torriani *et al.*, 1996) and these belong to the *L. sakei* clade (Felis & Dellaglio, 2007). These latter strains are of particular interest, because the genomes available for *L. curvatus* and *L. sakei* would indicate that these are non-motile and aflagellate species. Interestingly, gliding motility (motility without specialised organelles) has been described for *L. koreensis*, a kimchi isolate that is a member of the *L. parabrevis* clade (Bui *et al.*, 2011).

To date, the majority of the motile lactobacilli have been isolated from environmental sources and in particular, from foods and alcoholic beverages in which fermentation is part of the manufacturing process. *Lactobacillus ruminis* is the only motile species that has been frequently recovered from mammals including humans, pigs, horses and cows (Al Jassim, 2003; Sharpe *et al.*, 1973; Tannock *et al.*, 2000; Willing *et al.*, 2009).

1.2.1.13 L. ruminis – *history of a motile species.*

The nomenclature "*L. ruminis"* was first used to describe a motile, anaerobic, non-spore forming, homofermentative, Gram-positive, rod shaped bacterium that was recovered at high density from the bovine rumen (Sharpe *et al.*, 1973). Moreover, *Catenabacterium catenaforme*, an anaerobic species that was isolated from the human GI tract (described in (Reuter, 2001)) was later reclassified as *L. ruminis* (Hammes & Hertel, 2006; Reuter, 2001; Sharpe & Dellaglio, 1977).

The *Lactobacillus* motility phenotype was considered atypical of this genus when it was first observed, impelling requests for the validity of motility as a *Lactobacillus* trait to be "cleared up" (Mann & Oxford, 1954). Thus, despite several early publications describing the potential for motility among presumptive *Lactobacillus* isolates (Cunningham & Smith, 1939; Deibel & Niven Jr., 1958; Harrison Jr & Arne Hansen, 1950), flagellum mediated motility remained an unusual and poorly acknowledged *Lactobacillus* trait until recent times.

The motility phenotype of *L. ruminis* is curious because of its apparent strain variable expression. While *L. ruminis* strains of animal origin are often motile, those recovered from humans and sewage are not (Hammes & Vogel, 1995; Kandler & Weiss, 1986; Reuter, 2001; Sharpe *et al.*, 1973). Traditionally, the motility of *Lactobacillus* species such as *L. ruminis*, has been sluggish and difficult to demonstrate (Sharpe *et al.*, 1973). Moreover, motile strains of some *Lactobacillus* species lose their motility upon repeated subculture in a laboratory setting (Torriani *et al.*, 1996). Thus, for many years, the trait of motility in the genus *Lactobacillus* was superficially described and was limited to an acknowledgement of the presence of flagella or of detectable phenotypic motility (Gemmel & Hodgkiss, 1964; McGroarty, 1994; Sharpe *et al.*, 1973).

1.2.1.14 L. ruminis *as an autochthonous species of the mammalian GI tract.*

Autochthonous species are particular to a given niche. The autochthonous microorganisms that inhabit the GI tract may be defined as "… indigenous microorganisms that colonize particular regions of the tract early in life, multiply to high population levels soon after colonization … remain at those levels throughout the lives of healthy well-nourished animals … (and) should be found in essentially all individuals of a given animal species, irrespective of their geographical location" (Savage, 1972). *L. ruminis* is one of the few *Lactobacillus* species considered autochthonous to the mammalian intestines (Tannock *et al.*, 2000; Tannock, 2004), a fact that is supported by the frequent isolation of this organism from several mammalian hosts.

The previously described 18 month study involving ten healthy New Zealanders was among the first to demonstrate that *L. ruminis* was stably present in the human GI tract over time (Tannock *et al.*, 2000). A meta-analysis describing earlier research on the *Lactobacillus* composition and succession in the human GI tract, reported that *L. ruminis* was recovered from the stomach and the intestines of human adults and from the intestines of children aged 6 - 12 years (Reuter, 2001). A culture independent study of the faecal *Lactobacillus* community of 13 healthy adults from the Netherlands, similarly found that *L. ruminis*-like sequences were commonly recovered from the faecal and cecal DNA libraries, and that *L. ruminis* was found to dominate the *Lactobacillus* community of the sampled sites in two out of three of these adults over a period of twenty months as assessed by monitoring DGGE fingerprints (Heilig *et al.*, 2002). *L. ruminis* was recovered from one of two individuals that took part in a Spanish study that aimed to investigate the daily variation in the "most prominent and representative of the cultivable microbial

populations", particularly of the lactic acid bacteria, present in human faeces over a fifteen day period (Delgado *et al.*, 2004). Incidental isolations of *L. ruminis* from human faeces have also been reported (Kagkli *et al.*, 2007; Taweechotipatr *et al.*, 2009; Yun *et al.*, 2005).

Several studies on animals lend similar support for the establishment of *L. ruminis* as an autochthonous species of the mammalian intestine. *L. ruminis* was first isolated from the bovine rumen (Sharpe *et al.*, 1973), and 16S rRNA gene restriction fragment length polymorphism (RFLP) analysis and sequencing have identified this species as a dominant member of the *Lactobacillus* community of cows (Krause *et al.*, 2003).

Several reports have described the isolation of *L. ruminis* from the porcine GI tract, in which it has been the dominant lactic acid bacterium present (Al Jassim, 2003; Yin & Zheng, 2005). In fact, 8 of 13 isolates that were among 60 pure cultures recovered from the intestines of 6 healthy pigs post-mortem were identified as *L. ruminis* by morphological analyses and 16S rRNA gene sequencing. At least one *L. ruminis* strain was recovered from each pig (Al Jassim, 2003). Another study identified 10 *L. ruminis* isolates as part of a panel of 52 *Lactobacillus*-like isolates that were a subset of all the colonies recovered from the either the gut or faeces of five pigs (Yin & Zheng, 2005). Remarkably, in both of these studies, *L. ruminis* was the only *Lactobacillus* species detected. However, *L. ruminis* does not always become dominant in the *Lactobacillus* community of the porcine GI tract. In a comparative 16S rRNA gene analysis of the phylogenetic diversity of the luminal GI microbiota of 24 pigs of varying age, health and dietary status, *L. ruminis* was represented in the library by a single clone (Leser *et al.*, 2002). A total of 4,270 clones, representing 375 phylotypes constituted the library. A subset of twenty-four

OTUs represented the *Lactobacillus* species present in the porcine gut. Twenty-two of these twenty-four *Lactobacillus* OTUs represented lactobacilli of known identity, and *L. amylovorus* (191 clones), *L. johnsonii* (138 clones) and *L. reuteri* (91 clones) were most abundantly represented in the clone library (Leser *et al.*, 2002).

Species of the *L. salivarius* clade form a significant component of the equine GI microbiota (Morita *et al.*, 2009; Willing *et al.*, 2009). Swarming *L. ruminis* (11 isolates) was the dominant *Lactobacillus* species identified during screening of 40 colonies isolated from healthy thoroughbreds on a forage-concentrate diet. Other species of the *L. salivarius* clade identified included *L. salivarius* (5 isolates), *L. equi* (3 isolates) and *L. agilis* (1 isolate). This is consistent with an earlier study which recovered *L. salivarius* clade species, specifically *L. equi* (4/37 isolates), *L. salivarius* (3/37 isolates), *L. ruminis* (3/37 isolates), *L. agilis* (2/37 isolates), *L. hayakitensis* (1/37 isolates) and *L. saerimneri* (1/37 isolates) from equine faeces (Morita *et al.*, 2009). However, *L. hayakitensis*, *L. equigenerosi* and *L. equi* rather than *L. ruminis*, were identified as the prevalent *Lactobacillus* species present in the faeces of these nine horses by PCR-DGGE analysis and sequencing (Morita *et al.*, 2009).

L. ruminis may also dominate the GI microbiota of domestic cats (Desai *et al.*, 2009; Ritchie *et al.*, 2010). Although only 2 of 12 cats in one study were found to harbour *L. ruminis*, this species represented 21.95 % of the 364 clones in a 16S rRNA gene fragment library derived from the faeces of the 12 animals (Ritchie *et al.*, 2010). *L. ruminis* was found to dominate the faecal microbiotas of four outdoor cats, representing 63 % (736/1166 high-quality, sequenced chaperonin *cpn60* clones) of the "outdoor" clone library (Desai *et al.*, 2009). *L. ruminis* occurred at a lower frequency (46/1075 high-quality, sequenced clones) in the "indoor" clone library from five indoor cats faecal microbiota (Desai *et al.*, 2009).

L. ruminis therefore has a broad host-range that includes but is not limited to, humans and domesticated mammals. The identification and/or isolation of *L. ruminis* from the intestinal and faecal microbiotas of these animals appear to vary considerably from study to study. This variation may reflect different experimental approaches for the isolation and detection of the dominant and persistent species of a given microbial community, as well as for species assignment. Furthermore, local variation in diet and environment may impact on the abundance of these species in the intestinal microbiotas of target animals.

The enteric *Lactobacillus* population is a subdominant component of the human GI microbiota (Tannock *et al.*, 2000). Because *L. ruminis* constitutes only a fraction of the enteric *Lactobacillus* community, it must therefore be considered as a minority species in the gut.

1.2.1.15 *The predominant motile commensal bacteria of the human GI tract.*

The low GC% Gram positive species of the phylum *Firmicutes*, represent a major component of the GI microbiota of humans and animals (Arumugam *et al.*, 2011; Claesson *et al.*, 2011; Ley *et al.*, 2008). A number of lineages within this phylum include at least some species that are motile by means of flagella. The trait of flagellum-mediated motility is widespread in the genera *Bacillus* and *Clostridium*, but it is less common in the genus *Lactobacillus*. It would seem likely that some of the most abundant motile species present in the gut would belong to the phylum *Firmicutes*. Indeed, as may be inferred from box-plot graphs presented by Qin and colleagues, of the 56 bacterial species whose genomes were present at greater than 1 % coverage in at least 90 % of individuals in a cohort of 124 people, at least 30 belonged to the phylum *Firmicutes* and at least four of these species were known to be motile by means of flagella (Qin *et al.*, 2010). The species in question were

Eubacterium rectale (present in 123 individuals), *Eubacterium siraeum* (119 individuals), *Roseburia intestinalis* (124 individuals) and *Butyrivibrio crossotus* (123 individuals). These species and phylogenetically close taxa have also been identified as abundant in other studies of the human intestinal microbiota (Aminov *et al.*, 2006; Claesson *et al.*, 2011).

The functions of flagellum biogenesis and chemotaxis typically occur as minority or specialist functions among bacteria (Kurokawa *et al.*, 2007; Qin *et al.*, 2010; Turnbaugh *et al.*, 2009). Given their limited distribution, these functions are rarely represented at high proportions in the microbial gene catalogue of the intestinal microbiome. For example, the "cell motility" COG category was depleted in the collective gut microbiomes of Japanese adults, children and unweaned infants (13 individuals in total) (Kurokawa *et al.*, 2007). This COG category was identified as under-represented in the human gut microbiome by comparison to a non-redundant reference database of 243 microbial genomes. That non-redundant database included only one representative genome per species and also excluded the genomes of any microbes that were known to inhabit the human gut, so that any functional categories over- or under- represented in the human gut microbiome could be distinguished. Having observed a striking paucity of genes for flagellum biogenesis and chemotaxis, the authors surmised that these functions were not required for the persistence of microorganisms in an intestinal environment, and suggest that the loss of motility genes reflects an evolutionary adaptation of the enteric bacteria to their niche (Kurokawa *et al.*, 2007). Furthermore, cell motility genes were found to be depleted in the intestinal microbiomes of obese versus lean twins, and KEGG pathway analysis revealed that cell motility genes were more common in the "variable" rather than the "core" human gut microbiome (Turnbaugh *et al.*, 2009).

A later functional categorisation of the human gut microbiome introduced the concepts of the "minimal gut genome" and "minimal gut metagenome" (Qin *et al.*, 2011). The former includes the genetic functions encoded by all gut microbes to thrive in this niche, while the latter attempts to describe the genetic functions required for homeostasis of the entire gut ecosystem. As can be deduced from the graphs included with the published article, cell motility was not an abundant function in either the minimal gut genome or the minimal gut metagenome of the gene catalogue derived from the microbiomes of 124 individuals (Qin *et al.*, 2011).

In keeping with the low abundance of motility genes in the intestinal microbiome as assessed by metagenomics, "cell motility" mRNA transcripts also represented a minor proportion of the intestinal metatranscriptome of 10 healthy adults (Gosalbes *et al.*, 2011). Species of the phylum *Firmicutes* contributed the greatest proportion of these motility transcripts (Gosalbes *et al.*, 2011). Likewise, a study of the faecal metaproteome of two individuals identified "cell motility proteins" at very low numbers (Verberkmoes *et al.*, 2008). However, a more recent study reported that flagellin proteins formed part of the core metaproteome of healthy adults, and were identified at an unusually high level in one individual, in whom flagellins represented ~ 4 % of all identified spectra (Kolmeder *et al.*, 2012).

1.2.2 The genera *Eubacterium* and *Roseburia.*

1.2.2.1 *Phylogeny.*

Species of the genera *Eubacterium* and *Roseburia* are members of the family *Clostridiales* and are specifically associated with the suprageneric phylogenetic *Clostridium* clusters XIVa (*Roseburia* species, *E. eligens, E. rectale*) and IV (*E. siraeum*) (Collins *et al.*, 1994). The genus *Eubacterium* is loosely defined and as a result, its taxonomy is heterogeneous and unsatisfactory (Wade, 2006). Many

"*Eubacterium"* species could be reclassified and better integrated into other novel or extant genera on the basis of their molecular and phenotypic properties. In fact, several species formerly considered as *Eubacterium* isolates have already been reclassified to the genera *Dorea*, *Mogibacterium*, *Collinsella* and *Pseudoramibacter*, among others (Euzeby, 1997).

In contrast, the genus *Roseburia* comprises fewer member species and is more homogeneous. At this time, five *Roseburia* species are recognised (Euzeby, 1997) (www.bacterio.cict.fr/qr/roseburia.html). *E. rectale* is more closely related to these *Roseburia* species than it is to the *Eubacterium* type species, *E. limosum* (Aminov *et al.*, 2006). On the basis of this phylogenetic observation and additional phenotypic assessments, it has been proposed that *E. rectale* should be reclassified as *Roseburia rectale* (Duncan *et al.*, 2006), but as yet, the classification and nomenclature of this species has not been formally revised.

1.2.2.2 Eubacterium *and* Roseburia *populations in the gut are influenced by host diet.*

Eubacterium and *Roseburia* species are responsive to host dietary carbohydrate consumption, with reductions in these populations reported when individuals consume a "high-protein, low carbohydrate" diet (Duncan *et al.*, 2007; Russell *et al.*, 2011; Walker *et al.*, 2011). *E. rectale* and *Roseburia* species ferment plant-derived oligo- and poly- saccharides such as inulin, resistant starch and xylan (Duncan *et al.*, 2006). These substrates could serve as prebiotics for *Roseburia* and *Eubacterium* species (Scott *et al.*, 2011). For example, populations of bacteria related to *E. rectale* and *Roseburia* species increased in individuals who consumed diets rich in resistant starch (Walker *et al.*, 2011). As butyrate producing species, the presence of these bacteria in the intestine is thought to be desirable for maintaining gut health,

and dietary interventions which would stabilise or maintain these populations of butyrate producing bacteria could be useful.

1.2.2.3 *Motility of* Eubacterium *and* Roseburia *species*.

While all of the *Roseburia* species described to date are motile and bear flagella (Duncan *et al.*, 2002; Duncan *et al.*, 2006; Stanton *et al.*, 2009), many *Eubacterium* species, including the type species *Eubacterium limosum*, are apparently non-motile (Wade, 2009). Motility among these species could however, be a strain dependent trait. Although some motile *Eubacterium* and *Roseburia* species are among the most abundant species in the human intestinal microbiota (Aminov *et al.*, 2006; Claesson *et al.*, 2011; Qin *et al.*, 2010), the flagellum-mediated motility of these species has not been well characterised. While there is a paucity of information describing the biological and ecological roles of motility among the *Eubacterium* species, the motility of certain *Roseburia* species has been studied and discussed.

Motility was advantageous for *Roseburia cecicola* GM to colonise the murine cecum in the presence of other microbial species, but it was not required for colonisation of germ-free animals (Stanton & Savage, 1983; Stanton & Savage, 1984). This species has 20-35 sub-polar flagella (Stanton & Savage, 1983), which appear to be composed of a single flagellin protein, reportedly ~ 42 kDa in size (Martin & Savage, 1985). A later publication estimated the molecular weight of *R. cecicola* GM flagellin to be ~31 kDa (Martin & Savage, 1988). This could imply that the flagellin protein of this species is post-translationally modified. The first ten residues at the amino terminus of the *R. cecicola* GM flagellin protein were deduced to be MVVQQNMTAM (Martin & Savage, 1988), which is consistent with the amino-terminal sequences of flagellins from other *Clostridium* XIVa species

(Kalmokoff *et al.*, 2000). A genome sequence, which would prove a useful resource for studying the genetic basis for the motility of this species, is not yet available.

If a correlation exists between flagellin expression levels and phenotypic motility, *R. inulinivorans* may be more motile when grown on starch than on inulin. Specifically, expression of flagellin by *R. inulinivorans* is known to be substrate dependent, and is greater when the bacterium is grown on starch rather than on maltose, inulin or glucose (Scott *et al.*, 2011). Although the mechanism responsible for this increase in flagellin expression during growth on starch is presently unknown, this phenomenon may reflect a biological adaptation which allows *R. inulinivorans* to out-compete other species *in vivo*. Starch, unlike inulin, is thought to be insoluble when it reaches the intestines. Thus, the insoluble starch may serve as a chemo-attractant for *R. inulinivorans.* The motility of this speices may therefore benefit the bacterium by affording it early access to these insoluble starch particles (Scott *et al.*, 2011).

1.2.2.4 *Immunomodulatory properties of* Eubacterium *and* Roseburia *species and the role of their flagellin proteins in disease.*

The immunomodulatory properties of enteric *Eubacterium* and *Roseburia* species could potentially be harnessed to promote gut health. The ability of conditioned media (CM) from *E. rectale, Roseburia faecis, Roseburia hominis* and *R. intestinalis* to modulate NF-κB in epithelial and monocyte-like cell lines was evaluated *in vitro* (Lakhdari *et al.*, 2011). When compared to NF-κB activity in unstimulated control cells, CM from several *Roseburia* species significantly increased NF-κB activity in two IEC lines, while CM from *E. rectale* ATCC 33656 did not. Furthermore, CM from these *Roseburia* cultures had a significant ($P < 0.05$) stimulatory effect on activated NF-κB in these cell lines, while CM from *E. rectale*

ATCC 33656 had a significant inhibitory effect on stimulated NF-κB activity in only one of the two cell lines tested. A correlation between butyrate production and stimulation of NF-κB activity by *Roseburia* species derived CM was observed, suggesting that butyrate production by these species contributed to the NF-κB stimulatory activity (Lakhdari *et al.*, 2011). It was also suggested, without confirmation, that the flagellin proteins of these motile species could contribute to the enhancement of NF-κB activation by these flagellate species. In a separate study, when compared to NF-κB expression in untreated cells, heat-killed *E. rectale* A1-86 significantly induced expression of an NF-κB reporter via TLR5, suggesting that the flagellin component of this species was indeed pro-inflammatory (Erridge *et al.*, 2010).

Flagellin-TLR5 interactions involving these and other *Clostridium* cluster XIVa species could have pathological significance. *Clostridium* cluster XIVa flagellin proteins may have a role in the development of Crohn's disease (Duck *et al.*, 2007; Lodes *et al.*, 2004). Furthermore, butyrate biosynthesis and cell motility genes were depleted in metagenomes of individuals with type 2 diabetes when compared to the metagenomes of healthy controls (Qin *et al.*, 2012). Sequences representing *R. intestinalis*, *R. inulinivorans* and *E. rectale* were also enriched in controls compared to diabetes patients. Butyrate synthesising bacteria were deduced to be major contributors of cell motility functions to the intestinal metagenome. This implies that the *Eubacterium* and *Roseburia* species that were enriched in the metagenomes of the control individuals could be major contributors to the functions of butyrate biosynthesis and cell motility in the metagenomes of these individuals. Moreover, an analysis of the intestinal gut metaproteome revealed that flagellin proteins were

generally associated with species of the *E. rectale* and *R. inulinivorans* groups (Kolmeder *et al.*, 2012).

1.3 Summary of thesis contents.

As described herein, motility is of significant biological relevance as a microbe-host interaction trait. In this thesis, flagellum-mediated motility has been studied in the context of the bacterial genera that form a commensal association with the mammalian GI tract, and particularly in humans.

Chapter Two describes the genomic, transcriptomic, microbiological and immunological characterisation of flagellum-mediated motility and flagellum expression in *L. ruminis*, the only motile *Lactobacillus* species currently known to be autochthonous to the human GI tract. Chapter Three describes the genomic basis for flagellum biogenesis and its potential regulation in motile *Roseburia* and *Eubacterium* species, including those that are abundantly present in the human GI microbiota, and also outlines the inflammatory potential of the native flagellin proteins of these species.

1.3.1 Aims and Objectives.

- To characterise the flagellum-biogenesis genotype and motility phenotype of motile lactobacilli, with a particular focus on *Lactobacillus ruminis*.
- To use a comparative genomics approach for the characterisation of the flagellum biogenesis genotype of motile enteric *Eubacterium* and *Roseburia* species.
- To propose a model for the regulation of flagellum biogenesis in *L. ruminis* and to identify potential regulators of motility gene expression in *Eubacterium* and *Roseburia* species *in silico.*
- To determine the immune response of the human host to the flagellin proteins of several *Lactobacillus* species, *E. rectale* and *R. inulinivorans*.

- To determine if complete functional pathways for a specific process, such as flagellum biogenesis, can be identified for a particular target species in metagenome databases, such as those generated by the ELDERMET project.

1.4 References

Adamo, R., Sokol, S., Soong, G., Gomez, M. I. & Prince, A. (2004). *Pseudomonas aeruginosa* flagella activate airway epithelial cells through asialoGM1 and toll-like receptor 2 as well as toll-like receptor 5. *Am J Respir Cell Mol Biol* **30**, 627-634.

Adlerberth, I. & Wold, A. E. (2009). Establishment of the gut microbiota in Western infants. *Acta Paediatr* **98**, 229-238.

Agrawal, A., Singh, P. P., Bottazzi, B., Garlanda, C. & Mantovani, A. (2009). Pattern recognition by pentraxins. *Adv Exp Med Biol* **653**, 98-116.

Ahmed, S., Macfarlane, G. T., Fite, A., McBain, A. J., Gilbert, P. & Macfarlane, S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* **73**, 7435-7442.

Ahrne, S., Lonnermark, E., Wold, A. E., Aberg, N., Hesselmar, B., Saalman, R., Strannegard, I. L., Molin, G. & Adlerberth, I. (2005). Lactobacilli in the intestinal microbiota of Swedish infants. *Microbes Infect* **7**, 1256-1262.

Al Jassim, R. A. (2003). *Lactobacillus ruminis* is a predominant lactic acid producing bacterium in the caecum and rectum of the pig. *Lett Appl Microbiol* **37**, 213-217.

Aminov, R. I., Walker, A. W., Duncan, S. H., Harmsen, H. J., Welling, G. W. & Flint, H. J. (2006). Molecular diversity, cultivation, and improved detection by fluorescent *in situ* hybridization of a dominant group of human gut bacteria related to *Roseburia* spp. or *Eubacterium rectale*. *Appl Environ Microbiol* **72**, 6371-6376.

Andersen-Nissen, E., Smith, K. D., Strobe, K. L., Barrett, S. L., Cookson, B. T., Logan, S. M. & Aderem, A. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc Natl Acad Sci U S A* **102**, 9247-9252.

Andersen-Nissen, E., Smith, K. D., Bonneau, R., Strong, R. K. & Aderem, A. (2007). A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. *J Exp Med* **204**, 393-403.

Anderson, J. K., Smith, T. G. & Hoover, T. R. (2009). Sense and sensibility: flagellum-mediated gene regulation. *Trends Microbiol* **18**, 30-37.

Arumugam, M., Raes, J., Pelletier, E. & other authors (2011). Enterotypes of the human gut microbiome. *Nature* **473**, 174-180.

Atuma, C., Strugala, V., Allen, A. & Holm, L. (2001). The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *Am J Physiol Gastrointest Liver Physiol* **280**, G922-929.

Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920.

Badet, C. & Thebaud, N. B. (2008). Ecology of lactobacilli in the oral cavity: a review of literature. *Open Microbiol J* **2**, 38-48.

Bambou, J. C., Giraud, A., Menard, S., Begue, B., Rakotobe, S., Heyman, M., Taddei, F., Cerf-Bensussan, N. & Gaboriau-Routhiau, V. (2004). *In vitro* and *ex vivo* activation of the TLR5 signaling pathway in intestinal epithelial cells by a commensal *Escherichia coli* strain. *J Biol Chem* **279**, 42984 - 42992.

Barrett, E., Ross, R. P., O'Toole, P. W., Fitzgerald, G. F. & Stanton, C. (2012). gamma-Aminobutyric acid production by culturable bacteria from the human intestine. *J Appl Microbiol* **113**, 411-417.

Bauerl, C., Perez-Martinez, G., Yan, F., Polk, D. B. & Monedero, V. (2010). Functional analysis of the p40 and p75 proteins from *Lactobacillus casei* BL23. *J Mol Microbiol Biotechnol* **19**, 231-241.

Beatson, S. A., Minamino, T. & Pallen, M. J. (2006). Variation in bacterial flagellins: from sequence to structure. *Trends Microbiol* **14**, 151-155.

Begley, M., Gahan, C. G. & Hill, C. (2005). The interaction between bacteria and bile. *FEMS Microbiol Rev* **29**, 625-651.

Begley, M., Hill, C. & Gahan, C. G. (2006). Bile salt hydrolase activity in probiotics. *Appl Environ Microbiol* **72**, 1729-1738.

Belas, R. & Suvanasuthi, R. (2005). The ability of Proteus mirabilis to sense surfaces and regulate virulence gene expression involves FliL, a flagellar basal body protein. *J Bacteriol* **187**, 6789-6803.

Benjamin, J. L., Hedin, C. R., Koutsoumpas, A. & other authors (2012). Smokers with active Crohn's disease have a clinically relevant dysbiosis of the gastrointestinal microbiota. *Inflamm Bowel Dis* **18**, 1092-1100.

Benson, A. K., Kelly, S. A., Legge, R. & other authors (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**, 18933-18938.

Biagi, E., Nylund, L., Candela, M. & other authors (2010). Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One* **5**, e10667.

Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., Perez-Perez, G., Blaser, M. J. & Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci U S A* **103**, 732-737.

Blair, K. M., Turner, L., Winkelman, J. T., Berg, H. C. & Kearns, D. B. (2008). A molecular clutch disables flagella in the *Bacillus subtilis* biofilm. *Science* **320**, 1636-1638.

Blander, J. M. & Sander, L. E. (2012). Beyond pattern recognition: five immune checkpoints for scaling the microbial threat. *Nat Rev Immunol* **12**, 215-225.

Booijink, C. C., El-Aidy, S., Rajilic-Stojanovic, M., Heilig, H. G., Troost, F. J., Smidt, H., Kleerebezem, M., De Vos, W. M. & Zoetendal, E. G. (2010). High temporal and inter-individual variation detected in the human ileal microbiota. *Environ Microbiol* **12**, 3213-3227.

Botos, I., Segal, D. M. & Davies, D. R. (2011). The structural biology of Toll-like receptors. *Structure* **19**, 447-459.

Bown, R. L., Gibson, J. A., Sladen, G. E., Hicks, B. & Dawson, A. M. (1974). Effects of lactulose and other laxatives on ileal and colonic pH as measured by a radiotelemetry device. *Gut* **15**, 999-1004.

Bravo, J. A., Forsythe, P., Chew, M. V., Escaravage, E., Savignac, H. M., Dinan, T. G., Bienenstock, J. & Cryan, J. F. (2011). Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc Natl Acad Sci U S A* **108**, 16050-16055.

Brown, J., Faulds-Pain, A. & Aldridge, P. (2009). The coordination of flagellar gene expression and the flagellar assembly pathway. In *Pili and flagella, current research and future trends*, pp. 99-120. Edited by K. F. Jarrell. Norfolk, UK: Caister Academic Press.

Brown, M. T., Steel, B. C., Silvestrin, C., Wilkinson, D. A., Delalez, N. J., Lumb, C. N., Obara, B., Armitage, J. P. & Berry, R. M. (2012). Flagellar hook flexibility is essential for bundle formation in swimming *Escherichia coli* cells. *J Bacteriol* **194**, 3495-3501.

Buhnik-Rosenblau, K., Danin-Poleg, Y. & Kashi, Y. (2011). Predominant effect of host genetics on levels of *Lactobacillus johnsonii* bacteria in the mouse gut. *Appl Environ Microbiol* **77**, 6531-6538.

Bui, T. P., Kim, Y. J., In, J. G. & Yang, D. C. (2011). *Lactobacillus koreensis* sp. nov., isolated from the traditional Korean food kimchi. *Int J Syst Evol Microbiol* **61**, 772-776.

Burch, A. Y., Shimada, B. K., Mullin, S. W., Dunlap, C. A., Bowman, M. J. & Lindow, S. E. (2012). *Pseudomonas syringae* coordinates production of a motility-enabling surfactant with flagellar assembly. *J Bacteriol* **194**, 1287-1298.

Burnens, A. P., Stanley, J., Sack, R., Hunziker, P., Brodard, I. & Nicolet, J. (1997). The flagellin N-methylase gene *fliB* and an adjacent serovar-specific IS200 element in *Salmonella typhimurium*. *Microbiology* **143 (Pt 5)**, 1539-1547.

Burns, P., Sanchez, B., Vinderola, G., Ruas-Madiedo, P., Ruiz, L., Margolles, A., Reinheimer, J. & de los Reyes-Gavilan, C. G. (2010). Inside the adaptation process of *Lactobacillus delbrueckii* subsp. *lactis* to bile. *Int J Food Microbiol* **142**, 132-141.

Byun, R., Nadkarni, M. A., Chhour, K. L., Martin, F. E., Jacques, N. A. & Hunter, N. (2004). Quantitative analysis of diverse *Lactobacillus* species present in advanced dental caries. *J Clin Microbiol* **42**, 3128-3136.

Casadevall, A. & Pirofski, L. A. (2000). Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infect Immun* **68**, 6511-6518.

Caufield, P. W., Li, Y., Dasanayake, A. & Saxena, D. (2007). Diversity of lactobacilli in the oral cavities of young women with dental caries. *Caries Res* **41**, 2-8.

Chen, B. G., Turner, L. & Berg, H. C. (2007). The wetting agent required for swarming in *Salmonella enterica* serovar Typhimurium is not a surfactant. *J Bacteriol* **189**, 8750-8753.

Chen, S., Beeby, M., Murphy, G. E. & other authors (2011). Structural diversity of bacterial flagellar motors. *EMBO J* **30**, 2972-2981.

Chilcott, G. S. & Hughes, K. T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. *Microbiol Mol Biol Rev* **64**, 694-708.

Choi, Y. J., Im, E., Chung, H. K., Pothoulakis, C. & Rhee, S. H. (2010). TRIF mediates Toll-like receptor 5-induced signaling in intestinal epithelial cells. *J Biol Chem* **285**, 37570-37578.

Christensen, J. E., Pacheco, S. A. & Konkel, M. E. (2009). Identification of a *Campylobacter jejuni* secreted protein required for maximal invasion of host cells. *Molecular Microbiology* **73**, 650-662.

Chung, H., Pamp, S. J., Hill, J. A. & other authors (2012). Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578-1593.

Claesson, M. J., Cusack, S., O'Sullivan, O. & other authors (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4586-4591.

Claesson, M. J., Jeffery, I. B., Conde, S. & other authors (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184.

Collins, J. K., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. & Farrow, J. A. E. (1994). The phylogeny of the genus *Clostridium*: Proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* **44**, 812-826.

Corr, S. C., Li, Y., Riedel, C. U., O'Toole, P. W., Hill, C. & Gahan, C. G. (2007). Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. *Proc Natl Acad Sci U S A* **104**, 7617-7621.

Cotter, P. D., Hill, C. & Ross, R. P. (2005). Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol* **3**, 777-788.

Cunningham, A. & Smith, A. M. (1939). A motile *Lactobacillus* isolated from A.I.V. silage. *J Appl Microbiol* **2**, 7-8.

Dal Bello, F. & Hertel, C. (2006). Oral cavity as natural reservoir for intestinal lactobacilli. *Syst Appl Microbiol* **29**, 69-76.

Darnton, N. C. & Berg, H. C. (2007). Force-extension measurements on bacterial flagella: triggering polymorphic transformations. *Biophys J* **92**, 2230-2236.

Darnton, N. C., Turner, L., Rojevsky, S. & Berg, H. C. (2010). Dynamics of bacterial swarming. *Biophys J* **98**, 2082-2090.

Dasgupta, N., Wolfgang, M. C., Goodman, A. L., Arora, S. K., Jyot, J., Lory, S. & Ramphal, R. (2003). A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol Microbiol* **50**, 809-824.

De Keersmaecker, S. C., Braeken, K., Verhoeven, T. L., Perea Velez, M., Lebeer, S., Vanderleyden, J. & Hols, P. (2006). Flow cytometric testing of green fluorescent protein-tagged *Lactobacillus rhamnosus* GG for response to defensins. *Appl Environ Microbiol* **72**, 4923-4930.

Deibel, R. H. & Niven Jr., C. F. (1958). Microbiology of meat curing. *Appl Microbiol* **6**, 323-327.

Delgado, S., Suarez, A., Otero, L. & Mayo, B. (2004). Variation of microbiological and biochemical parameters in the faeces of two healthy people over a 15 day period. *Eur J Nutr* **43**, 375-380.

DePamphilis, M. L. & Adler, J. (1971). Fine structure and isolation of the hook-basal body complex of flagella from *Escherichia coli* and *Bacillus subtilis*. *J Bacteriol* **105**, 384-395.

Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **54**, 1469-1476.

Desai, A. R., Musil, K. M., Carr, A. P. & Hill, J. E. (2009). Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure. *Vet Microbiol* **137**, 120-128.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W. H., Lakshmanan, A. & Wade, W. G. (2010). The human oral microbiome. *J Bacteriol* **192**, 5002-5017.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. & Knight, R. (2010). Delivery mode shapes the acquisition and structure

of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* **107**, 11971-11975.

Duck, L. W., Walter, M. R., Novak, J., Kelly, D., Tomasi, M., Cong, Y. & Elson, C. O. (2007). Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis* **13**, 1191-1201.

Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. (2002). *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int J Syst Evol Microbiol* **52**, 1615-1620.

Duncan, S. H., Louis, P. & Flint, H. J. (2004). Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* **70**, 5810-5817.

Duncan, S. H., Aminov, R. I., Scott, K. P., Louis, P., Stanton, T. B. & Flint, H. J. (2006). Proposal of *Roseburia faecis* sp. nov., *Roseburia hominis* sp. nov. and *Roseburia inulinivorans* sp. nov., based on isolates from human faeces. *Int J Syst Evol Microbiol* **56**, 2437-2441.

Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A. M., Flint, H. J. & Lobley, G. E. (2007). Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl Environ Microbiol* **73**, 1073-1078.

Eaton, K. A., Morgan, D. R. & Krakowka, S. (1992). Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. *J Med Microbiol* **37**, 123-127.

Eaves-Pyles, T., Bu, H. F., Tan, X. D., Cong, Y., Patel, J., Davey, R. A. & Strasser, J. E. (2011). Luminal-applied flagellin is internalized by polarized intestinal epithelial cells and elicits immune responses via the TLR5 dependent mechanism. *PLoS One* **6**, e24869.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638.

Emerson, S. U., Tokuyasu, K. & Simon, M. I. (1970). Bacterial flagella: polarity of elongation. *Science* **169**, 190-192.

Erridge, C., Duncan, S. H., Bereswill, S. & Heimesaat, M. M. (2010). The induction of colitis and ileitis in mice is associated with marked increases in intestinal concentrations of stimulants of TLRs 2, 4, and 5. *PloS one* **5**, e9125.

Euzeby, J. P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* **47**, 590-592.

Fang, F., Li, Y., Bumann, M., Raftis, E. J., Casey, P. G., Cooney, J. C., Walsh, M. A. & O'Toole, P. W. (2009). Allelic variation of bile salt hydrolase genes in

Lactobacillus salivarius does not determine bile resistance levels. *J Bacteriol* **191**, 5743-5757.

FAO/WHO (2001). Health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria, pp. 1-34. Cordoba: Food and Agriculture Organisation of the United Nations, World Health Organisation.

Farhat, K., Riekenberg, S., Heine, H. & other authors (2008). Heterodimerization of TLR2 with TLR1 or TLR6 expands the ligand spectrum but does not lead to differential signaling. *J Leukoc Biol* **83**, 692-701.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of lactobacilli and bifidobacteria. *Curr Issues Intest Microbiol* **8**, 44-61.

Ferooz, J. & Letesson, J. J. (2010). Morphological analysis of the sheathed flagellum of *Brucella melitensis*. *BMC Res Notes* **3**, 333.

Ferooz, J., Lemaire, J., Delory, M., De Bolle, X. & Letesson, J. J. (2011). RpoE1, an extracytoplasmic function sigma factor, is a repressor of the flagellar system in *Brucella melitensis*. *Microbiology* **157**, 1263-1268.

Fierer, J. & Guiney, D. G. (2001). Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. *J Clin Invest* **107**, 775-780.

Fischbach, M. A. & Sonnenberg, J. L. (2010). Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* **10**, 336-347.

Forde, B. M. (2013). Genomics of commensal lactobacilli. In *Microbiology*, pp. 290. Cork: University College Cork.

Franceschi, C., Bonafe, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E. & De Benedictis, G. (2000). Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann N Y Acad Sci* **908**, 244-254.

Fred, E. B., Peterson, W. H. & Davenport, A. (1919). Acid fermentation of xylose. *J Biol Chem* **XXXIX**, 347-385.

Gaboriau-Routhiau, V., Rakotobe, S., Lecuyer, E. & other authors (2009). The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677-689.

Galkin, V. E., Yu, X., Bielnicki, J., Heuser, J., Ewing, C. P., Guerry, P. & Egelman, E. H. (2008). Divergence of quaternary structures among bacterial flagellar filaments. *Science* **320**, 382-385.

Geis, G., Suerbaum, S., Forsthoff, B., Leying, H. & Opferkuch, W. (1993). Ultrastructure and biochemical studies of the flagellar sheath of *Helicobacter pylori*. *J Med Microbiol* **38**, 371-377.

Gemmel, M. & Hodgkiss, W. (1964). The physiological characters and flagellar arrangement of motile homofermentative lactobacilli. *J Gen Microbiol* **35**, 519-526.

Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J. & Madara, J. L. (2001). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* **167**, 1882-1885.

Gewirtz, A. T., Vijay-Kumar, M., Brant, S. R., Duerr, R. H., Nicolae, D. L. & Cho, J. H. (2006). Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease. *Am J Physiol Gastrointest Liver Physiol* **290**, G1157-1163.

Ghelardi, E., Salvetti, S., Ceragioli, M., Gueye, S. A., Celandroni, F. & Senesi, S. (2012). Contribution of surfactin and SwrA to flagellin expression, swimming and surface motility in *Bacillus subtilis*. *Appl Environ Microbiol*.

Girgis, H. S., Liu, Y., Ryu, W. S. & Tavazoie, S. (2007). A comprehensive genetic characterization of bacterial motility. *PLoS Genet* **3**, 1644-1660.

Gomez-Gomez, L., Bauer, Z. & Boller, T. (2001). Both the extracellular leucine-rich repeat domain and the kinase activity of FSL2 are required for flagellin binding and signaling in *Arabidopsis*. *Plant Cell* **13**, 1155-1163.

Gomez, S. A., Arguelles, C. L., Guerrieri, D. & other authors (2009). Secretory leukocyte protease inhibitor: a secreted pattern recognition receptor for mycobacteria. *Am J Respir Crit Care Med* **179**, 247-253.

Gosalbes, M. J., Durban, A., Pignatelli, M., Abellan, J. J., Jimenez-Hernandez, N., Perez-Cobas, A. E., Latorre, A. & Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS one* **6**, e17447.

Gray, H. (1918). The large intestine. In *Anatomy of the human body*. Edited by W. H. Lewis. Philadelphia: Lea & Febiger.

Gross, E. L., Leys, E. J., Gasparovich, S. R., Firestone, N. D., Schwartzbaum, J. A., Janies, D. A., Asnani, K. & Griffen, A. L. (2010). Bacterial 16S sequence analysis of severe caries in young permanent teeth. *J Clin Microbiol* **48**, 4121-4128.

Grundling, A., Burrack, L., Bouwer, H. & Higgins, D. E. (2004). *Listeria monocytogenes* regulates flagellar motility gene expression through MogR, a transcriptional repressor required for virulence. *Proc Natl Acad Sci U S A* **101**, 12318-12323.

Hagbom, M., Istrate, C., Engblom, D. & other authors (2011). *Rotavirus* stimulates release of serotonin (5-HT) from human enterochromaffin cells and activates brain structures involved in nausea and vomiting. *PLoS Pathog* **7**, e1002115.

Hakalehto, E., Vilpponen-Salmela, T., Kinnunen, K. & von Wright, A. (2011). Lactic acid bacteria enriched from human gastric biopsies. *ISRN Gastroenterol* **2011**, 109183.

Hammes, W. P. & Vogel, R. F. (1995). The genus *Lactobacillus*. In *The lactic acid bacteria: the genera of lactic acid bacteria*, pp. 19-49. Edited by B. J. B. Wood & W. Holzapfel. Edmunds, Suffolk: Blackie academic and professional, an imprint of Chapman and Hall.

Hammes, W. P. & Hertel, C. (2006). The genera *Lactobacillus* and *Carnobacterium*. In *Prokaryotes*, pp. 320-403. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer & E. Stackebrandt. New York: Springer.

Hamon, E., Horvatovich, P., Izquierdo, E., Bringel, F., Marchioni, E., Aoude-Werner, D. & Ennahar, S. (2011). Comparative proteomic analysis of *Lactobacillus plantarum* for the identification of key proteins in bile tolerance. *BMC Microbiol* **11**.

Hamon, E., Horvatovich, P., Bisch, M., Bringel, F., Marchioni, E., Aoude-Werner, D. & Ennahar, S. (2012). Investigation of biomarkers of bile tolerance in *Lactobacillus casei* using comparative proteomics. *J Proteome Res* **11**, 109-118.

Harrison Jr, A. P. & Arne Hansen, P. (1950). Lactobacilli from turkeys. *J Bacteriol* **60**, 543-555.

Harshey, R. M. & Matsuyama, T. (1994). Dimorphic transition in *Escherichia coli* and *Salmonella typhimurium*: surface-induced differentiation into hyperflagellate swarmer cells. *Proc Natl Acad Sci U S A* **91**, 8631-8635.

Hartman, A. L., Lough, D. M., Barupal, D. K., Fiehn, O., Fishbein, T., Zasloff, M. & Eisen, J. A. (2009). Human gut microbiome adopts an alternative state following small bowel transplantation. *Proc Natl Acad Sci U S A* **106**, 17187-17192.

Hawn, T. R., Verbon, A., Lettinga, K. D. & other authors (2003). A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J Exp Med* **198**, 1563-1572.

Hawn, T. R., Wu, H., Grossman, J. M., Hahn, B. H., Tsao, B. P. & Aderem, A. (2005). A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. *Proc Natl Acad Sci U S A* **102**, 10593-10597.

Hayashi, F., Smith, K. D., Ozinsky, A. & other authors (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099-1103.

Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M. & Benno, Y. (2005). Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol* **54**, 1093-1101.

He, G., Shankar, R. A., Chzhan, M., Samouilov, A., Kuppusamy, P. & Zweier, J. L. (1999). Noninvasive measurement of anatomic structure and intraluminal

oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proc Natl Acad Sci U S A* **96**, 4586-4591.

Heijtz, R. D., Wang, S., Anuar, F., Qian, Y., Bjorkholm, B., Samuelsson, A., Hibberd, M. L., Forssberg, H. & Pettersson, S. (2011). Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A* **108**, 3047-3052.

Heilig, H. G., Zoetendal, E. G., Vaughan, E. E., Marteau, P., Akkermans, A. D. & de Vos, W. M. (2002). Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* **68**, 114-123.

Herskovits, A. A., Auerbuch, V. & Portnoy, D. A. (2007). Bacterial ligands generated in a phagosome are targets of the cytosolic innate immune system. *PLoS Pathog* **3**, e51.

Hirota, N., Kitada, M. & Imae, Y. (1981). Flagellar motors of alkalophilic *bacillus* are powered by an electrochemical potential gradient of Na+. *FEBS J* **132**, 278-280.

Hofmann, A. F. (1994). Intestinal absroption of bile acids and biliary constutients. The intestinal component of the enterohepatic ciruclation and the ingetrated system. In *Physiology of the gastrointestinal tract*, pp. 1845-1866. Edited by L. R. Johnson. New York: Raven Press.

Holt, P. G. & Jones, C. A. (2000). The development of the immune system during pregnancy and early life. *Allergy* **55**, 688-697.

Homma, M. & Iino, T. (1985). Locations of hook-associated proteins in flagellar structures of *Salmonella typhimurium*. *J Bacteriol* **162**, 183-189.

Hoskins, L. C., Agustines, M., McKee, W. B., Boulding, E. T., Kriaris, M. & Niedermeyer, G. (1985). Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J Clin Invest* **75**, 944-953.

Houry, A., Briandet, R., Aymerich, S. & Gohar, M. (2010). Involvement of motility and flagella in *Bacillus cereus* biofilm formation. *Microbiology* **156**, 1009-1018.

Hu, D. L., Zhu, G., Mori, F., Omoe, K., Okada, M., Wakabayashi, K., Kaneko, S., Shinagawa, K. & Nakane, A. (2007). *Staphylococcal* enterotoxin induces emesis through increasing serotonin release in intestine and it is downregulated by cannabinoid receptor 1. *Cell Microbiol* **9**, 2267-2277.

Hugo, A. A., De Antoni, G. L. & Perez, P. F. (2010). *Lactobacillus delbrueckii* subsp *lactis* (strain CIDCA 133) resists the antimicrobial activity triggered by molecules derived from enterocyte-like Caco-2 cells. *Lett Appl Microbiol* **50**, 335-340.

Huttenhower, C., Gevers, D., Knight, R. & other authors (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214.

Inagaki, T., Moschetta, A., Lee, Y. K. & other authors (2006). Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proc Natl Acad Sci U S A* **103**, 3920-3925.

Ito, M., Terahara, N., Fujinami, S. & Krulwich, T. A. (2005). Properties of motility in *Bacillus subtilis* powered by the H+-coupled MotAB flagellar stator, Na+-coupled MotPS or hybrid stators MotAS or MotPB. *J Mol Biol* **352**, 396-408.

Iyer, C., Kosters, A., Sethi, G., Kunnumakkara, A. B., Aggarwal, B. B. & Versalovic, J. (2008). Probiotic *Lactobacillus reuteri* promotes TNF-induced apoptosis in human myeloid leukemia-derived cells by modulation of NF-kappaB and MAPK signalling. *Cell Microbiol* **10**, 1442-1452.

Jakobsson, H. E., Jernberg, C., Andersson, A. F., Sjolund-Karlsson, M., Jansson, J. K. & Engstrand, L. (2010). Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS One* **5**, e9836.

Janeway, C. A. (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold spring harbour symposia on quantitative biology* **54**, 1-13.

Kagkli, D. M., Vancanneyt, M., Vandamme, P., Hill, C. & Cogan, T. M. (2007). Contamination of milk by enterococci and coliforms from bovine faeces. *J Appl Microbiol* **103**, 1393-1405.

Kalmokoff, M. L., Allard, S., Austin, J. W., Whitford, M. F., Hefford, M. A. & Teather, R. M. (2000). Biochemical and genetic characterization of the flagellar filaments from the rumen anaerobe *Butyrivibrio fibrisolvens* OR77. *Anaerobe* **6**, 93-109.

Kamada, N., Kim, Y. G., Sham, H. P., Vallance, B. A., Puente, J. L., Martens, E. C. & Nunez, G. (2012). Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* **336**, 1325-1329.

Kandler, O. & Weiss, N. (1986). Regular, nonsporing Gram positive rods, Bergey's manual of sytematic bacteriology. In *Bergey's manual of sytematic bacteriology*, pp. 1208-1234. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore.

Kang, J. Y., Nan, X., Jin, M. S. & other authors (2009). Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* **31**, 873-884.

Kankainen, M., Paulin, L., Tynkkynen, S. & other authors (2009). Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. *Proc Natl Acad Sci U S A* **106**, 17193-17198.

Kanno, T., Matsuki, T., Oka, M. & other authors (2009). Gastric acid reduction leads to an alteration in lower intestinal microflora. *Biochem Biophys Res Commun* **381**, 666-670.

Karlsson, C. L., Molin, G., Cilio, C. M. & Ahrne, S. (2011). The pioneer gut microbiota in human neonates vaginally born at term-a pilot study. *Pediatr Res* **70**, 282-286.

Kearns, D. B. (2010). A field guide to bacterial swarming motility. *Nat Rev Microbiol* **8**, 634-644.

Keestra, A. M., de Zoete, M. R., van Aubel, R. A. & van Putten, J. P. (2008). Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. *Mol Immunol* **45**, 1298-1307.

Kelly-Wintenberg, K., South, S. L. & Montie, T. C. (1993). Tyrosine phosphate in a- and b-type flagellins of *Pseudomonas aeruginosa*. *J Bacteriol* **175**, 2458-2461.

Kelly, D., Campbell, J. I., King, T. P., Grant, G., Jansson, E. A., Coutts, A. G., Pettersson, S. & Conway, S. (2004). Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. *Nat Immunol* **5**, 104 - 112.

Kofoed, E. M. & Vance, R. E. (2011). Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* **477**, 592-595.

Kofoed, E. M. & Vance, R. E. (2012). NAIPs: building an innate immune barrier against bacterial pathogens. NAIPs function as sensors that initiate innate immunity by detection of bacterial proteins in the host cell cytosol. *Bioessays* **34**, 589-598.

Kojima, S. & Blair, D. F. (2004). Solubilization and purification of the MotA/MotB complex of *Escherichia coli*. *Biochemistry* **43**, 26-34.

Kolmeder, C. A., de Been, M., Nikkila, J. & other authors (2012). Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PloS one* **7**, e29913.

Konkel, M. E., Klena, J. D., Rivera-Amill, V., Monteville, M. R., Biswas, D., Raphael, B. & Mickelson, J. (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. *J Bacteriol* **186**, 3296-3303.

Koort, J., Vandamme, P., Schillinger, U., Holzapfel, W. & Bjorkroth, J. (2004). *Lactobacillus curvatus* subsp. *melibiosus* is a later synonym of *Lactobacillus sakei* subsp. *carnosus*. *Int J Syst Evol Microbiol* **54**, 1621-1626.

Kostrzynska, M., Betts, J. D., Austin, J. W. & Trust, T. J. (1991). Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *J Bacteriol* **173**, 937-946.

Krause, D. O., Smith, W. J., Conlan, L. L., Gough, J. M., Williamson, M. A. & McSweeney, C. S. (2003). Diet influences the ecology of lactic acid bacteria and *Escherichia coli* along the digestive tract of cattle: neural networks and 16S rDNA. *Microbiology* **149**, 57-65.

Kurdi, P., Kawanishi, K., Mizutani, K. & Yokota, A. (2006). Mechanism of growth inhibition by free bile acids in lactobacilli and bifidobacteria. *J Bacteriol* **188**, 1979-1786.

Kurokawa, K., Itoh, T., Kuwahara, T. & other authors (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181.

Lakhdari, O., Tap, J., Beguet-Crespel, F. & other authors (2011). Identification of NF-kappaB modulation capabilities within human intestinal commensal bacteria. *J Biomed Biotechnol* **2011**, 282356.

Lavelle, E. C., Murphy, C., O'Neill, L. A. & Creagh, E. M. (2010). The role of TLRs, NLRs, and RLRs in mucosal innate immunity and homeostasis. *Mucosal Immunol* **3**, 17-28.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. (2008). Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* **72**, 728-764.

LeBlanc, J. G., Milani, C., Savoy de Giori, G., Sesma, F., van Sinderen, D. & Ventura, M. (2012). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotech* **EPUB AHEAD OF PRINT**.

Leser, T. D., Amenuvor, J. Z., Jensen, T. K., Lindecrona, R. H., Boye, M. & Moller, K. (2002). Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. *Appl Environ Microbiol* **68**, 673-690.

Ley, R. E., Hamady, M., Lozupone, C. & other authors (2008). Evolution of mammals and their gut microbes. *Science* **320**, 1647-1651.

Li, H., Qiu, T., Huang, G. & Cao, Y. (2010). Production of gamma-aminobutyric acid by *Lactobacillus brevis* NCL912 using fed-batch fermentation. *Microb Cell Fact* **9**, 85.

Lind Due, V., Bonde, J., Kann, T. & Perner, A. (2003). Extremely low oxygen tension in the rectal lumen of normal human subjects. *Acta Anaesthesiol Scand* **47**, 372-373.

Lindstedt, G., Lindstedt, S. & Gustafsson, B. E. (1965). Mucus in intestinal contents of germfree rats. *J Exp Med* **121**, 201-213.

Lodes, M. J., Cong, Y., Elson, C. O., Mohamath, R., Landers, C. J., Targan, S. R., Fort, M. & Hershberg, R. M. (2004). Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest* **113**, 1296-1306.

Logan, S. M., Trust, T. J. & Guerry, P. (1989). Evidence for posttranslational modification and gene duplication of *Campylobacter* flagellin. *J Bacteriol* **171**, 3031-3038.

Logan, S. M. (2009). Glycosylation of flagellins. In *Pili and flagella, current research and future trends*, pp. 155-172. Edited by K. F. Jarrell. Norfolk, UK: Caister Academic Press.

Lyte, M. (2004). Microbial endocrinology and infectious disease in the 21st century. *Trends Microbiol* **12**, 14-20.

Macfarlane, G. T., Gibson, G. & Cummings, C. A. (1992). Comparison of fermentation reactions in different regions of the human colon. *J Appl Microbiol* **72**, 57-64.

Macfarlane, G. T. & Macfarlane, S. (2012). Bacteria, colonic fermentation, and gastrointestinal health. *J AOAC Int* **95**, 50-60.

Macias-Rodriguez, M. E., Zagorec, M., Ascencio, F., Vazquez-Juarez, R. & Rojas, M. (2009). *Lactobacillus fermentum* BCS87 expresses mucus- and mucin-binding proteins on the cell surface. *J Appl Microbiol* **107**, 1866-1874.

Macnab, R. M. & Koshland, D. E., Jr. (1972). The gradient-sensing mechanism in bacterial chemotaxis. *Proc Natl Acad Sci U S A* **69**, 2509-2512.

Macnab, R. M. (1977). Bacterial flagella rotating in bundles: a study in helical geometry. *Proc Natl Acad Sci U S A* **74**, 221-225.

Macnab, R. M. (1996). Flagella and motility. In *Escherichia coli and Salmonella: Cellular and molecular biology*, pp. 123-145. Edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham & other authors. Washington D. C.: ASM Press.

Macnab, R. M. (1999). The bacterial flagellum: reversible rotary propellor and type III export apparatus. *J Bacteriol* **181**, 7149-7153.

Macnab, R. M. (2003). How bacteria assemble flagella. *Annu Rev Microbiol* **57**, 77-100.

Maki-Yonekura, S., Yonekura, K. & Namba, K. (2010). Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. *Nat Struct Mol Biol* **17**, 417-422.

Makivuokko, H., Tiihonen, K., Tynkkynen, S., Paulin, L. & Rautonen, N. (2010). The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition. *Br J Nutr* **103**, 227-234.

Mann, S. O. & Oxford, A. E. (1954). Studies of some presumptive lactobacilli isolated from the rumens of young calves. *J Gen Microbiol* **11**, 83-90.

Manson, M. D., Tedesco, P., Berg, H. C., Harold, F. M. & Van der Drift, C. (1977). A protonmotive force drives bacterial flagella. *Proc Natl Acad Sci U S A* **74**, 3060-3064.

Mariconda, S., Wang, Q. & Harshey, R. M. (2006). A mechanical role for the chemotaxis system in swarming motility. *Mol Microbiol* **60**, 1590-1602.

Marsh, P. D. (1994). Microbial ecology of dental plaque and its significance in health and disease. *Adv Dent Res* **8**, 263-271.

Marsh, P. D. (2003). Are dental diseases examples of ecological catastrophes? *Microbiology* **149**, 279-294.

Marteau, P., Pochart, P., Dore, J., Bera-Maillet, C., Bernalier, A. & Corthier, G. (2001). Comparative study of bacterial groups within the human cecal and fecal microbiota. *Appl Environ Microbiol* **67**, 4939-4942.

Martin, J. H. & Savage, D. C. (1985). Purification and characterisation of flagella from *Roseburia cecicola*, an obligately anaerobic bacterium. *J Gen Microbiol* **131**, 2075-2078.

Martin, J. H. & Savage, D. C. (1988). Cloning, nucleotide sequence, and taxonomic implication of flagellin gene of *Roseburia cecicola*. *J Bacteriol* **170**, 2612-2617.

Maukonen, J., Matto, J., Suihko, M. L. & Saarela, M. (2008). Intra-individual diversity and similarity of salivary and faecal microbiota. *J Med Microbiol* **57**, 1560-1568.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107-118.

McCaffrey, R. L., Fawcett, P., O'Riordan, M., Lee, K. D., Havell, E. A., Brown, P. O. & Portnoy, D. A. (2004). A specific gene expression program triggered by Gram-positive bacteria in the cytosol. *Proc Natl Acad Sci U S A* **101**, 11386-11391.

McGroarty, J. A. (1994). Cell surface appendages of lactobacilli. *FEMS Microbiol Lett* **124**, 405-410.

McGuckin, M. A., Linden, S. K., Sutton, P. & Florin, T. H. (2011). Mucin dynamics and enteric pathogens. *Nat Rev Microbiol* **9**, 265-278.

Merz, A. J., So, M. & Sheetz, M. P. (2000). Pilus retraction powers bacterial twitching motility. *Nature* **407**, 98-102.

Minamino, T. & Macnab, R. M. (1999). Components of the *Salmonella* flagellar export apparatus and classification of export substrates. *J Bacteriol* **181**, 1388-1394.

Minamino, T., Imada, K., Kinoshita, M., Nakamura, S., Morimoto, Y. V. & Namba, K. (2011). Structural insight into the rotational switching mechanism of the bacterial flagellar motor. *PLoS Biol* **9**, e1000616.

Minnick, M. F. & Anderson, B. E. (2006). The Genus *Bartonella*. In *The Prokaryotes*, pp. 467-492. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer & E. Stackebrandt. Minneapolis: Springer.

Morita, H., Toh, H., Fukuda, S. & other authors (2008). Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. *DNA Res* **15**, 151-161.

Morita, H., Nakano, A., Shimazu, M. & other authors (2009). *Lactobacillus hayakitensis*, *L. equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora of healthy thoroughbreds. *Anim Sci J* **80**, 339-346.

Moser, S. A. & Savage, D. C. (2001). Bile salt hydrolase activity and resistance to toxicity of conjugated bile salts are unrelated properties in lactobacilli *Appl Environ Microbiol* **67**, 3476-3480.

Mutlu, E. A., Gillevet, P. M., Rangwala, H., Sikaroodi, M., Naqvi, A., Engen, P. A., Kwasny, M., Lau, C. K. & Keshavarzian, A. (2012). Colonic microbiome is altered in alcoholism. *Am J Physiol Gastrointest Liver Physiol* **302**, G966-978.

Muzio, M., Bosisio, D., Polentarutti, N. & other authors (2000). Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: selective expression of TLR3 in dendritic cells. *J Immunol* **164**, 5998-6004.

Niehus, E., Gressmann, H., Ye, F. & other authors (2004). Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of *Helicobacter pylori*. *Mol Microbiol* **52**, 947-961.

Nielsen, D. S., Moller, P. L., Rosenfeldt, V., Paerregaard, A., Michaelsen, K. F. & Jakobsen, M. (2003). Case study of the distribution of mucosa-associated *Bifidobacterium* species, *Lacobacillus* species, and other lactic acid bactgeria in the human colon. *Appl Environ Microbiol* **69**, 7545-7548.

Nishimura, M. & Naito, S. (2005). Tissue-specific mRNA expression profiles of human toll-like receptors and related genes. *Biol Pharm Bull* **28**, 886-892.

O'Mahony, D. S., Pham, U., Iyer, R., Hawn, T. R. & Liles, W. C. (2008). Differential constitutive and cytokine-modulated expression of human Toll-like receptors in primary neutrophils, monocytes, and macrophages. *Int J Med Sci* **5**, 1-8.

O'Neill, L. A. & Bowie, A. G. (2007). The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* **7**, 353-364.

O'Neill, L. A. J., Bryant, C. E. & Doyle, S. L. (2009). Therapeutic targeting of Toll-like receptors for infectionus and inflammatory diseases and cancer. *Pharmacol Rev* **61**, 177 - 197.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* **5**, e177.

Peel, M., Donachie, W. & Shaw, A. (1988). Temperature-dependent expression of flagella of *Listeria monocytogenes* studied by electron microscopy, SDS-PAGE and western blotting. *J Gen Microbiol* **134**, 2171-2178.

Peschel, A. & Sahl, H. G. (2006). The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat Rev Microbiol* **4**, 529-536.

Petnicki-Ocwieja, T., Hrncir, T., Liu, Y. J., Biswas, A., Hudcovic, T., Tlaskalova-Hogenova, H. & Kobayashi, K. S. (2009). Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc Natl Acad Sci U S A* **106**, 15813-15818.

Power, M. E., Alm, R. A. & Trust, T. J. (1992). Biochemical and antigenic properties of the *Campylobacter* flagellar hook protein. *J Bacteriol* **174**, 3874-3883.

Proell, M., Riedl, S. J., Fritz, J. H., Rojas, A. M. & Schwarzenbacher, R. (2008). The Nod-like receptor (NLR) family: a tale of similarities and differences. *PLoS One* **3**, e2119.

Qin, J., Li, R., Raes, J. & other authors (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65.

Qin, J., Li, Y., Cai, Z. & other authors (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60.

Rad, R., Ballhorn, W., Voland, P. & other authors (2009). Extracellular and intracellular pattern recognition receptors cooperate in the recognition of *Helicobacter pylori*. *Gastroenterology* **136**, 2247-2257.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229-241.

Rao, C. V., Glekas, G. D. & Ordal, G. W. (2008). The three adaptation systems of *Bacillus subtilis* chemotaxis. *Trends Microbiol* **16**, 480-487.

Reikvam, D. H., Erofeev, A., Sandvik, A. & other authors (2011). Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS One* **6**, e17996.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Rhee, K. J., Sethupathi, P., Driks, A., Lanning, D. K. & Knight, K. L. (2004). Role of commensal bacteria in development of gut-associated lymphoid tissues and preimmune antibody repertoire. *J Immunol* **172**, 1118-1124.

Rhee, S. H., Im, E., Riegler, M., Kokkotou, E., O'Brien, M. & Pothoulakis, C. (2005). Pathophysiological role of Toll-like receptor 5 engagement by bacterial flagellin in colonic inflammation. *Proc Natl Acad Sci U S A* **102**, 13610-13615.

Riboulet-Bisson, E., Sturme, M. H., Jeffery, I. B. & other authors (2012). Effect of *Lactobacillus salivarius* bacteriocin Abp118 on the mouse and pig intestinal microbiota. *PLoS One* **7**, e31113.

Ridlon, J. M., Kang, D. J. & Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* **47**, 241-259.

Rinttila, T., Kassinen, A., Malinen, E., Krogius, L. & Palva, A. (2004). Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* **97**, 1166-1177.

Ritchie, L. E., Burke, K. F., Garcia-Mazcorro, J. F., Steiner, J. M. & Suchodolski, J. S. (2010). Characterization of fecal microbiota in cats using universal 16S rRNA gene and group-specific primers for *Lactobacillus* and *Bifidobacterium* spp. *Vet Microbiol* **144**, 140-146.

Roberfroid, M., Gibson, G. R., Hoyles, L. & other authors (2010). Prebiotic effects: metabolic and health benefits. *Br J Nutr* **104 Suppl 2**, S1-63.

Roberts, E. & Frankel, S. (1950). gamma-Aminobutyric acid in brain: its formation from glutamic acid. *J Biol Chem* **187**, 55-63.

Roos, S., Engstrand, L. & Jonsson, H. (2005). *Lactobacillus gastricus* sp. nov., *Lactobacillus antri* sp. nov., *Lactobacillus kalixensis* sp. nov. and *Lactobacillus ultunensis* sp. nov., isolated from human stomach mucosa. *Int J Syst Evol Microbiol* **55**, 77-82.

Roshchina, V. V. (2010). Evolutionary considerations of neurotransmitters in microbial, plant and animal cells. In *Microbial endocrinology, interkingdom signaling in infectious disease and health*, pp. 17-52. Edited by M. Lyte & P. P. Freestone. New york: Springer.

Round, J. L. & Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A* **107**, 12204-12209.

Round, J. L., Lee, S. M., Li, J., Tran, G., Jabri, B., Chatila, T. A. & Mazmanian, S. K. (2011). The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332**, 974-977.

Rozen, D. E., Philippe, N., Arjan de Visser, J., Lenski, R. E. & Schneider, D. (2009). Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecol Lett* **12**, 34-44.

Ruas-Madiedo, P., Gueimonde, M., Fernandez-Garcia, M., de los Reyes-Gavilan, C. G. & Margolles, A. (2008). Mucin degradation by *Bifidobacterium strains* isolated from the human intestinal microbiota. *Appl Environ Microbiol* **74**, 1936-1940.

Ruseler-van Embden, J. G., van Lieshout, L. M., Gosselink, M. J. & Marteau, P. (1995). Inability of *Lactobacillus casei* strain GG, *L. acidophilus*, and *Bifidobacterium bifidum* to degrade intestinal mucus glycoproteins. *Scand J Gastroenterol* **30**, 675-680.

Russell, W. R., Gratz, S. W., Duncan, S. H. & other authors (2011). High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *Am J Clin Nutr* **93**, 1062-1072.

Rust, M., Borchert, S., Niehus, E. & other authors (2009). The *Helicobacter pylori* anti-sigma factor FlgM is predominantly cytoplasmic and cooperates with the flagellar basal body protein FlhA. *J Bacteriol* **191**, 4824-4834.

Ryan, K. A., Jayaraman, T., Daly, P., Canchaya, C., Curran, S., Fang, F., Quigley, E. M. & O'Toole, P. W. (2008). Isolation of lactobacilli with probiotic properties from the human stomach. *Lett Appl Microbiol* **47**, 269-274.

Ryu, W. S., Berry, R. M. & Berg, H. C. (2000). Torque-generating units of the flagellar motor of *Escherichia coli* have a high duty ratio. *Nature* **403**, 444-447.

Samatey, F. A., Imada, K., Nagashima, S., Vonderviszt, F., Kumasaka, T., Yamamoto, M. & Namba, K. (2001). Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* **410**, 331-337.

Samatey, F. A., Matsunami, H., Imada, K. & other authors (2004). Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature* **431**, 1062-1068.

Sander, L. E., Davis, M. J., Boekschoten, M. V., Amsen, D., Dascher, C. C., Ryffel, B., Swanson, J. A., Muller, M. & Blander, J. M. (2011). Detection of prokaryotic mRNA signifies microbial viability and promotes immunity. *Nature* **474**, 385-389.

Savage, D. C. (1972). Associations and physiological interactions of indigenous microorganisms and gastrointestinal epithelia. *Am J Clin Nutr* **25**, 1372-1379.

Schaefer, L., Auchtung, T. A., Hermans, K. E., Whitehead, D., Borhan, B. & Britton, R. A. (2010). The antimicrobial compound reuterin (3-hydroxypropionaldehyde) induces oxidative stress via interaction with thiol groups. *Microbiology* **156**, 1589-1599.

Schlee, M., Harder, J., Koten, B., Stange, E. F., Wehkamp, J. & Fellermann, K. (2008). Probiotic lactobacilli and VSL#3 induce enterocyte beta-defensin 2. *Clin Exp Immunol* **151**, 528-535.

Scott, K. P., Martin, J. C., Chassard, C. & other authors (2011). Substrate-driven gene expression in *Roseburia inulinivorans*: importance of inducible enzymes in the utilization of inulin and starch. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4672-4679.

Sharpe, M. E., Latham, M. J., Garvie, E. I., Zirngibl, J. & Kandler, O. (1973). Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp. nov. and *Lactobacillus vitulinus* sp. nov. *J Gen Microbiol* **77**, 37-49.

Sharpe, M. E. & Dellaglio, F. (1977). Deoxyribonucleic acid homology in anaerobic lactobacilli and in possibly related species. *Int J Syst Bacteriol* **27**, 19-21.

Shen, A. & Higgins, D. E. (2006). The MogR transcriptional repressor regulates nonhierarchal expression of flagellar motility genes and virulence in *Listeria monocytogenes*. *PLoS Pathog* **2**, e30.

Shimoyama, T., Kato, S., Ishii, S. & Watanabe, K. (2009). Flagellum mediates symbiosis. *Science* **323**, 1574.

Sims, I. M., Frese, S. A., Walter, J. & other authors (2011). Structure and functions of exopolysaccharide produced by gut commensal *Lactobacillus reuteri* 100-23. *ISME J*.

Smith, K., McCoy, K. D. & Macpherson, A. J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin Immunol* **19**, 59-69.

Smith, K. D., Andersen-Nissen, E., Hayashi, F., Strobe, K., Bergman, M. A., Barrett, S. L., Cookson, B. T. & Aderem, A. (2003). Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat Immunol* **4**, 1247-1253.

Smith, T. G. & Hoover, T. R. (2009). Deciphering bacterial flagellar gene regulatory networks in the genomic era. *Adv Appl Microbiol* **67**, 257-295.

Snyder, L. A. S., Loman, N. J., Futterer, K. & Pallen, M. J. (2009). Bacterial flagellar diversity and evolution: seek simplicity and distrust it? *Trends Microbiol* **17**, 1-5.

Sokol, H., Pigneur, B., Watterlot, L. & other authors (2008). *Faecalibacterium parusnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731-16736.

Srinivasan, N. (2010). Telling apart friend from foe: discriminating between commensals and pathogens at mucosal sites. *Innate Immun* **16**, 391-404.

Stanton, T. B. & Savage, D. C. (1983). Colonization of gnotobiotic mice by *Roseburia cecicola*, a motile, obligately anaerobic bacterium from murine ceca. *Appl Environ Microbiol* **45**, 1677-1684.

Stanton, T. B. & Savage, D. C. (1984). Motility as a factor in bowel colonization by *Roseburia cecicola*, an obligately anaerobic bacterium from the mouse caecum. *J Gen Microbiol* **130**, 173-183.

Stanton, T. B., Duncan, S. H. & Flint, H. J. (2009). Genus XVI. *Roseburia* Stanton and Savage 1983a, 626. In *Bergey's manual of systematic bacteriology*, pp. 954-956. Edited by P. Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. H. Schleifer & W. B. Whitman. New York: Springer

Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A* **99**, 15451-15455.

Stsepetova, J., Sepp, E., Kolk, H., Loivukene, K., Songisepp, E. & Mikelsaar, M. (2011). Diversity and metabolic impact of intestinal *Lactobacillus* species in healthy adults and the elderly. *Br J Nutr* **105**, 1235-1244.

Sui, J., Lieighton, S., Busta, F. & Brady, L. (2002). 16S ribosomal DNA analysis of the faecal lactobacilli composition of human subjects consuming a probiotic strain *Lactobacillus acidophilus* NCFM. *J Appl Microbiol* **93**, 907-912.

Sun, Y. H., Rolan, H. G. & Tsolis, R. M. (2007). Injection of flagellin into the host cell cytosol by *Salmonella enterica* serotype *Typhimurium*. *J Biol Chem* **282**, 33897-33901.

Suzuki, H., Yonekura, K. & Namba, K. (2004). Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis. *J Mol Biol* **337**, 105-113.

Swanson, P. A., 2nd, Kumar, A., Samarin, S., Vijay-Kumar, M., Kundu, K., Murthy, N., Hansen, J., Nusrat, A. & Neish, A. S. (2011). Enteric commensal bacteria potentiate epithelial restitution via reactive oxygen species-mediated inactivation of focal adhesion kinase phosphatases. *Proc Natl Acad Sci U S A* **108**, 8803-8808.

Swidsinski, A., Loening-Baucke, V., Theissig, F., Engelhardt, H., Bengmark, S., Koch, S., Lochs, H. & Dorffel, Y. (2007). Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* **56**, 343-350.

Szurmant, H. & Ordal, G. W. (2004). Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol Mol Biol Rev* **68**, 301-319.

Takeda, K. & Akira, S. (2004). TLR signaling pathways. *Semin Immunol* **16**, 3-9.

Takeuchi, O. & Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell* **140**, 805-820.

Talarico, T. L., Casas, I. A., Chung, T. C. & Dobrogosz, W. J. (1988). Production and isolation of reuterin, a growth inhibitor produced by *Lactobacillus reuteri*. *Antimicrob Agents Chemother* **32**, 1854-1858.

Tallant, T., Deb, A., Kar, N., Lupica, J., de Veer, M. & DiDonato, J. (2004). Flagellin acting via TLR5 is the major activator of key signaling pathways leading to NF-kappaB and proinflammatory gene program activation in intestinal epithelial cells. *BMC Microbiol* **4**, 33.

Tannock, G. W., Munro, K., Harmsen, H. J., Welling, G. W., Smart, J. & Gopal, P. K. (2000). Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20. *Appl Environ Microbiol* **66**, 2578-2588.

Tannock, G. W. (2004). A special fondness for lactobacilli. *Appl Environ Microbiol* **70**, 3189-3194.

Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). A genomic perspective on protein families. *Science* **278**, 631-637.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D. & other authors (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.

Taweechotipatr, M., Iyer, C., Spinler, J. K., Versalovic, J. & Tumwasorn, S. (2009). *Lactobacillus saerimneri* and *Lactobacillus ruminis*: novel human-derived probiotic strains with immunomodulatory activities. *FEMS Microbiol Lett* **293**, 65-72.

Terashima, H., Fukuoka, H., Yakushi, T., Kojima, S. & Homma, M. (2006). The *Vibrio* motor proteins, MotX and MotY, are associated with the basal body of Na-driven flagella and required for stator formation. *Mol Microbiol* **62**, 1170-1180.

Thomas, C. M., Hong, T., van Pijkeren, J. P., Hemarajata, P., Trinh, D. V., Hu, W., Britton, R. A., Kalkum, M. & Versalovic, J. (2012). Histamine derived from probiotic *Lactobacillus reuteri* suppresses TNF via modulation of PKA and ERK signaling. *PLoS One* **7**, e31951.

Ting, J. P., Lovering, R. C., Alnemri, E. S. & other authors (2008). The NLR gene family: a standard nomenclature. *Immunity* **28**, 285-287.

Torriani, S., Van Reenen, G. A., Klein, G., Reuter, G., Dellaglio, F. & Dicks, L. M. (1996). *Lactobacillus curvatus* subsp. *curvatus* subsp. nov. and *Lactobacillus curvatus* subsp. *melibiosus* subsp. nov. and *Lactobacillus sake* subsp. *sake* subsp. nov. and *Lactobacillus sake* subsp. *carnosus* subsp. nov., new subspecies of *Lactobacillus curvatus* Abo-Elnaga and Kandler 1965 and *Lactobacillus sake* Katagiri, Kitahara, and Fukami 1934 (Klein et al. 1996, emended descriptions), respectively. *Int J Syst Bacteriol* **46**, 1158-1163.

Tronick, S. R. & Martinez, R. J. (1971). Methylation of the flagellin of *Salmonella typhimurium*. *J Bacteriol* **105**, 211-219.

Tuohy, K. M., Pinart-Gilberga, M., Jones, M., Hoyles, L., McCartney, A. L. & Gibosn, G. R. (2006). Survivability of a probiotic *Lactobacillus casei* in the

gastrointestinal tract of healthy human volunteers and its impact on the faecal microflora. *J Appl Microbiol* **102**, 1026-1032.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T. & other authors (2009). A core gut microbiome in obese and lean twins. *Nature* **457**, 480-484.

Turner, L., Zhang, R., Darnton, N. C. & Berg, H. C. (2010). Visualization of flagella during bacterial swarming. *J Bacteriol* **192**, 3259-3267.

Turpin, W., Humblot, C., Noordine, M. L., Thomas, M. & Guyot, J. P. (2012). *Lactobacillaceae* and cell adhesion: genomic and functional screening. *PLoS One* **7**, e38034.

UniGene UniGene: EST profile, Hs.604542 - Toll-like receptor 5: National Center for Biogechnology Information.

van Baarlen, P., Troost, F. J., van Hemert, S., van der Meer, C., de Vos, W. M., de Groot, P. J., Hooiveld, G. J., Brummer, R. J. & Kleerebezem, M. (2009). Differential NF-kB pathways induction by *Lactobacillus plantarum* in the duodenum of healthy humans correlating with immune tolerance. *Proc Natl Acad Sci U S A* **106**, 2371-2376.

van Houte, J. & Green, D. B. (1974). Relationship between the concentration of bacteria in saliva and the colonization of teeth in humans. *Infect Immun* **9**, 624-630.

Vance, R. E., Isberg, R. R. & Portnoy, D. A. (2009). Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. *Cell Host Microbe* **6**, 10-21.

Vankova, J. (1957). Motile catalase-producing strains of *Lactobacillus delbruckii*. *Nature* **179**, 204.

Verberkmoes, N. C., Russell, A. L., Shah, M. & other authors (2008). Shotgun metaproteomics of human distal gut microbiota. *ISME J* **3**, 179-189.

Vijay-Kumar, M., Aitken, J. D., Carvalho, F. A. & other authors (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228-231.

Vinzing, M., Eitel, J., Lippmann, J. & other authors (2008). NAIP and Ipaf control *Legionella pneumophila* replication in human cells. *J Immunol* **180**, 6808-6815.

von Schillde, M. A., Hormannsperger, G., Weiher, M. & other authors (2012). Lactocepin secreted by *Lactobacillus* exerts anti-inflammatory effects by selectively degrading proinflammatory chemokines. *Cell Host Microbe* **11**, 387-396.

Wade, W. G. (2006). The genus *Eubacterium* and related genera. In *The Prokaryotes*, pp. 823-835. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer & E. Stackebrandt. New York: Springer.

Walker, A. W., Duncan, S. H., McWilliam Leitch, E. C., Child, M. W. & Flint, H. J. (2005). pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Appl Environ Microbiol* **71**, 3692-3700.

Walker, A. W., Ince, J., Duncan, S. H. & other authors (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* **5**, 220-230.

Walter, J., Loach, D. M., Alqumber, M., Rockel, C., Hermann, C., Pfitzenmaier, M. & Tannock, G. W. (2007). D-alanyl ester depletion of teichoic acids in *Lactobacillus reuteri* 100-23 results in impaired colonization of the mouse gastrointestinal tract. *Environ Microbiol* **9**, 1750-1760.

Walter, J. (2008). Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol* **74**, 4985-4996.

Wang, Q., Suzuki, A., Mariconda, S., Porwollik, S. & Harshey, R. M. (2005). Sensing wetness: a new role for the bacterial flagellum. *Embo J* **24**, 2034-2042.

Watanabe, T., Nishio, H., Tanigawa, T. & other authors (2009). Probiotic *Lactobacillus casei* strain Shirota prevents indomethacin-induced small intestinal injury: involvement of lactic acid. *Am J Physiol Gastrointest Liver Physiol* **297**, G506-513.

Waters, R. C., O'Toole, P. W. & Ryan, K. A. (2007). The FliK protein and flagellar hook-length control. *Protein Sci* **16**, 769-780.

Way, S. S., Thompson, L. J., Lopes, J. E., Hajjar, A. M., Kollmann, T. R., Freitag, N. E. & Wilson, C. B. (2004). Characterization of flagellin expression and its role in *Listeria monocytogenes* infection and immunity. *Cell Microbiol* **6**, 235-242.

Willing, B., Voros, A., Roos, S., Jones, C., Jansson, A. & Lindberg, J. E. (2009). Changes in faecal bacteria associated with concentrate and forage-only diets fed to horses in training. *Equine Vet J* **41**, 908-914.

Wilson, C. M. (2011).*Lactobacillus* dialogues: the impact of cell signalling on gene expression and ecological behaviour of a gut commensal. In *Microbiology and Immunology*. Otago: University of Otago.

Wilson, C. M., Aggio, R. B., O'Toole, P. W., Villas-Boas, S. & Tannock, G. W. (2012). Transcriptional and metabolomic consequences of LuxS inactivation reveal a metabolic rather than quorum-sensing role for LuxS in *Lactobacillus reuteri* 100-23. *J Bacteriol* **194**, 1743-1746.

Woodmansey, E. J., McMurdo, M. E., Macfarlane, G. T. & Macfarlane, S. (2004). Comparison of compositions and metabolic activities of fecal microbiotas in young adults and in antibiotic-treated and non-antibiotic-treated elderly subjects. *Appl Environ Microbiol* **70**, 6113-6122.

Wu, G. D., Chen, J., Hoffmann, C. & other authors (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108.

Yan, F., Cao, H., Cover, T. L., Whitehead, R., Washington, M. K. & Polk, D. B. (2007). Soluble proteins produced by probiotic bacteria regulate intestinal epithelial cell survival and growth. *Gastroenterology* **132**, 562-575.

Yasuda, E., Serata, M. & Sako, T. (2008). Suppressive effect on activation of macrophages by *Lactobacillus casei* strain Shirota genes determining the synthesis of cell wall-associated polysaccharides. *Appl Environ Microbiol* **74**, 4746-4755.

Yin, Q. & Zheng, Q. (2005). Isolation and identification of the dominant *Lactobacillus* in gut and faeces of pigs using carbohydrate fermentation and 16S rDNA analysis. *J Biosci Bioeng* **99**, 68-71.

Yonekura, K., Maki, S., Morgan, D. G., DeRosier, D. J., Vonderviszt, F., Imada, K. & Namba, K. (2000). The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science* **290**, 2148-2152.

Yonekura, K., Maki-Yonekura, S. & Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643-650.

Yoon, S. I., Kurnasov, O., Natarajan, V., Hong, M., Gudkov, A. V., Osterman, A. L. & Wilson, I. A. (2012). Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859-864.

Yoon, S. S. & Mekalanos, J. J. (2008). Decreased potency of the *Vibrio cholerae* sheathed flagellum to trigger host innate immunity. *Infect Immun* **76**, 1282-1288.

Yuan, J. & Berg, H. C. (2008). Resurrection of the flagellar rotary motor near zero load. *Proc Natl Acad Sci U S A* **105**, 1182-1185.

Yun, J. H., Yim, D. S., Kang, J. Y. & other authors (2005). Identification of *Lactobacillus ruminus* SPM0211 isolated from healthy Koreans and its antimicrobial activity against some pathogens. *Arch Pharm Res* **28**, 660-666.

Zaura, E., Keijser, B. J. F., Huse, S. M. & Crielaard, W. (2009). Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* **9**.

Zeng, H., Carlson, A. Q., Guo, Y., Yu, Y., Collier-Hyams, L. S., Madara, J. L., Gewirtz, A. T. & Neish, A. S. (2003). Flagellin is the major proinflammatory determinant of enteropathogenic *Salmonella*. *J Immunol* **171**, 3668-3674.

Zhou, J. S., Gopal, P. K. & Gill, H. S. (2001). Potential probiotic lactic acid bacteria *Lactobacillus rhamnosus* (HN001), *Lactobacillus acidophilus* (HN017) and *Bifidobacterium lactis* (HN019) do not degrade gastric mucin *in vitro*. *Int J Food Microbiol* **63**, 81-90.

Zhou, K., Kanai, R., Lee, P., Wang, H. & Modis, Y. (2012). Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. *J Struct Biol* **177**, 402-409.

Zimmer, J., Lange, B., Frick, J. S., Sauer, H., Zimmermann, K., Schwiertz, A., Rusch, K., Klosterhalfen, S. & Enck, P. (2012). A vegan or vegetarian diet substantially alters the human colonic faecal microbiota. *Eur J Clin Nutr* **66**, 53-60.

Zoetendal, E. G., Von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. & De Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* **68**, 3401-3407.

Chapter 2

Characterization of pro-inflammatory flagellin proteins produced by *Lactobacillus ruminis* and related motile lactobacilli

Notes:

L. ruminis and *L. mali* genomes were assembled by B. M. Forde.

Experiments performed by T. Darby were included as additional experimental replicates for statistical analyses of the immune response of human intestinal epithelial cell lines to whole cell and flagellin stimuli.

This chapter has been published as a research article:

Neville, B. A., Forde, B. M., Claesson, M. J., Darby, T., Coghlan, A., Nally, K., Ross, R. P. & O'Toole, P. W. (2012). Characterization of pro-inflammatory flagellin proteins produced by *Lactobacillus ruminis* and related motile lactobacilli. *PLoS One* **7**, e40592.

Table of Contents

Chapter 2

CHARACTERIZATION of PRO-INFLAMMATORY FLAGELLIN PROTEINS PRODUCED by *LACTOBACILLUS RUMINIS* and RELATED MOTILE LACTOBACILLI.

B. Anne Neville[1], Brian M. Forde[1], Marcus J. Claesson[1], Trevor Darby[2], Avril Coghlan[1], Kenneth Nally[2], R. Paul Ross[3] and Paul W. O'Toole[1,2*].

[1] Department of Microbiology, University College Cork, Cork, Ireland.

[2] Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland.

[3]Teagasc, Moorepark Food Research Centre, Moorepark, Fermoy, Co, Cork, Ireland.

*To whom correspondence should be addressed. E-mail: pwotoole@ucc.ie Tel: +353-21-4903997. Fax: +353-21-4903101

Abstract.

Lactobacillus ruminis is one of at least twelve motile but poorly characterized species found in the genus *Lactobacillus*. Of these, only *L. ruminis* has been isolated from mammals, and this species may be considered as an autochthonous member of the gastrointestinal microbiota of humans, pigs and cows. Nine *L. ruminis* strains were investigated here to elucidate the biochemistry and genetics of *Lactobacillus* motility. Six strains isolated from humans were non-motile while three bovine isolates were motile. A complete set of flagellum biogenesis genes was annotated in the sequenced genomes of two strains, ATCC25644 (human isolate) and ATCC27782 (bovine isolate), but only the latter strain produced flagella. Comparison of the *L. ruminis* and *L. mali* DSM20444^{T} motility loci showed that their genetic content and gene-order were broadly similar, although the *L. mali* motility locus was interrupted by an 11.8 Kb region encoding rhamnose utilization genes that is absent from the *L. ruminis* motility locus. Phylogenetic analysis of 39 motile bacteria indicated that *Lactobacillus* motility genes were most closely related to those of motile carnobacteria and enterococci. Transcriptome analysis revealed that motility genes were transcribed at a significantly higher level in motile *L. ruminis* ATCC27782 than in non-motile ATCC25644. Flagellin proteins were isolated from *L. ruminis* ATCC27782 and from three other *Lactobacillus* species, while recombinant flagellin of aflagellate *L. ruminis* ATCC25644 was expressed and purified from *E. coli*. These native and recombinant *Lactobacillus* flagellins, and also flagellate *L. ruminis* cells, triggered interleukin-8 production in cultured human intestinal epithelial cells in a manner suppressed by short interfering RNA directed against Toll-Like Receptor 5. This

study provides genetic, transcriptomic, phylogenetic and immunological insights into the trait of flagellum-mediated motility in the lactobacilli.

2.1 Introduction.

Bacterial flagella are long, thin, proteinaceous structures that form rigid helices which rotate in a counterclockwise direction to propel the cell forward (Macnab, 2003; Yuan *et al.*, 2010). The filament of the bacterial flagellum is composed of polymers of a protein called flagellin, a microbe associated molecular pattern (MAMP) that is recognized by toll-like receptor 5 (TLR5) of the host (Hayashi *et al.*, 2001), and which activates the nuclear factor-κβ (NFκβ) signaling-pathway in epithelial and innate immune cells (Gewirtz *et al.*, 2001; Letran *et al.*, 2011; Tallant *et al.*, 2004). While lactobacilli have been researched intensively because of their food and health-related applications (Walter, 2005), to date, neither their flagella nor their capacity for flagellum-mediated motility has been formally characterized.

From a microbial perspective, flagellate species may have a competitive advantage over non-motile species with respect to niche colonisation (Lane *et al.*, 2007), biofilm formation (Houry *et al.*, 2010; Lemon *et al.*, 2007) and for the secretion of virulence proteins by pathogens (Konkel *et al.*, 2004). However, the production of flagella requires a substantial investment of resources by the bacterium, with ~0.1 % and ~2 % of the cell's energy devoted to flagellum rotation and biosynthesis respectively (Macnab, 1996). For this reason, and also presumably to avoid stimulating immune responses, some species, such as *Listeria monocytogenes* suppress motility gene expression *in vivo* (Grundling *et al.*, 2004; Peel *et al.*, 1988).

Recently, the crystal structures of human TLR5 (hTLR5) (Zhou *et al.*, 2012) and also of *Danio rerio* (zebrafish) TLR5 (drTLR5) in complex with two *Salmonella* flagellin molecules (Yoon *et al.*, 2012) have been determined, providing insight into

the molecular-basis for flagellin recognition by TLR5. In the absence of its ligand, hTLR5 exists as an asymmetric homodimer (Zhou *et al.*, 2012) and like the drTLR5, this structure is proposed to bind two flagellin molecules (Zhou *et al.*, 2012). Previously, alanine scanning mutagenesis had identified specific flagellin residues that reduced TLR5 recognition of flagellin by 76-97 % when mutated (Smith *et al.*, 2003). Several of these critical residues were found on the convex surface of the flagellin molecule and include core flagellin residues R90, L94 and Q97 (Smith *et al.*, 2003). The structural studies confirmed that R90, L94 and Q97 were indeed involved in the TLR5 interaction, and bonded with several TLR5 residues at the primary binding interface B of the drTLR5 structure (Yoon *et al.*, 2012). In particular, the LRR9 loop of drTLR5 formed the major flagellin binding site (Yoon *et al.*, 2012). Specifically, flagellin residues R90 and E114 have been shown to interact with several drTLR5 residues, including TLR5 S271 (Yoon *et al.*, 2012), a naturally varying TLR5 residue that is involved in establishing the flagellin recognition profile of a given species (Andersen-Nissen *et al.*, 2007).

Several flagellate bacterial pathogens have evolved flagellin proteins that are not recognized by human TLR5 (Andersen-Nissen *et al.*, 2005). Residues 89-96 of the amino-terminal D1 domain of *S. typhimurium* flagellin constitute the flagellin pattern recognized by TLR5. These residues are also involved in flagellin polymerization (Andersen-Nissen *et al.*, 2005; Smith *et al.*, 2003). Particular substitutions within this region enable selected flagellate α- and ε-proteobacterial pathogens, including *H. pylori* and *C. jejuni*, to evade immune-recognition without compromising their motility (Andersen-Nissen *et al.*, 2005). While the recognition of *Lactobacillus* cells and their associated molecules by TLRs 2, 4 and 9 has been

well established (Lebeer *et al.*, 2008; Wells, 2011) a direct TLR5-*Lactobacillus* interaction has not been demonstrated previously.

At present, at least twelve motile species have been officially recognized in the genus *Lactobacillus* (Table S2.1), and the *L. salivarius* phylogenetic clade (Felis & Dellaglio, 2007) includes the largest proportion of these motile species. Eleven of these twelve motile *Lactobacillus* species have been isolated from food or environmental sources (Table S2.1). *L. curvatus* subsp. *curvatus* and *L. sakei* subsp. *carnosus* which were described as motile in one study (Torriani *et al.*, 1996), are located in the *L. sakei* clade.

Lactobacilli represent a subdominant element of the human intestinal microbiota. By culture-dependent methods, the faecal *Lactobacillus* population of adult humans was estimated at approximately 10^4-10^8 CFU/g faeces (wet weight) (Dal Bello & Hertel, 2006; Kimura *et al.*, 1997; Kimura *et al.*, 2010). Culture-independent enumeration approaches, such as real-time PCR (Rinttila *et al.*, 2004; Stsepetova *et al.*, 2011) similarly suggest that the lactobacilli are present in the faecal microbiota at concentrations of 10^6-10^8 targets/g faeces (wet weight). Therefore, the lactobacilli represent at most ~ 0.01 % - 0.6 % of the faecal microbiota, and this proportion varies significantly from individual to individual (Maukonen *et al.*, 2008; Tannock *et al.*, 2000).

Certain *Lactobacillus* species and strains have been developed as probiotics for human consumption (Dommels *et al.*, 2009). Accordingly, the potential *in vivo* functions of cell-surface appendages of lactobacilli, such as pili (Kankainen *et al.*, 2009) and surface layer proteins (Jakava-Viljanen *et al.*, 2002; Sun *et al.*, 2012) have been investigated. The flagellate *Lactobacillus* species however, have not attracted much scientific attention thus far, and the perception that the lactobacilli are

uniformly non-motile persists as a consequence (Snyder *et al.*, 2009; Vizoso Pinto *et al.*, 2009). Moreover, well characterized, aflagellate *Lactobacillus* species have been engineered to display *Salmonella* flagellin proteins on their cell-surface with the intention of developing these recombinant strains as vaccine delivery vectors with enhanced adjuvancy (Kajikawa & Igimi, 2010; Kajikawa *et al.*, 2011). Naturally flagellate *Lactobacillus* species have been overlooked for this purpose to date.

L. ruminis has been identified as part of the intestinal microbiota of several mammals, including humans (Delgado *et al.*, 2004; Heilig *et al.*, 2002; Reuter, 2001; Tannock *et al.*, 2000), cows (Krause *et al.*, 2003; Sharpe *et al.*, 1973) and pigs (Al Jassim, 2003; Yin & Zheng, 2005), and it is the only formally recognized motile *Lactobacillus* species which is also autochthonous to the gastrointestinal tract of humans (Reuter, 2001; Tannock *et al.*, 2000). *L. ruminis* is an obligately homofermentative rod-shaped bacterium that tends to form end-to-end filaments or chains of cells (Kandler & Weiss, 1986). Bovine *L. ruminis* isolates, including strain ATCC27782 which is one of the strains analyzed in this study, are motile by means of peritrichous flagella (Sharpe *et al.*, 1973). Although the first report of *L. ruminis* (then called *Catenabacterium catenaforme*) isolation from humans described the bacterium as being motile (Lerche & Reuter, 1960), the strain lodged, (later identified as *L. ruminis* ATCC25644) was non-motile (Reuter, 2001).

Lactobacillus mali (Carr & Davies, 1970) is one of the twelve motile species found in the twenty-five- member *L. salivarius* clade. Historically, the formal description of this species has been complicated by several revisions of its nomenclature (Carr *et al.*, 1977; Kaneuchi *et al.*, 1988). In general, the scientific literature describing the biology of *L. mali* is sparse. Nevertheless, its status as a

wine and cider associated lactic acid bacterium has yielded some insight into the metabolism of the species. *L. mali* is a malolactic bacterium (Dicks & Endo, 2009) capable of producing volatile phenols (Buron *et al.*, 2011; Couto *et al.*, 2006) and biogenic amines (Coton *et al.*, 2010; Landete *et al.*, 2007) which influence the organoleptic properties and the safety of fermented alcoholic beverages. *L. mali* strains are also known to produce exopolysaccharide (Seto *et al.*, 2006) and menaquinones (Kandler & Weiss, 1986; Morishita *et al.*, 1999) suggesting that some *L. mali* strains may be suitable for industrial applications.

We have determined the genomic basis for flagellum-mediated motility in the genus *Lactobacillus* by investigating the motility loci of two *L. ruminis* strains and comparing them to the motility locus of *L. mali*. We have considered the phylogenetic origins of the *L. ruminis* motility genes among the motile species of the order *Lactobacillales*. We have also investigated the non-motility of *L. ruminis* isolates of human origin, and we have characterized the flagellin proteins of, and the innate immune response to, these motile and non-motile autochthonous gastrointestinal (GI) commensals.

2.2 Results.

2.2.1 Motility of Lactobacilli.

Of nine *L. ruminis* strains examined, only strains isolated from cows, (ATCC27780, ATCC27781 and ATCC27782) were flagellate and motile (Figure 2.1, Table S2.2). These strains displayed unicellular motility, but also often formed chains of two to five cells which were motile as a unit (SI movie S2.1). The observed motility of *L. ruminis* ATCC27782 typically decreased with increasing cell density during growth. During early exponential phase, cultures were often “very motile” to “moderately motile” when viewed with a phase contrast microscope. At

late exponential phase, cultures became “weakly motile”, while stationary phase cultures were typically non-motile. The motility of stationary phase cells could be restored by adding fresh medium, suggesting that nutrient depletion may influence the motility phenotype. Three other species of the *L. salivarius* clade tested, specifically *L. ghanensis* L489^{T}, *L. mali* DSM20444^{T} and *L. nagelii* DSM13675^{T} were also motile. *L. ruminis* isolates from humans, including ATCC25644, were not motile when grown in MRS media as determined by phase-contrast microscopy and motility agar plates.

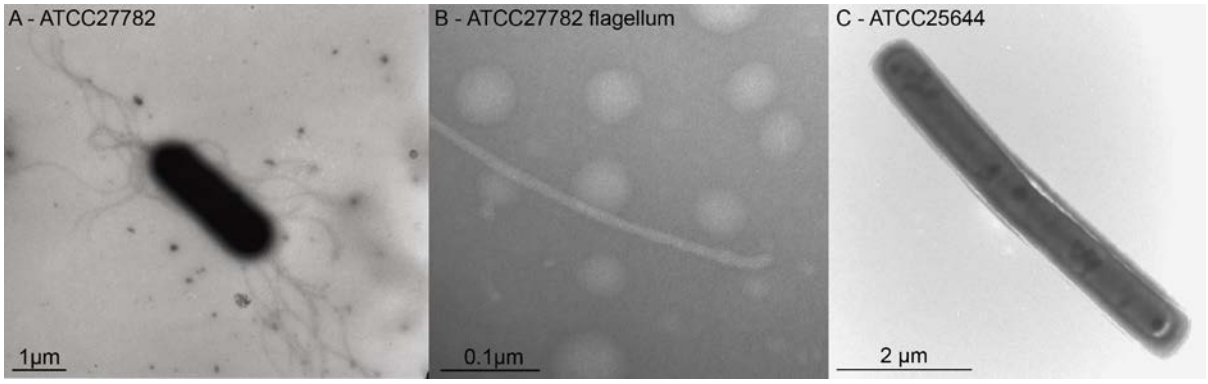


Figure 2.1: Transmission electron micrographs of *L. ruminis* whole cells and flagella. A: *L. ruminis* ATCC27782; 1% ammonium molybdate, 6000x magnification. B: *L. ruminis* ATCC27782 flagellum filament with visible attached hook structure; 2% ammonium molybdate, 250000x magnification. C: *L. ruminis* ATCC25644; 0.25 % ammonium molybdate, 15000x magnification.

2.2.2 Annotation of *L. ruminis* and *L. mali* motility genes.

“High-quality draft” and “finished” (Chain *et al.*, 2009) genome sequences were generated for *L. ruminis* ATCC25644 and ATCC27782 respectively (Forde *et al.*, 2011). Motility-related genes were annotated in loci spanning 45,687 bp and 48,062 bp in *L. ruminis* ATCC25644 and ATCC27782 respectively. The [GC] % of these genomic regions was 44.79 % (ATCC25644) and 44.04 % (ATCC27782),

consistent with the overall values for each genome of 43.76 % and 43.34 % respectively.

A "standard draft" genome 2.6 Mb in size, consisting of 200 contigs assembled into 95 scaffolds, was generated for *L. mali* DSM20444^T. Its motility locus spans 51, 309 bp, and it encodes forty three motility and chemotaxis genes. Its [GC]% of 36.2 % is consistent with the [GC]% of 36 % for the whole genome.

2.2.3 Comparative genomics and phylogeny of *Lactobacillus* motility genes.

The motility loci of *L. ruminis* ATCC25644 and ATCC27782 both encode forty-five predicted flagellum biogenesis and chemotaxis genes (Figure S2.1; Table S2.3), but with a second copy of the *fliC* gene (*fliC*2) and a glycosyltransferase pseudogene being additionally present in the ATCC27782 genome. These motility loci otherwise display conserved order and orientation in the two genomes and are 97 % identical to each other at nucleotide level.

The *L. mali* genome encodes a similar set of motility genes, and its genetic arrangement is broadly similar to that of the *L. ruminis* ATCC27782 motility locus (Figure S2.2). The *L. mali* motility genes are divided between two loci that are separated between *flgL* and *fliC*, by an ~11.8 Kb region that encodes a number of metabolic genes, including those for rhamnose metabolism (*rhaTBMAD*). Relative to the *L. ruminis* motility locus, other notable variations in the *L. mali* motility region include the inversion of the *motAB* gene pair, the absence of homologs for *flaG* and a potential negative regulator (LRC_15730/ANHS_518; see below) of motility gene expression and the presence of only one flagellin gene in a strain that is known to be motile.

The relatedness of the *L. ruminis* motility proteins to those of 39 other *Firmicutes*, *Proteobacteria* and *Thermotogae* was evaluated by generation of

phylogenetic trees for 39 motility protein families. A well supported *L. ruminis-Carnobacterium-Enterococcus* clade was present in 22 motility protein trees and in the 16S rRNA gene tree (Figure S2.3). These data suggest that the *L. ruminis* motility genes were probably acquired by vertical descent via the *L. ruminis-Carnobacterium-Enterococcus* last common ancestor.

2.2.4 Proposed regulation of *L. ruminis* motility gene transcription and bioinformatic analysis of ANHS_518/LRC_15730.

The transcription of motility genes in *L. ruminis* ATCC25644 and ATCC27782 was examined by microarray analysis during the motile and non-motile growth phases, to investigate the genetic basis for their different motility phenotypes. During the motile phase of growth, transcription of most genes in the motility locus was significantly higher in ATCC27782 than in ATCC25644 (Table S2.4), with the exception of LRC_15730, a gene on the opposite DNA strand (Figure S2.1). This gene was transcribed at a much higher level in ATCC25644 (annotated as ANHS_518 in ATCC25644 (Table S2.4)), suggesting that it may function as a negative regulator of flagellum biogenesis. Microarray data also showed that flagellin genes (*fliC1* and *fliC2*) were transcribed at much higher levels in ATCC27782 than in ATCC25644. Since the *fliC1* and *fliC2* genes are 98 % identical and could not be distinguished by their hybridization to the array oligonucleotides, allele-specific qRT-PCR was used to show that only *fliC*2 was transcribed at high levels in ATCC27782, with Cq values of ~23, compared to Cq values of ~37 (indicating very low expression) and 0 (non-expression) for *fli*C1 in ATCC27782 and ATCC25644 respectively. Thus *fliC2* is the majorly expressed flagellin gene in the motile strain, and *fliC1* may be expressed at very low levels only in ATCC27782. The *era* gene comparator had a Cq value in the range of 18-

19.5. Microarray data was also verified by qRT-PCR for *fliM* and LRC_15730/ANHS_518.

Four sigma factors, σ^{70}, σ^{54}, σ^{28}, and $\sigma^{70'}$ (a σ^{70}-like putative extracytoplasmic function sigma factor) were annotated in the *L. ruminis* genome (Forde *et al.*, 2011). These genes are shown schematically in Figure S2.4). Relative transcription of σ^{28} (the flagellin-specific *fliA* sigma factor) was significantly higher in ATCC27782 than in ATCC25644. In contrast, the $\sigma^{70'}$ gene was expressed at a much higher level in ATCC25644 than in ATCC27782. The genes for the housekeeping sigma factors σ^{70} and σ^{54} were not differentially expressed between strains. Likewise there was no differential expression of the gene LRC_15720 which is part of the motility locus and may encode a sigma factor, but which is not formally annotated as such.

The transcription start (+1) sites of several candidate regulatory and effector genes involved in flagellum biogenesis were mapped by 5′ rapid amplification of cDNA ends (RACE) to assist promoter identification *in silico*. Both flagellin genes *fliC*1 and *fliC*2 of *L. ruminis* ATCC27782 appear to be under the control of a σ^{28}-dependent promoter (Figure S2.4). The *fliC*1 gene transcription +1 site could not be experimentally determined because this gene is not highly expressed in either ATCC25644 or ATCC27782. Nevertheless, the sequence identity and spacing of the predicted -35 and -10 boxes at the promoters of *fliC*1 and *fliC*2 was broadly conserved (Figure S2.4). Thus, *fliC*1 is predicted to share the same promoter configuration and type as that of *fliC*2. A short palindromic sequence AGATCT, a known transcription factor binding site (Speck *et al.*, 1999), and the recognition sequence of the restriction enzyme BglII, was identified between the -35 and -10

elements of the *fliC*1 in both *L. ruminis* ATCC27782 and ATCC25644. This palindrome was not present at the ATCC27782 *fliC*2 promoter.

Consistent with other *Lactobacillus* genes and transcripts (McCracken *et al.*, 2000), a purine residue (G) appears in the +1 position of the ANHS_518 gene transcript. However, the promoter region for this gene did not conform to the consensus sequences or motifs predicted for known *Lactobacillus* promoters (McCracken *et al.*, 2000) and so could not be reliably identified *in silico*.

Bioinformatic analyses revealed that the LRC_15730/ANHS_518 gene is unique to *L. ruminis* strains, meaning that it has no homologs in any other sequenced species. Iterative PSI-BLASTp searches revealed weak homology to a *B. cereus* ComK transcription factor (ZP_03237802.1, E-value 0.002) after seven iterations.

2.2.5 Characterization of flagellin proteins extracted from motile *Lactobacillus* species.

Flagellin proteins were extracted from *L. ruminis* ATCC27782, *L. ghanensis* L489^T, *L. mali* DSM20444^T and *L. nagelii* DSM13675^T (Figure 2.2). These candidate flagellin proteins ranged in size from ~25 kDa to ~38 kDa (Figure 2.2). Flagellin was not recovered when *L. ruminis* ATCC25644 (Figure 2.2) cells were subjected to the same extraction procedure. Fifteen amino-terminal residues of the *L. ruminis* ATCC27782 ~26 kDa protein product were sequenced, while 10 residues of the other isolated flagellins were sequenced. All sequences were identical to the amino-terminus of the major *L. ruminis* flagellin proteins except for a Val7Ile substitution in *L. mali* and an Ala8Ser substitution in the *L. ghanensis*, *L. mali* and *L. nagelii* proteins. The chemically determined *L. ruminis* ATCC27782 flagellin protein sequence corresponded exactly to that predicted from the highly expressed *fliC2* gene from *L. ruminis* ATCC27782, but also to the predicted (but not

expressed) *fliC1* gene products of both strains. The ~34 kDa protein present in both the ATCC27782 and ATCC25644 protein samples (Figure 2.2, panel A) was also amino-terminally sequenced. It was identified as glyceraldehyde 3-phosphate dehydrogenase (GAPDH), which is commonly found on the surface of Gram–positive bacteria (Saad *et al.*, 2009).

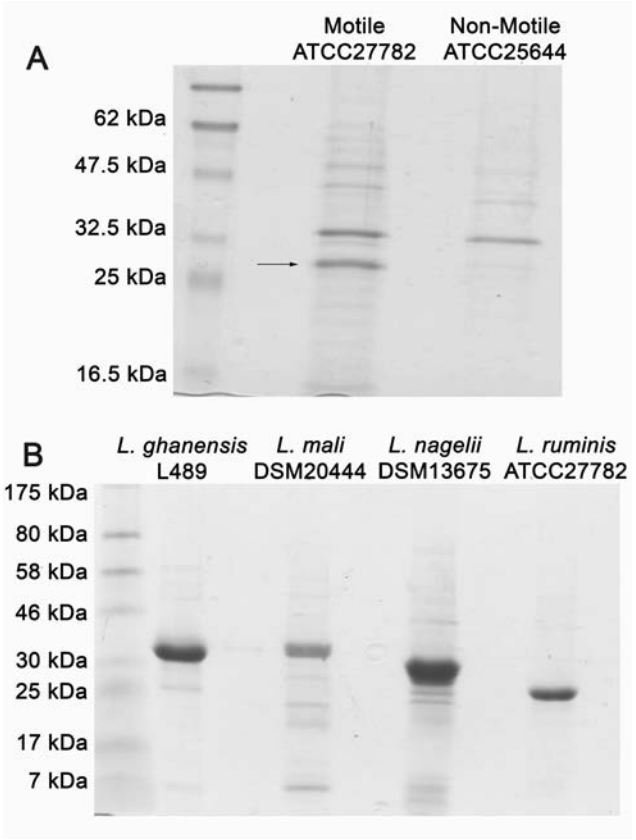


Figure 2.2: Extraction of *Lactobacillus* flagellin proteins and analysis on 10% SDS-PAGE gels. A: Pilot extraction of surface proteins from motile and non-motile *L. ruminis* strains. The arrow points to the ATCC27782 flagellin protein. The larger protein at ~34 kDa is GAPDH. B: Flagellin proteins extracted from various *Lactobacillus* species as indicated.

2.2.6 Flagellate *Lactobacillus* cells, native and recombinant *Lactobacillus* flagellin proteins stimulate IL8 secretion from human intestinal epithelial cell lines.

Interleukin 8 (IL8) production is a central component of the inflammatory response to bacterial flagellin proteins (Gewirtz *et al.*, 2001). Flagellate *L. ruminis* ATCC27782 cells elicited significantly more IL8 secretion from three human colonic epithelial cell lines (T84; HT-29, Caco-2), than the aflagellate ATCC25644 strain (Figure 2.3) ($P \leq 0.01$). In fact, ATCC25644 did not induce significant IL8 secretion from any of the cell lines tested. Furthermore, the flagellin proteins of *L. ghanensis*, *L. mali* and *L. nagelii* also induced significant IL8 secretion from HT-29 cells when compared to the untreated control ($P < 0.01$) (Figure S2.5). Taken together, these data suggested that the flagellin protein, rather than another cell surface component, was responsible for much of the IL8 secretion. To demonstrate that the *Lactobacillus* flagellin-TLR5 interaction was indeed responsible for the IL8 secretion observed, HT-29 cells were transfected with siRNA targeting TLR5 to reduce the expression of this receptor. These cells were stimulated post-transfection with either *L. ruminis* ATCC25644 or ATCC27782 whole cells (Figure 2.3). Knockdown of TLR5 expression was confirmed by qRT-PCR and siRNA treatment did not impact on the viability of the epithelial cells (data not shown). IL8 secretion by HT-29 cells in response to whole flagellate *L. ruminis* ATCC27782 was significantly lower when TLR5 expression was reduced ($P = 0.02$) (Figure 2.3). Thus, *Lactobacillus* flagellin elicits IL8 secretion from intestinal epithelial cells in a TLR5 dependent manner.

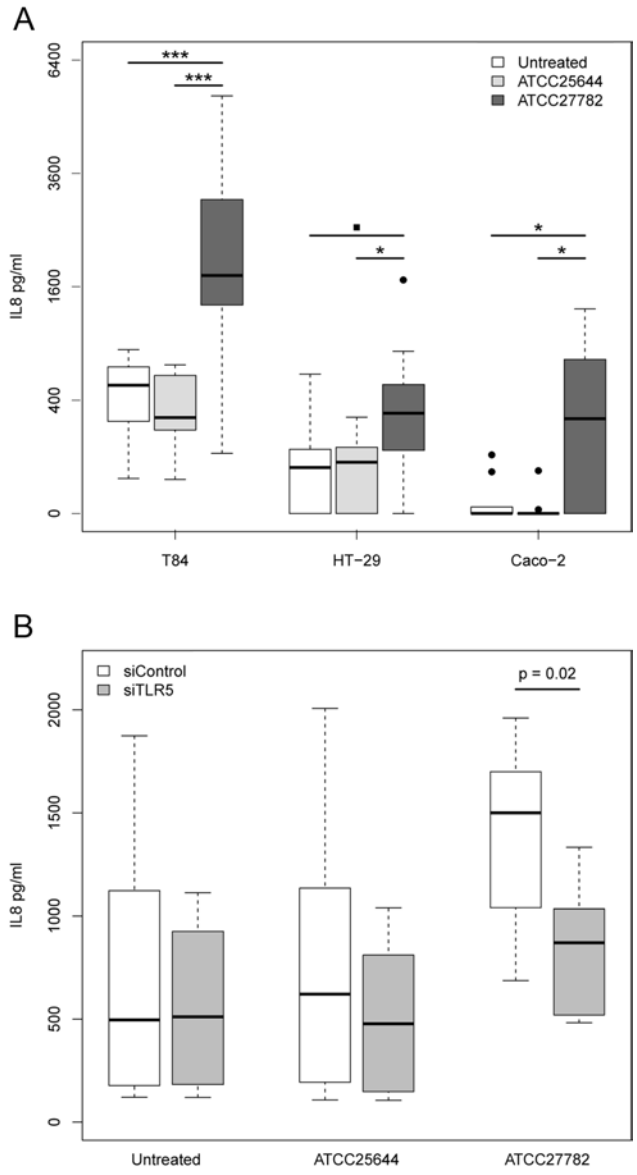


Figure 2.3: Flagellate *L. ruminis* cells or *Lactobacillus* flagellin stimulate IL8 production by intestinal epithelial cells in a TLR5-dependent manner.

A: Flagellate *L. ruminis* ATCC27782 cells induce significant IL8 secretion from human intestinal epithelial cell lines. T84, n = 9; HT-29, n = 7; Caco-2, n = 6. The increments on the Y-axis follow a square-root scale. B: Following transfection with siRNA directed against TLR5, IL8 secretion from HT-29 cells was significantly lower in response to ATCC27782 whole bacteria. Boxplots are based on six experimental replicates. Box-plots show the median and the interquartile range. Solid black dots represent outlier values. Statistical significance was determined using a one-tailed Mann-Whitney U test. Horizontal black bars indicate significant differences between samples. The level of statistical significance is given by P-values over these bars, or by the following symbols: $P < 0.001$ = ***; $P < 0.05$ = *; $P < 0.1$= ■.

A recombinant FliC-GST fusion protein was recovered from *E. coli* inclusion bodies to test if the non-expressed FliC1 protein of the aflagellate *L. ruminis* ATCC25644 strain would also elicit a proinflammatory response from T84 cells. Significantly more IL8 was secreted by T84 cells in response to the recombinant GST-flagellin fusion protein than in response to purified glutathione-S-transferase (GST) protein (P = 0.015), (Figure S2.5). The concentration of IL8 secreted by the T84 cells in response to the fusion protein ranged from 341.22 – 2065.64 pg/ml (n = 7) as determined by ELISA. This large range of IL8 secretion may be attributable to the nature of the fusion protein, which tended to lose its solubility and probably also its functional conformation upon concentration and cold-storage.

Several attempts were also made to express ATCC27782 *fliC1* recombinantly in *E. coli*. However, the *fliC1* and *fliC2* genes of ATCC27782 are 98% identical at a nucleotide level, and 98 % identical at a protein level. As a result, during PCR to amplify the target gene for cloning purposes, chimeric molecules were formed. This meant that the amplified product would encode a flagellin protein which had some unique residues corresponding to *fliC1* and others corresponding to *fliC2*. For this reason, ATCC27782 *fliC1* was not expressed recombinantly.

2.2.7 *L. ruminis* ATCC25644 may be motile *in vivo*.

To determine if *L. ruminis* ATCC25644 might regain motility *in vivo*, isogenic, rifampicin resistant ATCC25644 and ATCC27782 strains were individually fed to mice and were later recovered from their faeces. Following passage through the murine GI tract, a tumbling motility phenotype was imparted to some of the recovered *L. ruminis* ATCC25644 bacteria. Thus, some bacteria in these ATCC25644 cultures displayed tumbling, but not running motility when

grown in MRS broth under standard conditions, as assessed by phase-contrast microscopy (Movie S2.2). These bacteria were confirmed as ATCC25644 by strain specific PCR, 16S rRNA gene sequencing and carbohydrate profiling (data not shown).

To assess the impact of growth medium composition on the motility of ATCC25644, MRS motility agar plates were prepared from basic ingredients to which were added alternative carbon, protein and phosphate sources. None of these modifications imparted motility to ATCC25644 cultures as determined by phase-contrast light microscopy and semi-solid agar plates (Figure S2.6).

The motility of ATCC27782 was similarly evaluated in semi-solid agar in Falcon tubes with various supplements in a further attempt to simulate conditions in the gut that might promote motility. MRS agar supplemented with 1% porcine or bovine bile were tested which rendered the bacteria non-motile, while 0.5 mM EDTA, 1% NaCl, half-strength MRS did not abolish the motility of this strain as assessed by growth in semi-solid agar under anaerobic conditions at 37 °C. None of these conditions induced motility in ATCC25644 in semi-solid agar (data not shown).

Altering the incubation temperature influenced the growth rate of ATCC27782, but did not impact on flagellin production. ATCC27782 that had been incubated in MRS broth at 25 °C, 30 °C, 37 °C and 42 °C achieved optical densities (600 nm) of ~0.1, ~0.2, ~1.36 and ~1.93 respectively after 16 hr incubation in MRS broth, (n = 3). Flagellin could be recovered from cultures that had been incubated at 30 °C, 37 °C and 42 °C as demonstrated by Western blotting (data not shown). Motile bacteria were observed in the cultures incubated at 25 °C, 30 °C and 37 °C after 16 hrs. Cultures grown at 42 °C reached higher cell densities than those grown

at lower temperatures, but lost motility at these high cell densities. Too few motile bacteria were present in the culture grown at 25 °C to allow the recovery of flagellin.

2.3 Discussion.

As the number of *Lactobacillus* species described as being motile has been steadily increasing in recent years, (Table S2.1), the biological significance of "motility" and of "flagella" in this genus merits investigation. From an evolutionary perspective, the almost complete restriction of flagellate species to the *L. salivarius* clade is also noteworthy.

"Motility" could potentially confer a number of competitive advantages to flagellate *Lactobacillus* species. Among the accepted motile species found in the *L. salivarius* clade, most have been isolated from environmental sources (Table S2.1), where the advantages of motility for nutrient acquisition and niche colonisation justify the associated metabolic costs of flagellum production and energization. *L. ruminis*, however, is currently the only motile *Lactobacillus* species known to be autochthonous to the mammalian intestine, (Tannock *et al.*, 2000) and its potential production of flagella and therefore of flagellin, a well characterized MAMP *in vivo* is particularly relevant since it is known that flagellate commensal bacteria could have a significant effect on gut physiology (Sitaraman *et al.*, 2005; Vijay-Kumar *et al.*, 2010).

At least 43 genes are involved in the *Lactobacillus* flagellum-biogenesis and chemotaxis system, and these are chiefly found at a single locus in the *L. ruminis* genome. Although the deletion of genes involved in flagellum synthesis is known to be rapid in an environment where motility confers no evolutionary advantage (Langridge *et al.*, 2009), the *L. ruminis* ATCC25644 (non-motile, human isolate) and ATCC27782 (motile, bovine isolate) motility loci share 97 % nucleotide identity

over their entire length. The absence of gross deletion, despite a typically non-motile phenotype, suggests that *L. ruminis* strains isolated from humans may produce flagella and be restored to full motility under particular environmental condition(s), for example *in vivo*. The historical report of the original ATCC25644 strain being motile would support such a theory, although fully motile ATCC25644 cultures have not been achieved in the laboratory to date. Nevertheless, a motility phenotype of some ATCC25644 cells was partially acquired (tumbling, without runs) following passage through the murine GI tract. Thus, it could be possible that the composition of the MRS growth medium limits the motility of ATCC25644. However, when MRS motility agar plates were prepared with alternative protein, carbon and phosphate sources, full running motility was not imparted to this strain.

Because it was not possible to demonstrate full motility in ATCC25644, and also on the basis of genomic and transcriptomic data, we hereby propose a model to partially explain the genetic basis for the differential motility of *L. ruminis* strains (Figure S2.4). Although the genetic complement of the motility loci of the two *L. ruminis* genomes studied is very similar, the motility locus of *L. ruminis* ATCC27782 (motile) differs from the motility locus of *L. ruminis* ATCC25644 (non-motile) by harboring an extra flagellin gene. Significantly, it is this additional flagellin gene *fliC*2, and not *fliC*1 which is common to both genomes, that is expressed at high levels by the motile strain. We propose that the short palindromic sequence AGATCT (a known ATP-DnaA binding site in *E. coli* (Speck *et al.*, 1999)) that was identified at the *L. ruminis fliC*1 promoter, might serve as a transcription-factor binding site which, when occupied, would prevent transcription from this promoter.

We also suggest that the presumptive regulator ANHS_518/LRC_15730 is under the control of $\sigma^{70'}$ (LRC_04220/ANHS_51c), a sigma factor that is highly expressed in ATCC25644 (Table S2.3). *In silico* examination of the predicted secondary structure of ANHS_518/LRC_15730 identified potential for forming a partial leucine zipper DNA-binding domain at its carboxy-terminus. This protein has no characterized homologs and may represent an atypical class of flagellum regulator that binds to motility gene promoters and operators to repress their transcription. The extent of the proposed negative regulation by ANHS_518/LRC_15730 is unclear, and may include a master regulator, since all of the genes/transcription units in the motility locus of ATCC25644 are down-regulated.

The motility gene content and arrangement of the *L. ruminis* and *L. mali* motility loci are broadly similar. It seems likely that the conserved motility operon structure resulted from selection for efficient flagellum assembly, which is known to be hierarchically regulated at the level of transcription in many species (Smith & Hoover, 2009). Of note is the organization of the motility genes in a single, uninterrupted locus in *L. ruminis*, particularly since the motility genes of *B. subtilis*, (Kunst *et al.*, 1997) and a number of other motile *Firmicutes* are distributed around their respective genomes or their motility loci are interrupted by genes with no known role in motility (Forde et al., in preparation). The inclusion of genes for rhamnose utilization in the motility locus of *L. mali* may indicate that the flagellin proteins of this species are modified by glycosylation, since this sugar has been found as a post-translational modification of flagellin in other bacteria (Schirm *et al.*, 2004; Takeuchi *et al.*, 2007).

Recently, draft *L. vini* (Luckwu de Lucena *et al.*, 2012) and *L. acidipiscis* (Kim *et al.*, 2011; Naser *et al.*, 2006) genomes were published, and these also harbour motility genes. While the motility of *L. vini* has previously been acknowledged (Rodas *et al.*, 2006), *L. acidipiscis* has not been formally described as a motile species (Tanasupawat *et al.*, 2000). Thus, as more *Lactobacillus* genomes are sequenced, the true penetrance of motility among the lactobacilli will emerge. It will also be interesting to determine if motility genes are indeed confined to species of the *L. salivarius* phylogenetic clade which would have significant evolutionary implications.

The observed congruence of motility protein based phylogenetic trees and the 16S rRNA gene tree strongly suggests that the *L. ruminis* motility genes were inherited by vertical descent from the *L. ruminis-Carnobacterium-Enterococcus* last common ancestor. Deeper studies of gene and protein phylogeny would be required to better resolve the more ancestral origins of the *Lactobacillus* motility system.

In the mammalian gut, it has been shown that MAMPs of the commensal luminal bacteria are recognised by the host's pattern recognition receptors (PRR), such as the TLRs (Rakoff-Nahoum *et al.*, 2004). Recognition of the commensal microbiota by host PRRs may contribute to homeostasis and protection from injury in the gut (Rakoff-Nahoum *et al.*, 2004). *Lactobacillus*-TLR engagement may influence expression of these receptors (Vizoso Pinto *et al.*, 2009) and downstream signalling through them (Wells, 2011). Although *Lactobacillus* engagement of TLRs 2, 4 and 9 has been well established in the literature (Lebeer *et al.*, 2008; Wells, 2011), the significance of *Lactobacillus* flagellin-TLR5 signalling had not been studied previously. Thus, the observation of a robust proinflammatory response of human IECs to flagellate *L. ruminis* cells and flagellin proteins is

biologically significant, and the secretion of IL8 is consistent with flagellin-TLR5 interactions for other species (Tallant *et al.*, 2004). Of note however, is a report describing how aflagellate *L. plantarum* BFE 1685 and *L. rhamnosus* GG could increase expression of TLR5 in HT-29 cells through an unknown mechanism (Vizoso Pinto *et al.*, 2009). Since *L. plantarum* and *L. rhamnosus* are both species that may be probiotic, the ability of *L. ruminis* to signal to TLR5 does not necessarily exclude this species from being used as a probiotic. For example, *L. ruminis* could contribute to basal stimulation of the innate immune system (Rakoff-Nahoum *et al.*, 2004) and also sensitize IECs so that they respond better to flagellate pathogens.

In summary, we suggest that the flagellate, mammalian GI commensal *L. ruminis,* may serve as a model species for the study of motility among the lactobacilli, a trait that has been poorly characterized until now. This characterization of the innate immune response to flagellate *Lactobacillus* cells and *Lactobacillus* flagellin proteins has allowed us to predict the *in vivo* responses of the human host to this autochthonous GI bacterium. This immunological information may become particularly relevant if flagellate *Lactobacillus* species are to be developed for probiotic, pharmaceutical, or vaccination applications.

2.4 Materials & Methods.

2.4.1 Bacterial strains, growth conditions and motility evaluation.

Lactobacillus strains (Table S2.2) were cultured anaerobically at 37 °C (*L. ruminis*) or 30 °C (*L. ghanensis, L. mali* and *L. nagelii*) in MRS medium under anaerobic conditions for 48-96 hours. *E. coli* strains were grown aerobically (with agitation) at 37 °C in LB broth supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol.

Standard motility agar plates comprised MRS broth (Oxoid) supplemented with 0.3% - 0.5% (w/v) agarose. MRS broth prepared from first principles followed the composition of standard MRS broth (Oxoid). Glucose was added separately after sterilization. The following substitutions were made to prepare MRS media (500 ml) with alternative protein, phosphate and carbon sources: 9 g of Bactocasitone or Bactopeptone in place of peptone (5 g) and Lab Lemco powder (4 g); 1 g of β-glycerophosphate in place of dipotassium hydrogen phosphate; 10 g of preferred carbohydrate in place of glucose. The motility plates were inoculated with a 10 µl aliquot of an overnight *L. ruminis* culture that had been concentrated by slow centrifugation and resuspension in 1 ml sterile PBS. The inoculum was allowed to dry onto the surface of the plate for 3 min before the plate was moved. Plates were incubated for at least 48 hrs.

Culture motility was evaluated qualitatively by phase-contrast microscopy. Glass capillary tubes were filled with an aliquot of the bacterial culture to be tested. The capillary tube was placed on a heated microscope stage that was maintained at 37 °C for evaluation of culture motility. When every bacterium in a field of view of the microscope was either running or tumbling and moving quickly, the culture was considered “very motile”. When most of the bacteria in a field of view were motile and moving at a moderate speed, the culture was considered “moderately motile”. A culture was considered “weakly motile” when most of the bacteria in a field of view were not motile but some motile cells were also present. If no motile bacteria were observed in the fields of view examined, the culture was considered “non-motile”. Motility could be restored to stationary phase *L. ruminis* cultures of bovine origin by the addition of fresh (1.5 vol) MRS broth followed by anaerobic incubation for 2-3 hours.

2.4.2 Genome sequencing, assembly and comparative genomics.

L. ruminis ATCC25644 and ATCC27782 genomes were each sequenced to at least 21-fold coverage by generation of 200,000 454-FLX reads of average length 125-150 nt, with 3 kb paired-end data and with a half lane of Illumina 3 kb mate-pair and 400 nt paired end libraries. The average Illumina read length was 38 nt. Genomes were assembled using GS Assembler (Roche) and MIRA. The assembly, automated annotation and manual curation of these genomes has been described elsewhere (Forde *et al.*, 2011). These *L. ruminis* genomes have been issued the following GenBank accession numbers: AFYE00000000.1, (ATCC25644) and CP003032.1 (ATCC27782).

The *L. mali* DSM20444^T genome was also subjected to *de novo* assembly using Velvet (version 0.7.58) (Zerbino & Birney, 2008). Two-hundred contigs were assembled into 95 scaffolds with an N50 of 92,201 nt. A paired-end library consisting of 90 bp reads separated by 500 nt assisted the assembly. The draft *L. mali* genome was annotated using the Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) which is available on the NCBI website. An automated annotation of the motility contig (contig1) was separately subjected to manual curation and improvement. Artemis Comparison Tool, (Carver *et al.*, 2005) BLAST (Altschul *et al.*, 1990) were used to compare the *L. ruminis* and *L. mali* motility loci. This *L. mali* sequencing project has been assigned the BioProject number PRJNA84435 and GenBank accession number AKKT00000000.

2.4.3 Flagellin extraction.

Flagella were recovered and flagellin was purified as outlined previously for *C. jejuni* (Logan & Trust, 1983) with modifications as described here. Motility was verified by light-microscopy. The entire culture volume (0.5-4 L) was centrifuged

(1750 x *g*, 20 mins, 4 °C). Cell pellets were re-suspended in 80 ml of cold PBS. Cell-suspensions were homogenized for 30 seconds using a Waring commercial blender (Waring Blendor) at the high speed setting. Homogenization was repeated three times. Cell suspensions were incubated on ice for 30 seconds between blending. Cellular debris was removed by centrifugation, (10,000 *g*, 20 mins, 4 °C). Crude flagellin protein was concentrated by ultracentrifugation, (100,000 *g*, 1 hr, 4 °C). Resulting pellets were re-suspended in 100-200 µl of sterile-distilled water. Protein concentrations were determined using the Pierce BCA protein assay (ThermoScientific) according to manufacturer's instructions. For amino terminal sequencing, flagellin proteins were transferred to Immobilon membrane. The first ten to fifteen amino-terminal residues were sequenced at Aberdeen Proteomics (Aberdeen, Scotland).

2.4.4 Transmission electron microscopy.

Electron microscopy was carried out at the EM facility, University College Cork, and at the University of Birmingham, UK. Specimens were negatively stained with ammonium molybdate solution, 0.25 %-2 %, pH 7.

To harvest and concentrate the bacteria without damaging the cells or the flagella, motile *L. ruminis* ATCC27782 cultures in early exponential phase were passed through a 0.2 µm filter (Millipore). Bacteria were recovered by rinsing the filter with 0.15M NaCl. The bacterial suspension was centrifuged at 100 x *g* for 45 min at room temperature, the supernatant was discarded and the cell pellet was gently re-suspended in 300 µl of 0.15 M NaCl. Non-motile *L. ruminis* ATCC25644 cells, also in early exponential phase, were harvested by centrifugation (660 x *g*, 20 min; room temperature) and the cell pellet was resuspended in 300 µl of 0.15 M NaCl.

Copper grids were either floated on, or immersed in, the bacteria suspension for 30 s to 5 min. The grids were stained with an ammonium molybdate solution (0.25 % - 2 %), pH 7 for 20 s. The 1% and 2% ammonium molybdate solutions were supplemented with a wetting agent (70 µg/ml bacitracin). Before and after staining, excess liquid was removed from the surface of the grid using filter paper. Grids were air-dried and were viewed using a transmission electron microscope (JEOL Transmission Electron Microscope (JEOL Ltd., Tokyo, Japan), (JEM 2000FXII at University College Cork), (JEOL JEM-1200EX at University of Birmingham).

2.4.5 Epithelial cell response to *Lactobacillus* flagella and flagellin proteins.

Human intestinal epithelial cell lines, HT-29 (ATCC HTB-38); Caco-2 (ATCC HTB-37) and T84 (ATCC CCL-248) were cultured at 37 °C in DMEM (Invitrogen) supplemented with 10 % foetal calf serum and 1% penicillin and streptomycin (Stock concentrations of these antibiotics were 10,000 U/ml penicillin, 10 mg/ml streptomycin).

For routine stimulations, human IECs were seeded in triplicate at either 1 x 10^4 or 2 x 10^4 cells per well of a flat-bottomed 96 well tissue-culture plate and were incubated under standard conditions for 24 hr. Then, the IECs were exposed either to *L. ruminis* cells (MOI, 10:1) or to *Lactobacillus* or *Salmonella* flagellin proteins (final concentration: 0.1 µg/ml) for twelve hours. The bacteria and protein stimuli were prepared in DMEM with 10% foetal calf serum and antibiotics. The bacteria were washed twice with PBS before being resuspended in the tissue culture medium.

The amount of IL8 secreted in response to these various stimuli were measured with either the ELISA human IL8 Duo Kit (R&D systems) or MSD (Meso Scale Discovery) immunoassay plates which were used according to the

manufacturer's instructions. Four parameter logistic curves were used to generate standard curves from which the IL8 concentrations of the supernatants harvested from stimulated IECs were derived.

For siRNA experiments, non-polarized HT-29 cells were seeded in triplicate at either 1 x 10^4 or 2 x 10^4 cells per well of a 96 well plate, in antibiotic free DMEM with 10 % foetal calf serum. The plate was incubated under standard conditions overnight.

HT-29 cells were transfected with either siRNA targeting TLR5 or non-targeting control siRNA at a final concentration of 100 nM siRNA per well. The transfection was performed according to the manufacturer's guidelines. Briefly, equal volumes of siRNA and OptiMEM transfection medium were mixed in the same tube. In a separate tube, DharmaFECT transfection reagent was diluted with OptiMEM medium. After 5 min incubation at room temperature, the contents of both tubes were mixed and incubated for 20 min at room temperature. Antibiotic free DMEM with added foetal bovine serum (FBS) was used to adjust the final volume of the transfection medium. A 100 µl aliquot of this transfection medium was added to the HT-29 cells in place of the antibiotic free medium that had bathed the cells overnight. The HT-29 cells were incubated in this transfection medium for 48 hr under standard conditions before stimulation with whole *L. ruminis* cells (2 x 10^5 bacteria/well) resuspended in DMEM with 10 % FBS and 1 % penicillin/streptomycin.

Cell viability assays were performed using the "CellTiter-Glo® Luminescent Cell Viability Assay" (Promega) according to the manufacturer's instructions.

The efficiency of siRNA gene silencing was investigated by qRT-PCR. Epithelial cell lysates were harvested after 12 hr exposure to whole *L. ruminis* cells.

PCR primers and probes were designed using the Universal ProbeLibray Assay Design Centre (https://www.roche-applied-science.com/sis/rtpcr/ulp/adc.jsp) (sequences in Table S2.5). β-actin (ACTB) was used as the calibrator gene. Amplification reactions were prepared with the FastStart TaqMan Probe Master kit (Roche), using 900 nM of each primer in a total volume of 10 µl. Reactions were performed in triplicate on the LightCycler 480 System (Roche) under the following cycling conditions: Pre-incubation: 1 cycle; 95 °C, 10 min. Amplification: 45 cycles; 95 °C, 10 s; 60 °C, 45 s; 72 °C, 1 s. Cooling: 1 cycle; 40 °C, 30 s. Relative changes in gene expression were calculated according to the $2^{-\Delta\Delta Cq}$ method (Livak & Schmittgen, 2001).

2.4.6 Statistical analyses.

A one-tailed Mann-Whitney U test was used for statistical analysis of all cytokine data. Data scaling was applied where appropriate to normalize data from independent experiments. This scaling required the conversion of the IL8 concentration determined for each variable in a replicate experiment to a proportion. This was achieved by dividing the IL8 concentration calculated for each variable in a single experiment by the sum of IL8 secretion for all variables in the same experiment. Statistical tests were applied to these proportions.

2.4.7 Expression and purification of recombinant *L. ruminis* ATCC25644 FliC1.

An *L. ruminis* ATCC25644 *fliC*1 PCR product was engineered to include a SmaI and an XhoI restriction site at its 5′ and 3′ termini respectively. Primer sequences are given in (Table S2.5). The amplified product was restricted and ligated into the pGEX-6-P-3 expression vector (GE Healthcare), immediately downstream of the gene encoding GST. The vector was transformed into *E. coli*

Top Ten for plasmid maintenance and into *E. coli* Rosetta 2(DE3) pLysS (Novagen) for expression.

The target GST-FliC1 fusion protein was isolated from inclusion bodies. Expression of the gene for the recombinant protein was induced by addition of 0.1 mM IPTG (final concentration), to a 50-75 ml *E. coli* Rosetta culture in the exponential growth phase (OD_{600} 0.4). The culture was incubated aerobically at 37°C with agitation for at least 16 hours. Cell pellets were harvested by centrifugation followed by an optional lysozyme 50 mg/ml treatment, and were eventually resuspended in 10 ml wash buffer (1x PBS; 1% Triton-X-100, 1 mM EDTA, pH 7.4). The cell suspension was lysed by passing it through a French press (10,000 psi, twice). The inclusion bodies were recovered from the lysate by centrifugation (15,000 *g*, 30 min, 4°C). The pellet was resuspended in 10 ml wash buffer and the centrifugation step was repeated. The cell pellet was similarly twice washed in chilled water and was incubated overnight at 4°C in solubilization buffer (20 mM TrisHCl, 6 M Urea, 500 mM NaCl, 5 mM β-mercaptoethanol, pH 8). The remaining insoluble material was removed by centrifugation (18,000 *g*, 45 min, 4°C). The supernatant which contained the target protein was refolded by dialysis against a series of solutions containing 0.1 M Tris, 0.5 M arginine and decreasing concentrations (4 M – 0 M) of urea. Dialysis against each solution took place for 24 hours at 4°C. Optionally, post-dialysis, the protein samples were washed with 3 volumes of 0.1 M Tris and concentrated with an Amicon 10K ultra centrifugal filter (Millipore Ltd).

2.4.8 Transcriptome analysis.

L. ruminis ATCC27782 and ATCC25644 were cultured anaerobically at 37°C for 15 hours in 20 ml aliquots of MRS media. Each culture was centrifuged at room temperature to harvest the cells that were immediately resuspended in 500 µl of RNAprotect Bacteria Reagent (Qiagen). Total RNA was isolated according to the protocol for difficult-to-lyse bacteria outlined in the RNAprotect Bacteria Reagent (Qiagen) handbook, but with an extended proteinase K incubation (40 mins). The RNeasy Mini kit (Qiagen) was used to complete the extraction procedure. Contaminating DNA was removed with the Turbo DNA-free kit (Ambion).

cDNA for microarray analysis was prepared by reverse transcribing 10 µg of total RNA using random nonomers (MWG-Biotech, Germany) and the ULS cDNA synthesis and labeling Kit (Kreatech). The details of probe hybridization, (60 °C, 20 hours) microarray scanning and analysis are described elsewhere (Forde *et al.*, 2011). Genes with an expression ratio ≥ 5 and a p-value of $\leq 1.0x10^{-4}$, were considered significantly up or down regulated. The microarray data is available through the GEO website and has been assigned accession number, GSE31556.

For qRT-PCR, RNA was reverse transcribed using Superscript II (Invitrogen) kit and random primers. All reactions were performed in triplicate using the Lightcycler FastStart DNA Master^{plus} SYBR green I kit (Roche) and the LightCycler 480 System (Roche). Relative changes in gene expression were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001).

For several genes, qRT-PCR was performed to verify the result of the microarray analysis. cDNA was purified following overnight reverse transcription using the Illustra CyScribe GFX Purification Kit (GE Healthcare). The purified

cDNA was quantified with a Nanodrop (Thermo Scientific). A 10 ng aliquot of cDNA was used as template per qRT-PCR reaction. Primers were used at a final concentration of 0.2 M per reaction. Primer sequences were designed for *groEL*, *fliM* and LRC_15730 (Table S2.5). The qRT-PCR program was as follows: Pre-incubation: 1 cycle; 95 °C, 10 min. Amplification: 40 cycles; 95 °C, 10 s; 59 °C, 10 s; 72 °C, 10 s. Melting curve: 1 cycle; 95 °C, 5 s; 55 °C, 1 min; 97 °C, continuous. Cooling: 1 cycle; 40 °C, 30 s.

To distinguish transcription from ATCC27782 flagellin genes *fliC*1 (LRC_15700) and *fliC*2 (LRC_15680), RNA from motile ATCC27782 cells was reverse transcribed as previously described. Primer pairs specific to each flagellin gene and *era* were designed (Table S2.5). The qRT-PCR program was as follows: Pre-incubation: 1 cycle; 95 °C, 10 min. Amplification: 45 cycles; 95 °C, 10 s; 65 °C, 5 s;72°C, 5 s. Melting curve: 1 cycle; 95 °C, 5 s; 55 °C, 1 min; 97 °C, continuous. Cooling: 1 cycle; 40 °C, 30 s.

2.4.9 5' Rapid Amplification of cDNA Ends.

The transcription start sites of three target genes were determined using the 5′ RACE 2nd generation kit (Roche) according to the manufacturer's protocol. RNA was extracted from exponential phase *L. ruminis* cultures (OD_{600} ~0.8) as previously described. RNA from ATCC25644 was reverse transcribed using primers specific for the potential negative regulator (ANHS_518) and the sigma$^{70'}$ (ANHS_51c). RNA from ATCC27782 was reverse transcribed using flagellin specific primers (Table S2.5). The cDNA generated was purified and tailed using the High Pure Purification kit (Roche). A series of subsequent PCR reactions, in which the product of an earlier reaction provided template for the next reaction, generated products that

when sequenced, facilitated identification of the desired transcription start sites. DreamTaq (Fermentas) was used for PCR reactions, and the T_A was 50 °C.

2.4.10 Phylogenetic analyses and protein alignments.

Motility proteins from UniProt and NCBI protein database (accession numbers given in Table S2.6) were identified in each species using a combination of existing annotation, BLAST searches (Altschul *et al.*, 1990) and gene order. Protein sequences were aligned using MUSCLE (Edgar, 2004). Any multiple sequence alignment columns that contained at least one gap were removed. Thus the sequences included in the resulting file were all exactly the same length. An appropriate substitution model for tree construction was selected according to the output of Modelgenerator (Keane *et al.*, 2006). Trees were constructed using PHYML (Guindon & Gascuel, 2003) with 100 bootstrap replications. 16S rRNA gene trees were generated with TreeBuilder on the Ribosomal database project website (Cole *et al.*, 2007). A clade was considered to be strongly supported if its bootstrap value was ≥ 90.

2.4.11 Bioinformatic analysis of ANHS518/LRC_15730.

The ANHS_518 and LRC_15730 protein sequences were subjected to BLASTp, PSI-BLASTp analyses on the NCBI website. These sequences were also used as Interproscan and Pfam queries to identify if any conserved domains were present.

2.4.12 Generation of rifampicin tagged *L. ruminis* strains, mouse trial and motility evaluation.

Rifampicin-resistant *L. ruminis* ATCC25644 and *L. ruminis* ATCC27782 variants were produced by serially subculturing these strains in MRS media with increasing concentrations of rifampicin until bacteria resistant to 200 µg/ml

rifampicin were recovered. During the mouse trial, these bacteria were cultured in MRS broth (20 ml) with 200 µg/ml rifampicin daily. The bacterial cultures were centrifuged to harvest the bacteria, which were washed and resuspended in 2 ml sterile PBS. In addition to a standard rodent diet, two groups of five BALB/c mice (6-8 weeks old) were fed a 200 µl aliquot of either rifampicin resistant ATCC25644 or ATCC27782 by oral gavage once daily for five days. Fecal pellets from each group were collected daily. Faecal pellets from each group were pooled, homogenized and serially diluted in PBS. The rifampicin resistant strains were recovered by plating the various dilutions on MRS agar plates containing 50 µg/ml rifampicin. Plates were incubated anaerobically at 37°C for two days. Single colonies recovered from these plates were added to 2 ml MRS broth aliquots containing 50 µg/ml rifampicin and were incubated anaerobically at 37°C for 24 hr, until the cultures became turbid. Glycerol stocks were prepared for each culture. For motility screening, these stocks were used to inoculate MRS broth containing 50 µg/ml rifampicin. The motility of the *L. ruminis* ATCC25644 rifampicin resistant strains that were recovered was evaluated by phase-contrast microscopy and flagellin isolation followed by Western blotting.

For Western blotting, proteins were transferred from an SDS-PAGE gel to PVDF membrane using an EC140 mini-blot apparatus according to the manufacturer's instructions. Briefly, a gel sandwich was prepared using sponges and filter paper soaked in Towbin buffer (25 mM Tris, 192 mM glycine, pH 8.3). The PVDF membrane was activated in methanol before use. Transfer took place at 15 V constant voltage for 1 hour. The membrane was blocked with a solution of 1 % skimmed milk in 1X TBS (20 mM Tris, 0.15 M NaCl, pH 7.6). An anti-*L. ruminis* flagellin primary antibody was used at a concentration of 1 µg/ml. This

custom designed antibody was raised in rabbits by GenScript, USA. The antibody was designed to target the *L. ruminis* flagellin sequence GLTQAKRNAQDGISC. The secondary antibody used was an anti-rabbit IgG peroxidise conjugate. The presence of flagellin was confirmed by colorimetric development of each blot, which was achieved by the addition of 30 mg chloronapthol dissolved in 10 ml methanol to 50 ml 1X TBS with 33 µl hydrogen peroxide. A blue/black result confirmed the presence of flagellin protein.

The identity of the tumbling ATCC25644 strains recovered was confirmed by strain-specific PCR, 16S rRNA sequencing and API carbohydrate utilization profiling.

L. ruminis ATCC25644 specific primer sequences (1054561-1055463; 653884-654800; 626252-627921; 380111-380820; 1654861-1656335; 1338430-1339559; 644444-645201 Table S2.5) were designed to target ATCC25644 genome sequences that were not present in the ATCC27782 genome. PCR was performed using BioTaq (Bioline), using a 50°C annealing temperature, and a one minute extension time.

The ATCC25644 16S rRNA gene product for sequencing was generated by PCR using standard primers (27F and 1492R, Table S2.5). PCR was performed using a 50 °C annealing temperature.

A set of API50CH “Research strips for investigation carbohydrate metabolism in bacteria” (Biomerieux), was inoculated with the tumbling ATCC25644 according to the manufacturers instructions. Carbohydrate utilization was evaluated after 72 hours anaerobic incubation at 37°C.

2.5 Acknowledgements.

The authors thank G. Tannock, G. Felis and D. Nielsen for strains; S. Moore for protein refolding advice; S. Crotty, M. Cotter and C. W. Penn for TEM imaging; J. Cooney for protein structure analysis; S. Melgar for assistance with mouse trial. We thank past and present personnel of UCC microbiology labs 339, 4.11 and 4.25 for advice and assistance.

2.6 Funding Declaration.

This work was supported by a Principal Investigator Award (07/IN.1/B1780) from Science Foundation Ireland to PWOT. BAN was the recipient of an Embark studentship from the Irish Research Council for Science Engineering and Technology. TD and KN were supported by the Alimentary Pharmabiotic Centre, funded by Science Foundation Ireland.

2.7 References.

Al Jassim, R. A. (2003). *Lactobacillus ruminis* is a predominant lactic acid producing bacterium in the caecum and rectum of the pig. *Lett Appl Microbiol* **37**, 213-217.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Andersen-Nissen, E., Smith, K. D., Strobe, K. L., Barrett, S. L., Cookson, B. T., Logan, S. M. & Aderem, A. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc Natl Acad Sci U S A* **102**, 9247-9252.

Andersen-Nissen, E., Smith, K. D., Bonneau, R., Strong, R. K. & Aderem, A. (2007). A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. *J Exp Med* **204**, 393-403.

Buron, N., Coton, M., Desmarais, C., Ledauphin, J., Guichard, H., Barillier, D. & Coton, E. (2011). Screening of representative cider yeasts and bacteria for volatile phenol-production ability. *Food Microbiol* **28**, 1243-1251.

Carr, J. G. & Davies, P. A. (1970). Homofermentative lactobacilli of ciders including *Lactobacillus mali* nov. spec. *J Appl Bacteriol* **33**, 768-774.

Carr, J. G., Davies, P. A., Dellaglio, F., Vescovo, M. & Williams, A. W. (1977). The relationship between *Lactobacillus mali* from cider and *Lactobacillus yamanashiensis* from wine. *J Appl Microbiol* **42**, 219-228.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3423.

Chain, P. S., Grafham, D. V., Fulton, R. S. & other authors (2009). Genomics. Genome project standards in a new era of sequencing. *Science* **326**, 236-237.

Chao, S. H., Tomii, Y., Sasamoto, M., Fujimoto, J., Tsai, Y. C. & Watanabe, K. (2008). *Lactobacillus capillatus* sp. nov., a motile bacterium isolated from stinky tofu brine. *Int J Syst Evol Microbiol* **58**, 2555-2559.

Cole, J. R., Chai, B., Farris, R. J. & other authors (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**, D169-172.

Coton, M., Romano, A., Spano, G., Ziegler, K., Vetrana, C., Desmarais, C., Lonvaud-Funel, A., Lucas, P. & Coton, E. (2010). Occurrence of biogenic amine-forming lactic acid bacteria in wine and cider. *Food Microbiol* **27**, 1078-1085.

Couto, J. A., Campos, F. M., Figueiredo, A. R. & Hogg, T. A. (2006). Ability of lactic acid bacteria to produce volatile phenols. *Am J Enol Vitic* **57**, 166-171.

Dal Bello, F. & Hertel, C. (2006). Oral cavity as natural reservoir for intestinal lactobacilli. *Syst Appl Microbiol* **29**, 69-76.

Delgado, S., Suarez, A., Otero, L. & Mayo, B. (2004). Variation of microbiological and biochemical parameters in the faeces of two healthy people over a 15 day period. *Eur J Nutr* **43**, 375-380.

Dicks, L. M. T. & Endo, A. (2009). Taxonomic status of lactic acid bacteria in wine and key characteristics to differentiate species. *S Afr J Enol Vitic* **30**, 72-90.

Dommels, Y. E., Kemperman, R. A., Zebregs, Y. E., Draaisma, R. B., Jol, A., Wolvers, D. A., Vaughan, E. E. & Albers, R. (2009). Survival of *Lactobacillus reuteri* DSM 17938 and *Lactobacillus rhamnosus* GG in the human gastrointestinal tract with daily consumption of a low-fat probiotic spread. *Appl Environ Microbiol* **75**, 6198-6204.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Edwards, C. G., Collins, M. D., Lawson, P. A. & Rodriguez, A. V. (2000). *Lactobacillus nagelii* sp. nov., an organism isolated from a partially fermented wine. *Int J Syst Evol Microbiol* **50** 699-702.

Endo, A. & Okada, S. (2005). *Lactobacillus satsumensis* sp. nov., isolated from mashes of shochu, a traditional Japanese distilled spirit made from fermented rice and other starchy materials. *Int J Syst Evol Microbiol* **55**, 83-85.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of lactobacilli and bifidobacteria. *Curr Issues Intest Microbiol* **8**, 44-61.

Forde, B. M., Neville, B. A., O'Donnell, M. M., Riboulet-Bisson, E., Claesson, M. J., Coghlan, A., Ross, R. P. & O'Toole, P. W. (2011). Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. *Microb Cell Fact* **10** S13.

Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J. & Madara, J. L. (2001). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* **167**, 1882-1885.

Grundling, A., Burrack, L., Bouwer, H. & Higgins, D. E. (2004). *Listeria monocytogenes* regulates flagellar motility gene expression through MogR, a transcriptional repressor required for virulence. *Proc Natl Acad Sci U S A* **101**, 12318-12323.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704.

Hayashi, F., Smith, K. D., Ozinsky, A. & other authors (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099-1103.

Heilig, H. G., Zoetendal, E. G., Vaughan, E. E., Marteau, P., Akkermans, A. D. & de Vos, W. M. (2002). Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* **68**, 114-123.

Houry, A., Briandet, R., Aymerich, S. & Gohar, M. (2010). Involvement of motility and flagella in *Bacillus cereus* biofilm formation. *Microbiology* **156**, 1009-1018.

Irisawa, T. & Okada, S. (2009). *Lactobacillus sucicola* sp. nov., a motile lactic acid bacterium isolated from oak tree (Quercus sp.) sap. *Int J Syst Evol Microbiol* **59**, 2662-2665.

Jakava-Viljanen, M., Avall-Jaaskelainen, S., Messner, P., Sleytr, U. B. & Palva, A. (2002). Isolation of three new surface layer protein genes (*slp*) from *Lactobacillus brevis* ATCC14869 and characterization of the change in their expression under aerated and anaerobic conditions. *J Bacteriol* **184**, 6786-6795.

Kajikawa, A. & Igimi, S. (2010). Innate and acquired immune responses induced by recombinant *Lactobacillus casei* displaying flagellin-fusion antigen on the cell-surface. *Vaccine* **28**, 3409-3415.

Kajikawa, A., Nordone, S. K., Zhang, L., Stoeker, L. L., LaVoy, A. S., Klaenhammer, T. R. & Dean, G. A. (2011). Dissimilar properties of two recombinant *Lactobacillus acidophilus* strains displaying *Salmonella* FliC with different anchoring motifs. *Appl Environ Microbiol* **77**, 6587-6596.

Kandler, O. & Weiss, N. (1986). Regular, nonsporing Gram positive rods, Bergey's manual of sytematic bacteriology. In *Bergey's manual of sytematic bacteriology*, pp. 1208-1234. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore.

Kaneuchi, C., Seki, M. & Komagata, K. (1988). Taxonomic study of *Lactobacillus mali* Carr and Davis 1970 and related Strains: Validation of *Lactobacillus mali* Carr and Davis 1970 over *Lactobacillus yamanashiensis* Nonomura 1983. *Int J Syst Evol Microbiol* **38**, 269-272.

Kankainen, M., Paulin, L., Tynkkynen, S. & other authors (2009). Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. *Proc Natl Acad Sci U S A* **106**, 17193-17198.

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McLnerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* **6**, 29.

Kim, D. S., Choi, S. H., Kim, D. W., Kim, R. N., Nam, S. H., Kang, A., Kim, A. & Park, H. S. (2011). Genome sequence of *Lactobacillus cypricasei* KCTC 13900. *J Bacteriol* **193**, 5053-5054.

Kimura, K., McCartney, A. L., McConnell, M. A. & Tannock, G. W. (1997). Analysis of fecal populations of bifidobacteria and lactobacilli and investigation of the immunological responses of their human hosts to the predominant strains. *Appl Environ Microbiol* **63**, 3394-3398.

Kimura, K., Nishio, T., Mizoguchi, C. & Koizumi, A. (2010). Analysis of the composition of lactobacilli in humans. *Bioscience Microflora* **29**, 47-50.

Konkel, M. E., Klena, J. D., Rivera-Amill, V., Monteville, M. R., Biswas, D., Raphael, B. & Mickelson, J. (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. *J Bacteriol* **186**, 3296-3303.

Krause, D. O., Smith, W. J., Conlan, L. L., Gough, J. M., Williamson, M. A. & McSweeney, C. S. (2003). Diet influences the ecology of lactic acid bacteria and *Escherichia coli* along the digestive tract of cattle: neural networks and 16S rDNA. *Microbiology* **149**, 57-65.

Kunst, F., Ogasawara, N., Moszer, I. & other authors (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249-256.

Landete, J. M., Ferrer, S. & Pardo, I. (2007). Biogenic amine production by lactic acid bacteria, acetic bacteria and yeast isolated from wine. *Food Control* **18**, 1569-1574.

Lane, M. C., Alteri, C. J., Smith, S. N. & Mobley, H. L. T. (2007). Expression of flagella is coincident with uropathogenic *Escherichia coli* ascension to the upper urinary tract. *Proc Natl Acad Sci U S A* **104**, 16669-16674.

Langridge, G., Phan, M.-D., Turner, D. & other authors (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* **19**, 2308-2316.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. (2008). Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* **72**, 728-764.

Lemon, K. P., Higgins, D. E. & Kolter, R. (2007). Flagellar motility is critical for *Listeria monocytogenes* biofilm formation. *J Bacteriol* **189**, 4418-4424.

Lerche, M. & Reuter, G. (1960). A contribution to the method of isolation and differentiation of aerobic "lactobacilli" (Genus "*Lactobacillus Beijerinck*"). *Zentralbl Bakteriol* **179**, 354-370.

Letran, S., Lee, S., Atif, S., Flores-Langarica, A., Uematsu, S., Akira, S., Cunningham, A. & McSorley, S. (2011). TLR5-deficient mice lack basal inflammatory and metabolic defects but exhibit impaired CD4 T cell responses to a flagellated pathogen. *J Immunol* **186**, 5406-5412.

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408.

Logan, S. M. & Trust, T. J. (1983). Molecular identification of surface protein antigens of *Campylobacter jejuni*. *Infect Immun* **42**, 675-682.

Luckwu de Lucena, B. T., Silva, G. G., Manoel Dos Santos, B. & other authors (2012). Genome sequences of the ethanol-tolerant *Lactobacillus vini* strains LMG 23202T and JP7.8.9. *J Bacteriol* **194**, 3018.

Macnab, R. M. (1996). Flagella and motility. In *Escherichia coli and Salmonella: Cellular and molecular biology*, pp. 123-145. Edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham & other authors. Washington D. C.: ASM Press.

Macnab, R. M. (2003). How bacteria assemble flagella. *Annu Rev Microbiol* **57**, 77-100.

Manes-Lazaro, R., Ferrer, S., Rossello-Mora, R. & Pardo, I. (2008). *Lactobacillus uvarum* sp. nov.--a new lactic acid bacterium isolated from Spanish Bobal grape must. *Syst Appl Microbiol* **31**, 425-433.

Manes-Lazaro, R., Ferrer, S., Rossello-Mora, R. & Pardo, I. (2009a). *Lactobacillus oeni* sp. nov., from wine. *Int J Syst Evol Microbiol* **59**, 2010-2014.

Manes-Lazaro, R., Song, J., Pardo, I., Cho, J. C. & Ferrer, S. (2009b). *Lactobacillus aquaticus* sp. nov., isolated from a Korean freshwater pond. *Int J Syst Evol Microbiol* **59**, 2215-2218.

Maukonen, J., Matto, J., Suihko, M. L. & Saarela, M. (2008). Intra-individual diversity and similarity of salivary and faecal microbiota. *J Med Microbiol* **57**, 1560-1568.

McCracken, A., Turner, M., Giffard, P., Hafner, L. & Timms, P. (2000). Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species. *Arch Microbiol*, 383-389.

Morishita, T., Tamura, N., Makino, T. & Kudo, S. (1999). Production of menaquinones by lactic acid bacteria. *J Dairy Sci* **82**, 1897-1903.

Naser, S. M., Vancanneyt, M., Hoste, B., Snauwaert, C. & Swings, J. (2006). *Lactobacillus cypricasei* Lawson et al. 2001 is a later heterotypic synonym of *Lactobacillus acidipiscis* Tanasupawat et al. 2000. *Int J Syst Evol Microbiol* **56**, 1681-1683.

Nielsen, D. S., Schillinger, U., Franz, C. M., Bresciani, J., Amoa-Awua, W., Holzapfel, W. H. & Jakobsen, M. (2007). *Lactobacillus ghanensis* sp. nov., a motile lactic acid bacterium isolated from Ghanaian cocoa fermentations. *Int J Syst Evol Microbiol* **57**, 1468-1472.

Peel, M., Donachie, W. & Shaw, A. (1988). Temperature-dependent expression of flagella of *Listeria monocytogenes* studied by electron microscopy, SDS-PAGE and western blotting. *J Gen Microbiol* **134**, 2171-2178.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229-241.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Rinttila, T., Kassinen, A., Malinen, E., Krogius, L. & Palva, A. (2004). Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* **97**, 1166-1177.

Rodas, A. M., Chenoll, E., Macian, M. C., Ferrer, S., Pardo, I. & Aznar, R. (2006). *Lactobacillus vini* sp. nov., a wine lactic acid bacterium homofermentative for pentoses. *Int J Syst Evol Microbiol* **56**, 513-517.

Saad, N., Urdaci, M., Vignoles, C., Chaignepain, S., Tallon, R., Schmitter, J. M. & Bressollier, P. (2009). *Lactobacillus plantarum* 299v surface-bound GAPDH: a new insight into enzyme cell walls location. *J Microbiol Biotechnol* **19**, 1635-1643.

Schirm, M., Arora, S. K., Verma, A., Vinogradov, E., Thibault, P., Ramphal, R. & Logan, S. M. (2004). Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J Bacteriol* **186**, 2523-2531.

Seto, A., Saito, Y., Matsushige, M. & other authors (2006). Effective cellulose production by a coculture of *Gluconacetobacter xylinus* and *Lactobacillus mali*. *Appl Microbiol Biotechnol* **73**, 915-921.

Sharpe, M. E., Latham, M. J., Garvie, E. I., Zirngibl, J. & Kandler, O. (1973). Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp. nov. and *Lactobacillus vitulinus* sp. nov. *J Gen Microbiol* **77**, 37-49.

Sitaraman, S. V., Klapproth, J. M., Moore, D. A., 3rd, Landers, C., Targan, S., Williams, I. R. & Gewirtz, A. T. (2005). Elevated flagellin-specific immunoglobulins in Crohn's disease. *Am J Physiol Gastrointest Liver Physiol* **288**, G403-406.

Smith, K. D., Andersen-Nissen, E., Hayashi, F., Strobe, K., Bergman, M. A., Barrett, S. L., Cookson, B. T. & Aderem, A. (2003). Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat Immunol* **4**, 1247-1253.

Smith, T. G. & Hoover, T. R. (2009). Deciphering bacterial flagellar gene regulatory networks in the genomic era. *Adv Appl Microbiol* **67**, 257-295.

Snyder, L. A. S., Loman, N. J., Futterer, K. & Pallen, M. J. (2009). Bacterial flagellar diversity and evolution: seek simplicity and distrust it? *Trends Microbiol* **17**, 1-5.

Speck, C., Weigel, C. & Messer, W. (1999). ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *EMBO J* **18**, 6169-6176.

Stsepetova, J., Sepp, E., Kolk, H., Loivukene, K., Songisepp, E. & Mikelsaar, M. (2011). Diversity and metabolic impact of intestinal *Lactobacillus* species in healthy adults and the elderly. *Br J Nutr* **105**, 1235-1244.

Sun, Z., Kong, J., Hu, S., Kong, W., Lu, W. & Liu, W. (2012). Characterization of a S-layer protein from *Lactobacillus crispatus* K313 and the domains responsible for binding to cell wall and adherence to collagen. *Appl Microbiol Biotechnol* **97**, 1941-1952.

Takeuchi, K., Ono, H., Yoshida, M. & other authors (2007). Flagellin glycans from two pathovars of *Pseudomonas syringae* contain rhamnose in D and L configurations in different ratios and modifided 4-amino-4,6-dideoxyglucose. *J Bacteriol* **189**, 6945-6956.

Tallant, T., Deb, A., Kar, N., Lupica, J., de Veer, M. & DiDonato, J. (2004). Flagellin acting via TLR5 is the major activator of key signaling pathways leading to NF-kappaB and proinflammatory gene program activation in intestinal epithelial cells. *BMC Microbiol* **4**, 33.

Tanasupawat, S., Shida, O., Okada, S. & Komagata, K. (2000). *Lactobacillus acidipiscis* sp. nov. and *Weissella thailandensis* sp. nov., isolated from fermented fish in Thailand. *Int J Syst Evol Microbiol* **50 Pt 4**, 1479-1485.

Tannock, G. W., Munro, K., Harmsen, H. J., Welling, G. W., Smart, J. & Gopal, P. K. (2000). Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20. *Appl Environ Microbiol* **66**, 2578-2588.

Torriani, S., Van Reenen, G. A., Klein, G., Reuter, G., Dellaglio, F. & Dicks, L. M. (1996). *Lactobacillus curvatus* subsp. *curvatus* subsp. nov. and *Lactobacillus curvatus* subsp. *melibiosus* subsp. nov. and *Lactobacillus sake* subsp. *sake* subsp. nov. and *Lactobacillus sake* subsp. *carnosus* subsp. nov., new subspecies of *Lactobacillus curvatus* Abo-Elnaga and Kandler 1965 and *Lactobacillus sake* Katagiri, Kitahara, and Fukami 1934 (Klein et al. 1996, emended descriptions), respectively. *Int J Syst Bacteriol* **46**, 1158-1163.

Vijay-Kumar, M., Aitken, J. D., Carvalho, F. A. & other authors (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228-231.

Vizoso Pinto, M. G., Rodriguez Gomez, M., Seifert, S., Watzl, B., Holzapfel, W. H. & Franz, C. M. (2009). Lactobacilli stimulate the innate immune response and

modulate the TLR expression of HT29 intestinal epithelial cells *in vitro*. *Int J Food Microbiol* **133**, 86-93.

Walter, J. (2005). The microecology of lactobacilli in the gastrointestinal tract. In *Probiotics and prebiotics: scientific aspects*, pp. 51-82. Edited by G. W. Tannock. Norfolk, U.K: Caister Academic Press.

Weiss, N., Schillinger, U., Laterneser, M. & Kandler, O. (1981). *Lactobacillus sharpeae* sp. nov and *Lactobacillus agilis* sp. nov., two new species of homofermentative, meso-diaminopimelic acid containing lactobacilli isolated from sewage. *Zentralbl Bakteriol Mikrobiol Hyg l Abt C* **2**, 242-253.

Wells, J. (2011). Immunomodulatory mechanisms of lactobacilli. *Microb Cell Fact* **10**.

Yin, Q. & Zheng, Q. (2005). Isolation and identification of the dominant *Lactobacillus* in gut and faeces of pigs using carbohydrate fermentation and 16S rDNA analysis. *J Biosci Bioeng* **99**, 68-71.

Yoon, S. I., Kurnasov, O., Natarajan, V., Hong, M., Gudkov, A. V., Osterman, A. L. & Wilson, I. A. (2012). Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859-864.

Yuan, J., Fahrner, K. A., Turner, L. & Berg, H. C. (2010). Asymmetry in the clockwise and counterclockwise rotation of the bacterial flagellar motor. *Proc Natl Acad Sci U S A* **107**, 12846-12849.

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829.

Zhou, K., Kanai, R., Lee, P., Wang, H. & Modis, Y. (2012). Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. *J Struct Biol* **177**, 402-409.

2.8 Supporting Information

Movie S2.1: Microscopic observation of motile *L. ruminis* ATCC27782 cells. Bacteria were grown in MRS broth and were visualised using a light-microscope. Video was recorded with a USB eyepiece camera. Motile single cells and chains of motile cells are visible.

Movie S2.2: **Microscopic observation of tumbling *L. ruminis* ATCC25644 cells in MRS broth.** Video was recorded with a USB eyepiece camera attached to a phase-contrast microscope.

Movies available at the following website:
www.plosone.org/article/infor:doi/10.1371/journal.pone0040592

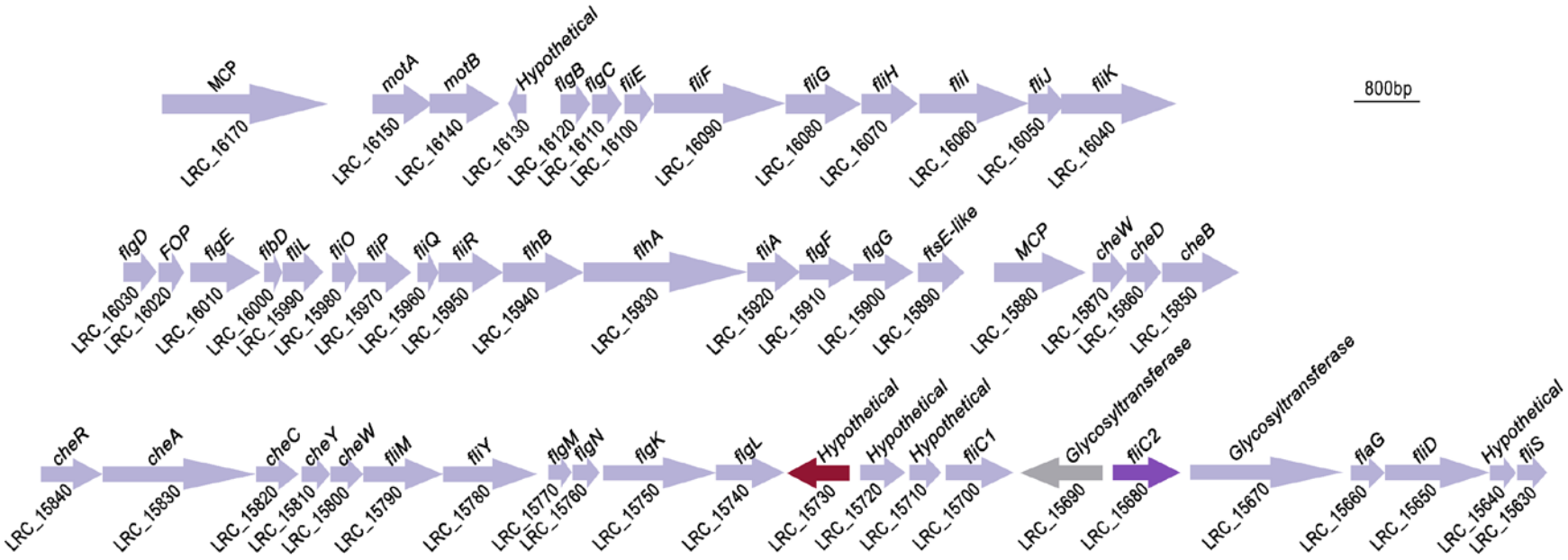


Figure S2.1: Genetic organization of the *L. ruminis* ATCC27782 motility locus.

Motility genes are arranged contiguously and span 48 Kb of the genome. MCP = Methyl accepting chemotaxis protein. Locus tags are given below each gene arrow. The glycosyltransferase (LRC_15690) colored gray is frameshifted. The additional *fliC*2 gene (LRC_15680) present in ATCC27782, but absent from the ATCC25644 motility locus is colored purple. The LRC_15730 gene, (shown in red here), is a homolog of ANHS_518, the only gene at the *L. ruminis* ATCC25644 motility locus to be differentially transcribed in the non-motile strain.

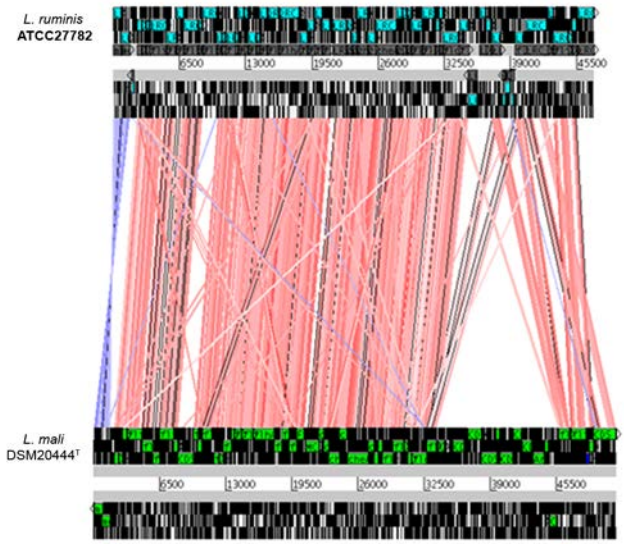
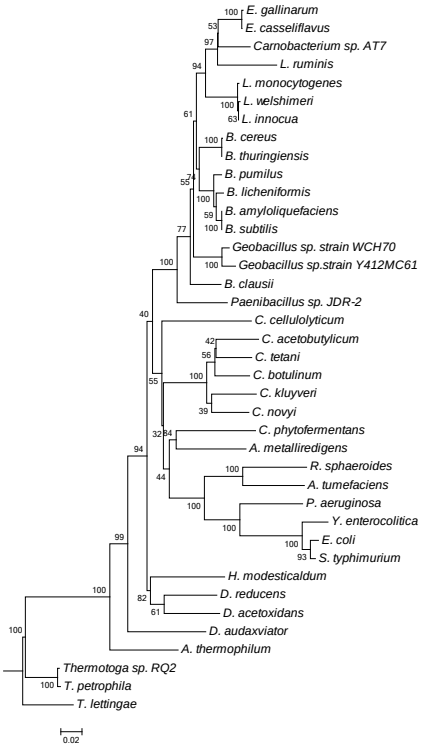
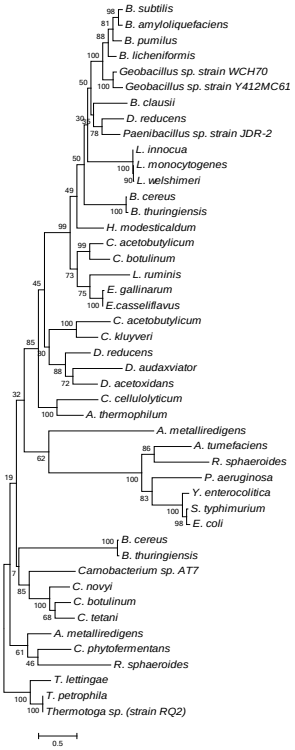


Figure S2.2: Artemis Comparison Tool (ACT) alignment of *L. ruminis* ATCC27782 and *L. mali* DSM20444^T motility loci.

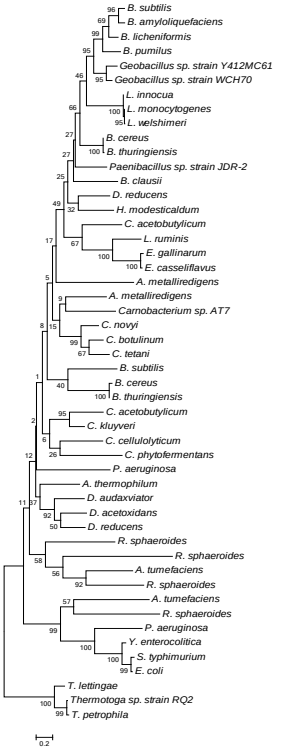
The *L. ruminis* motility locus (top) is aligned to the *L. mali* motility locus (bottom) using tBLASTx. Red lines indicate regions of similar sequence with the same orientation in both genomes. Blue lines indicate regions of similarity sequence that have an inverted configuration. A large, 11.8 kb insertion in the *L. mali* locus relative to the *L. ruminis* motility locus is also evident.



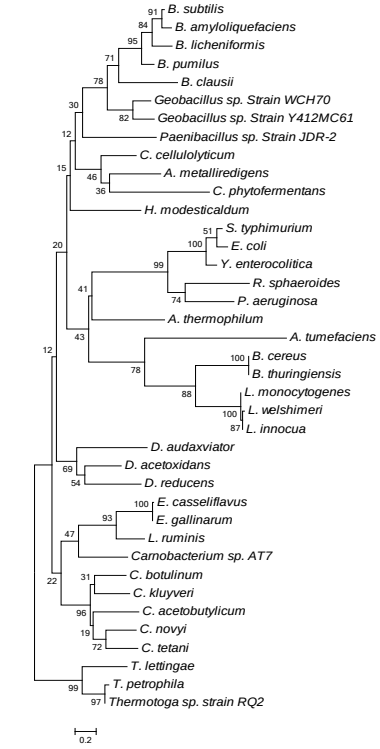
16S rRNA



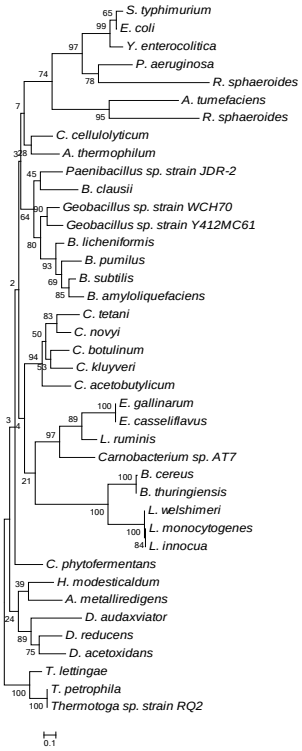
MotA



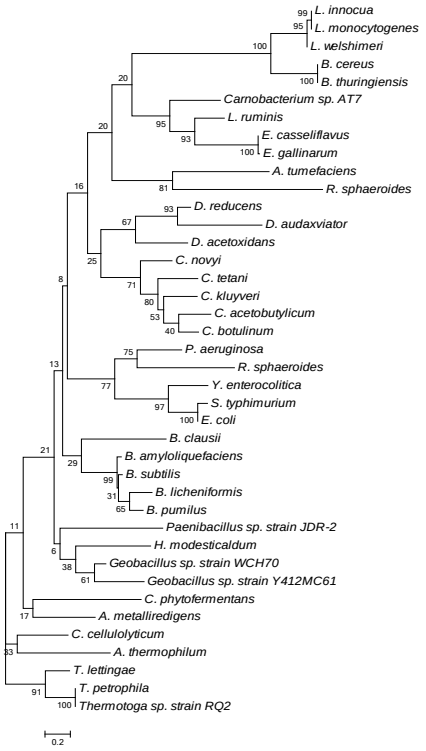
MotB



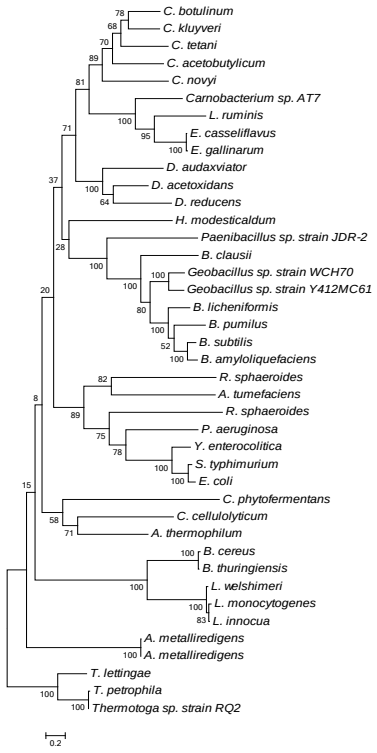
FlgB



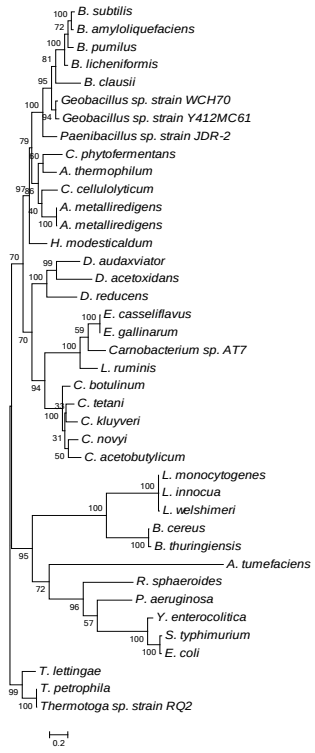
FlgC



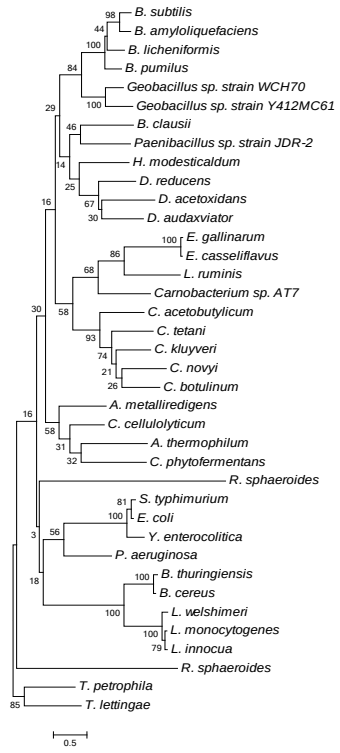
FliE



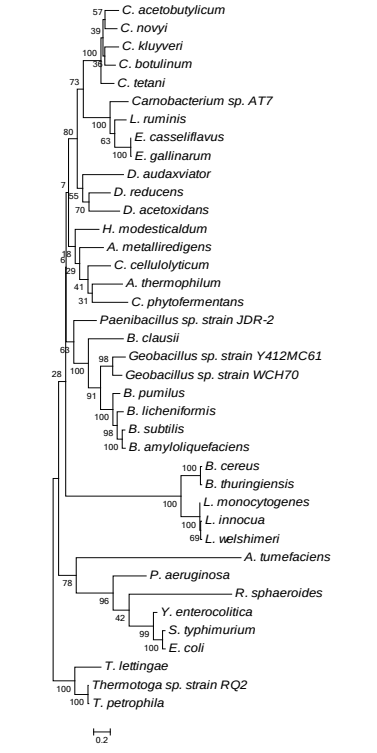
FliF



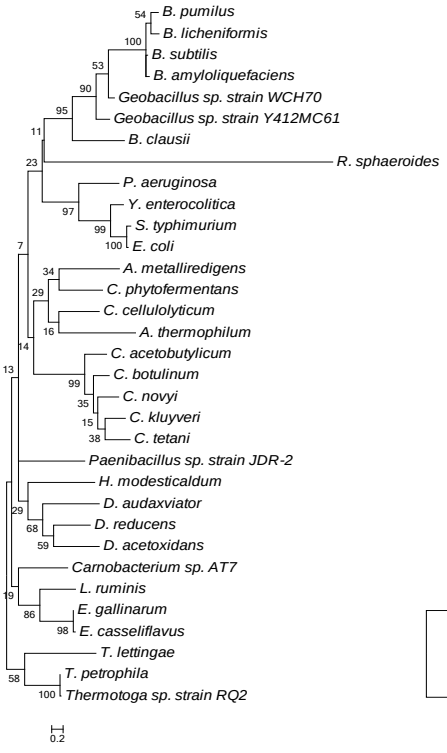
FliG



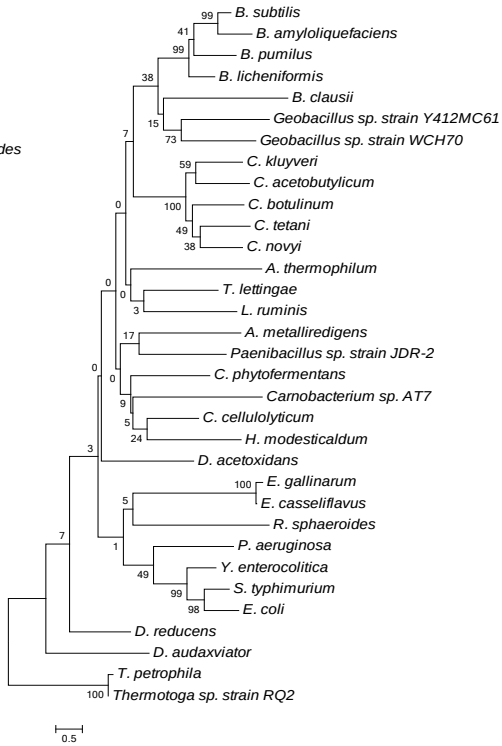
FliH



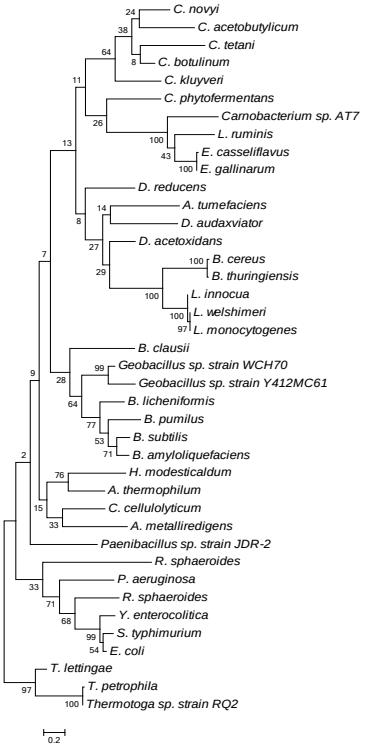
FliI



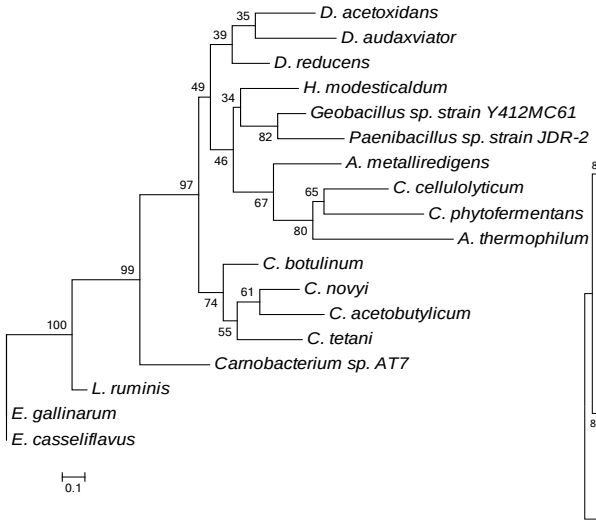
FliJ



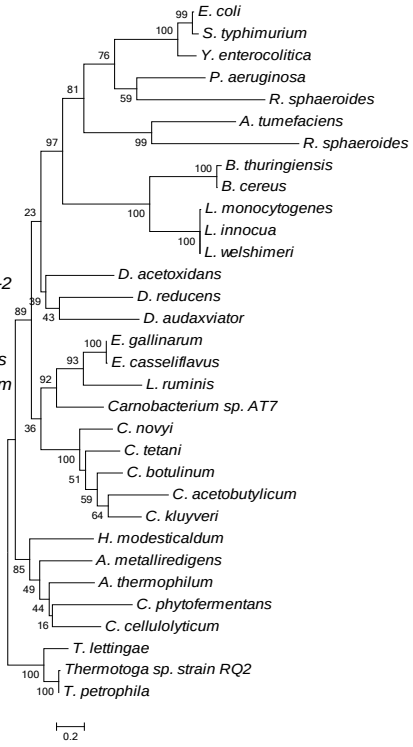
FliK



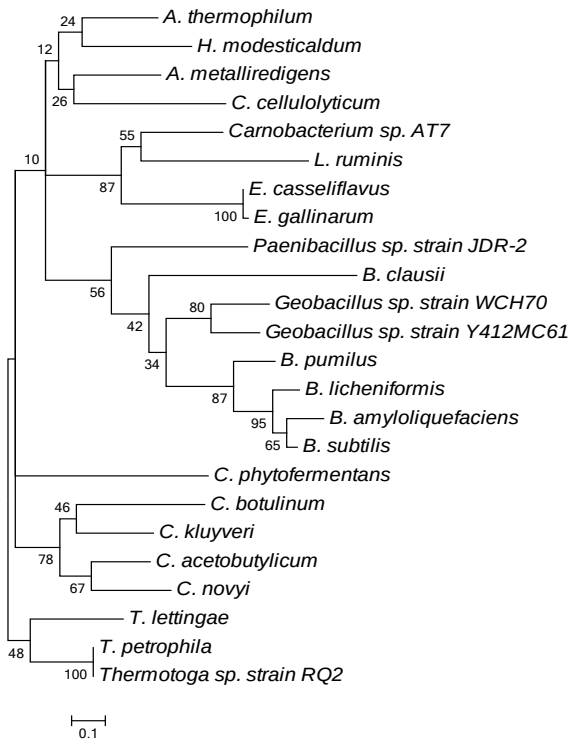
FlgD



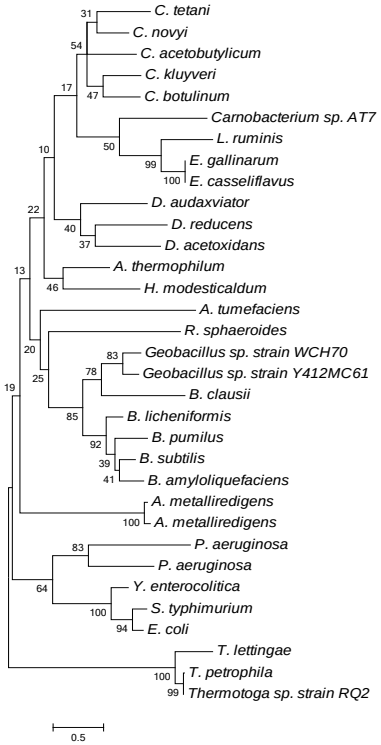
Flagellar Operon Protein



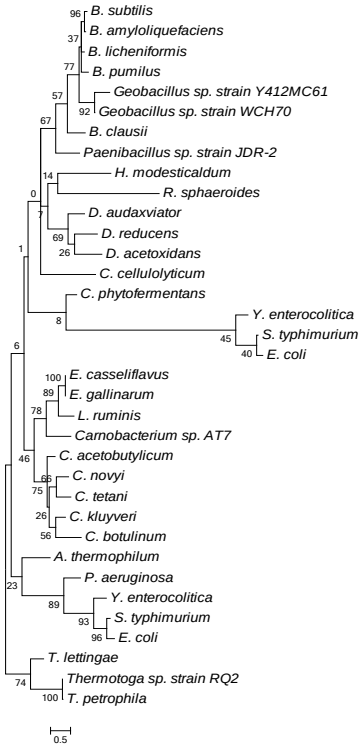
FlgE



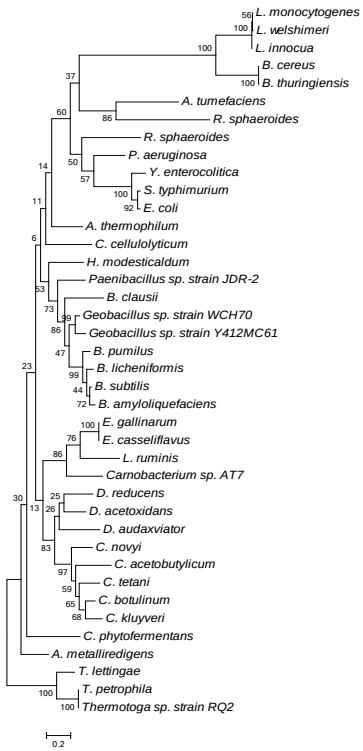
FlbD



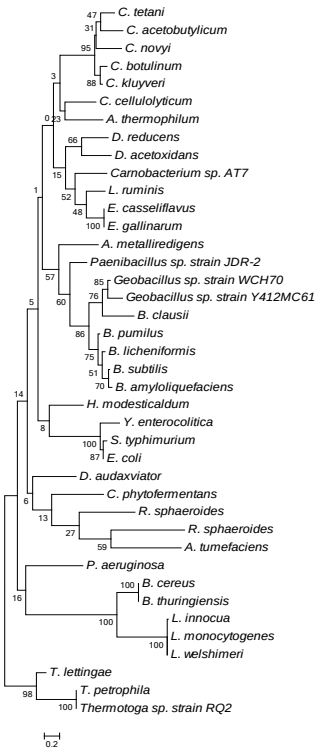
FliL



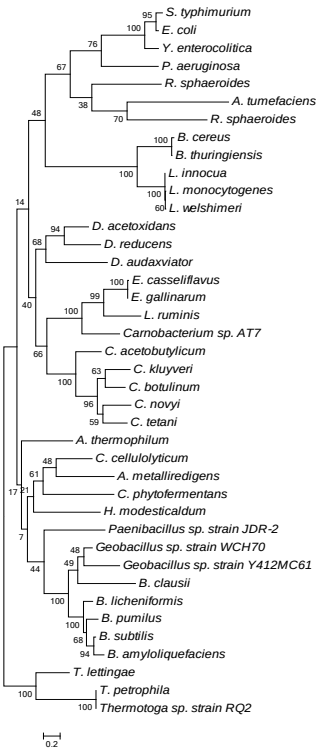
FliO/Z



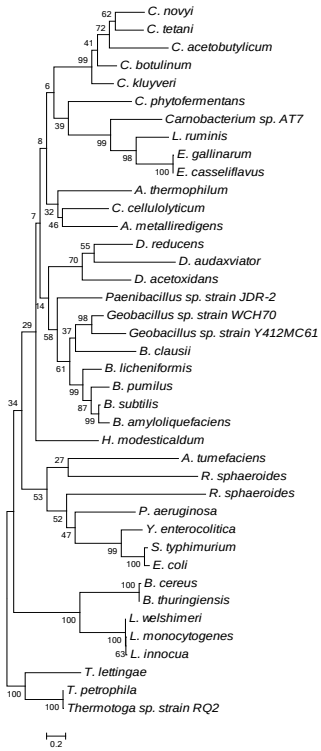
FliP



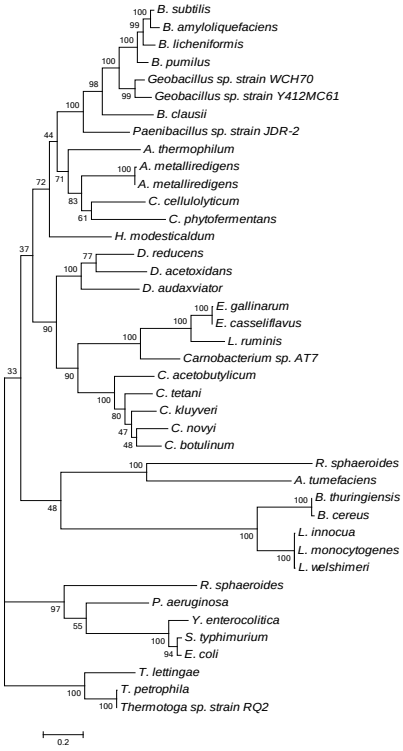
FliQ



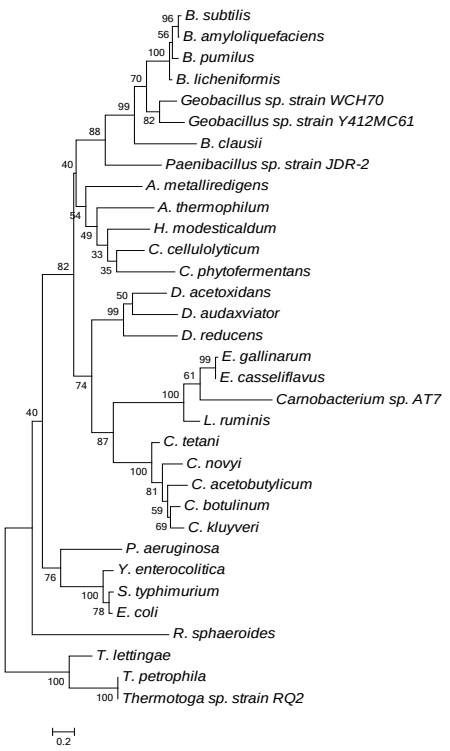
FliR



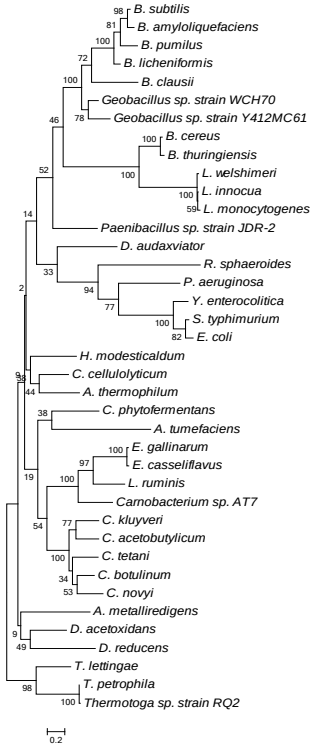
FlhB



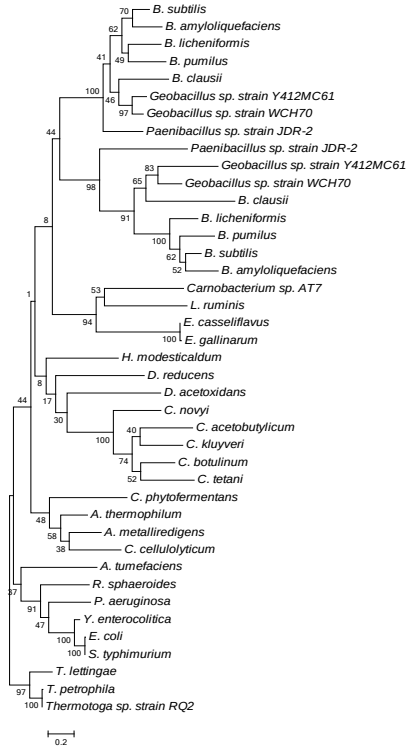
FlhA



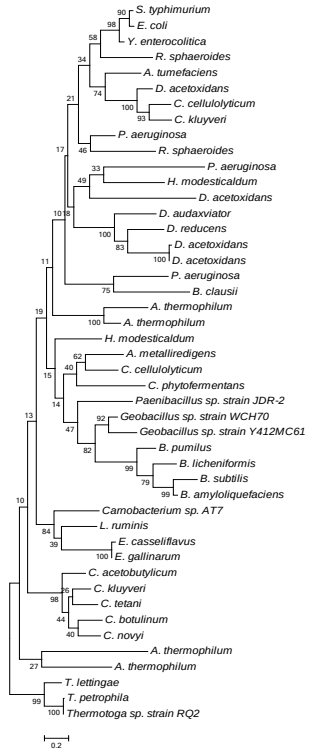
FliA



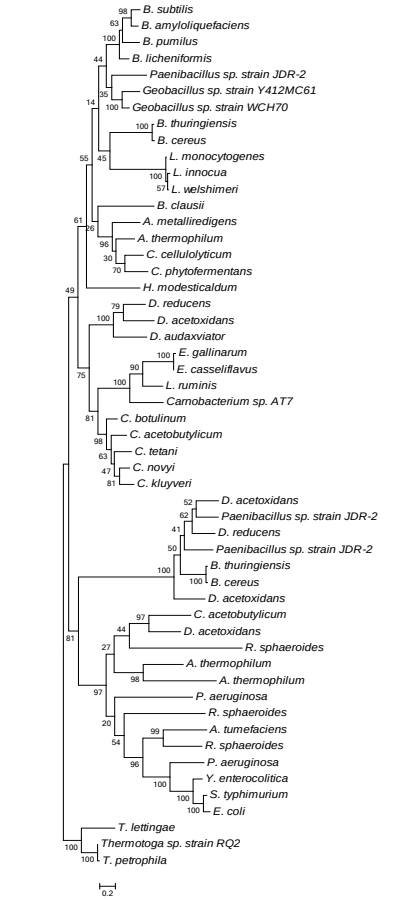
FlgF



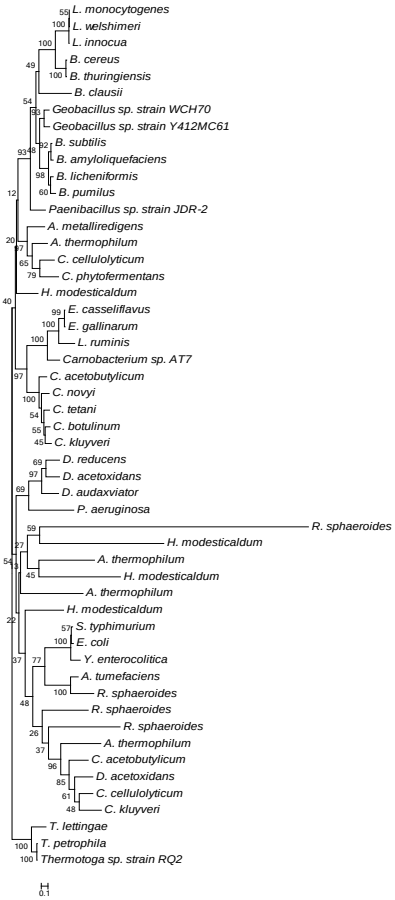
FlgG



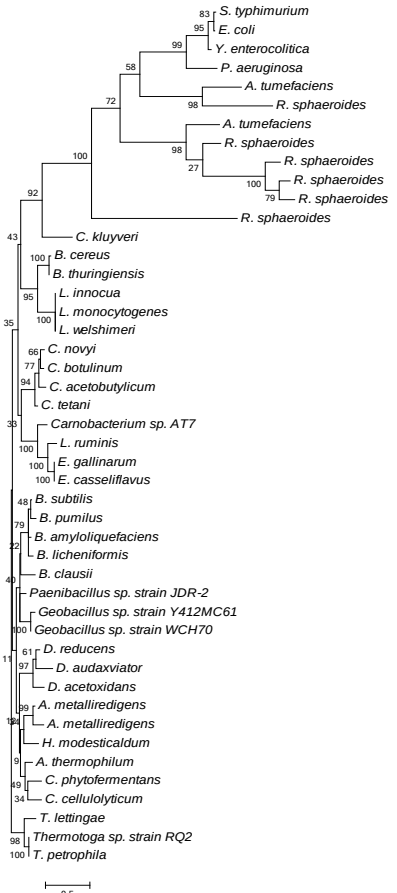
CheB



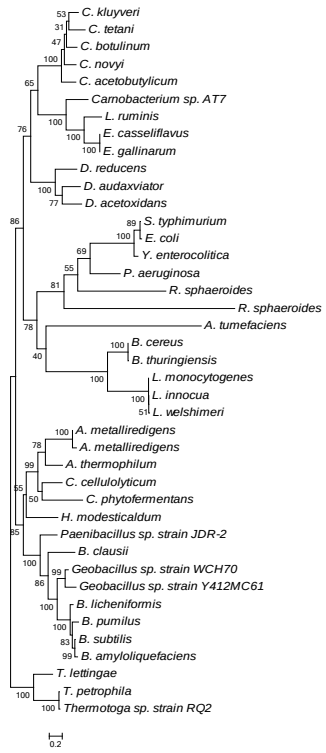
CheR



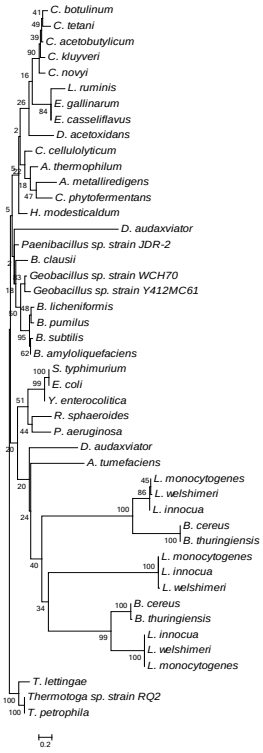
CheA



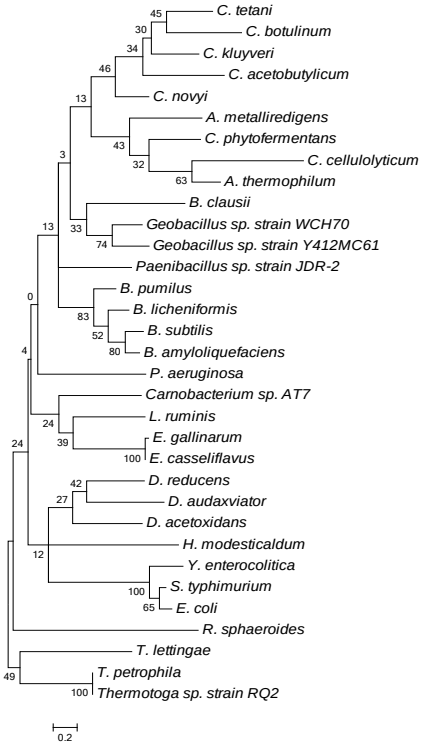
CheY



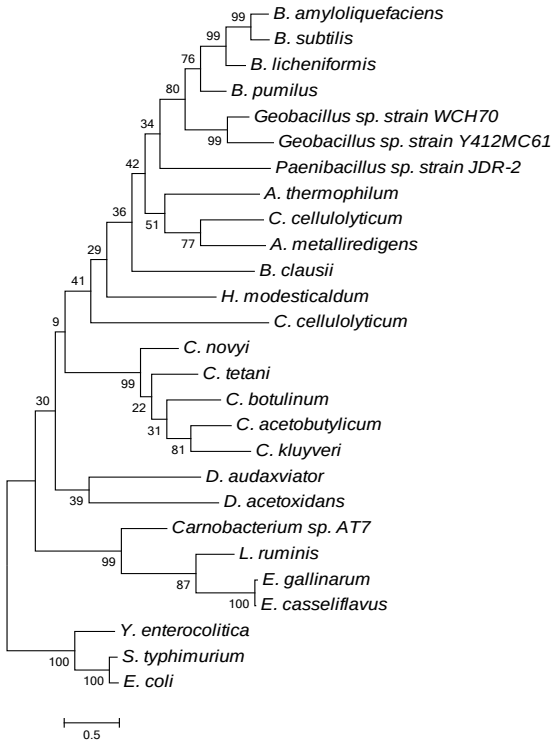
FliM



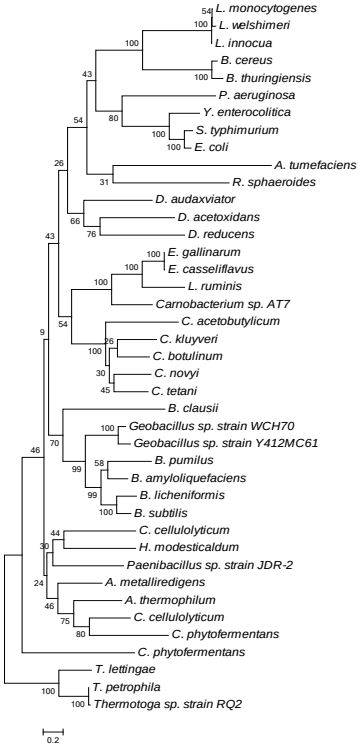
FliN/Y



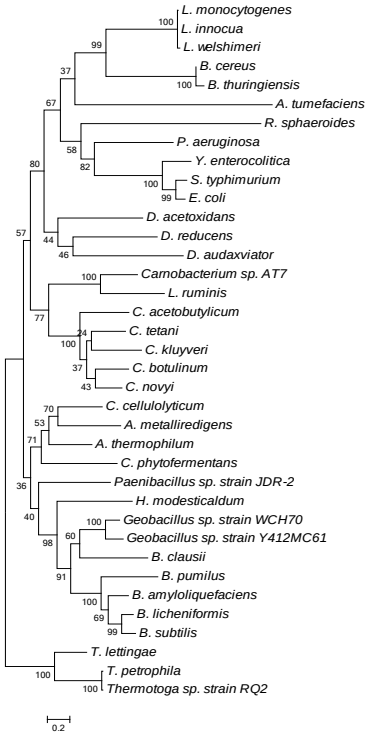
FlgM



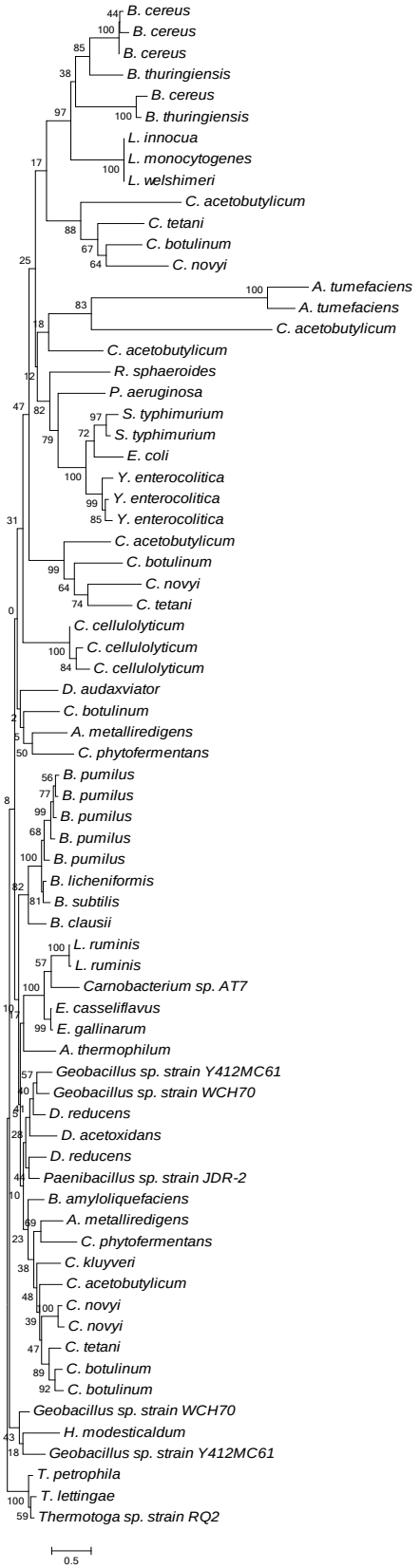
FlgN



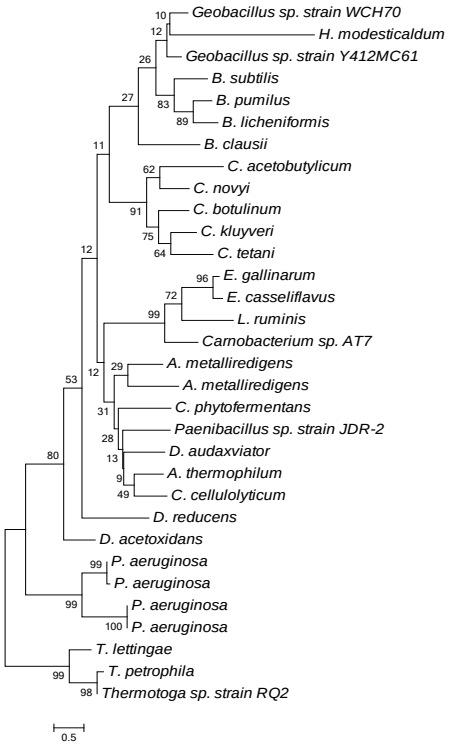
FlgK



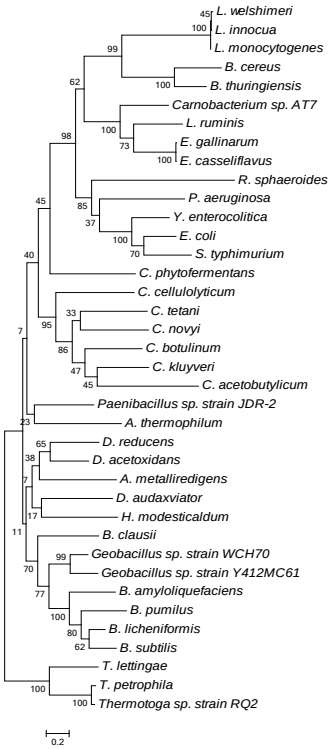
FlgL



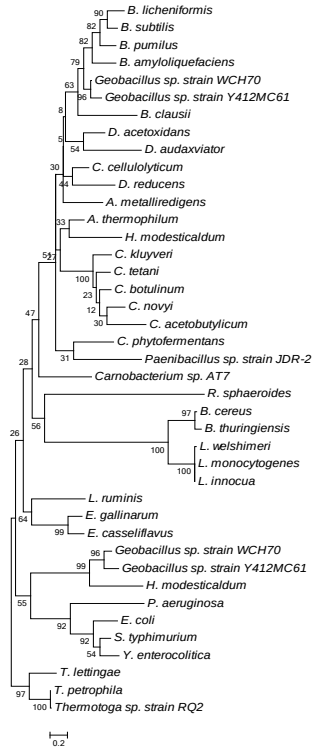
FliC



FlaG



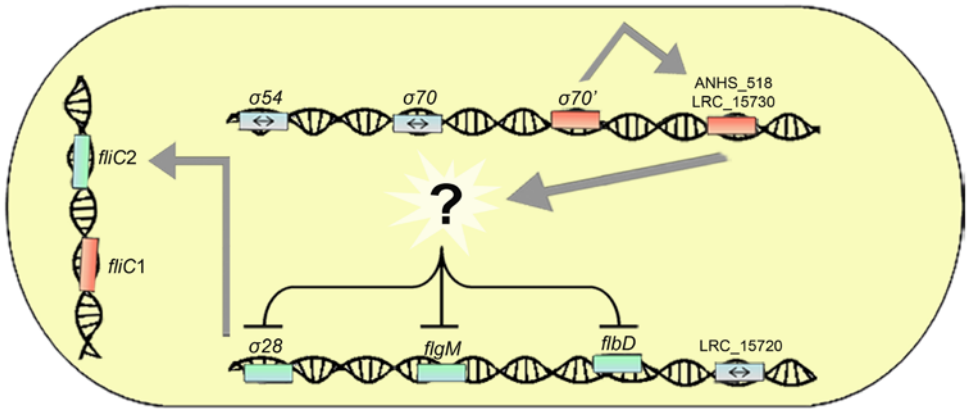
FliD



FliS

Figure S2.3: 16S rRNA gene tree and motility protein based phylogenetic trees.

Trees were constructed using PHYML. Bootstrap values are given at each node.



ATCC27782 *fliC1* LRC_15700:
5' - TTAAAAGATCTGAGATGCTGACGATATATAATCATGTAATCAGTTAATGATACAAAAGAAATGGAGGAATTTAACATG - 3'

ATCC27782 *fliC2* LRC_15680:
5' - TTAAACAAATTGTCGTGTTGCCGATATATAATCATGTAATCAGTTAATGATACAAGAGAATGGAGGAATTTAACATG - 3'

ATCC25644 flagellin ANHS_515c:
5' - TTAAAAGATCTGAGATGCTGACGATATATAATCATGTAATCAGGTAATGATACAAGAGAAATGGAGGAATTTAATATG - 3'

ATCC25644 negative regulator ANHS_518:
5' - TTGATTTAATTTAAATTACATTTTAAACTGAAATTGTTCTTTTCCAAAAGAACGCCAAAATTTAACTGTTATTGGGGGAATTATATG- 3'

Figure S2.4: Proposed model for regulation of flagellum biogenesis.

Sigma 70′ is responsible for transcription of ANHS_518/LRC_15730, which acts via an unknown mechanism to inhibit transcription of the genes regulating flagellum biogenesis. Genes shown in red and green are upregulated in ATCC25644 and ATCC27782 respectively during exponential phase, when ATCC27782 is motile. Genes shown in blue are not differentially transcribed. The nucleotide sequence immediately upstream of *L. ruminis* flagellin genes and the putative negative regulator, ANHS_518 are shown. Start codons are coloured green. Predicted -10 and -35 sequences are blue. The +1 transcription start sites mapped by 5' RACE are orange. Likely ribosome binding sites are underlined.

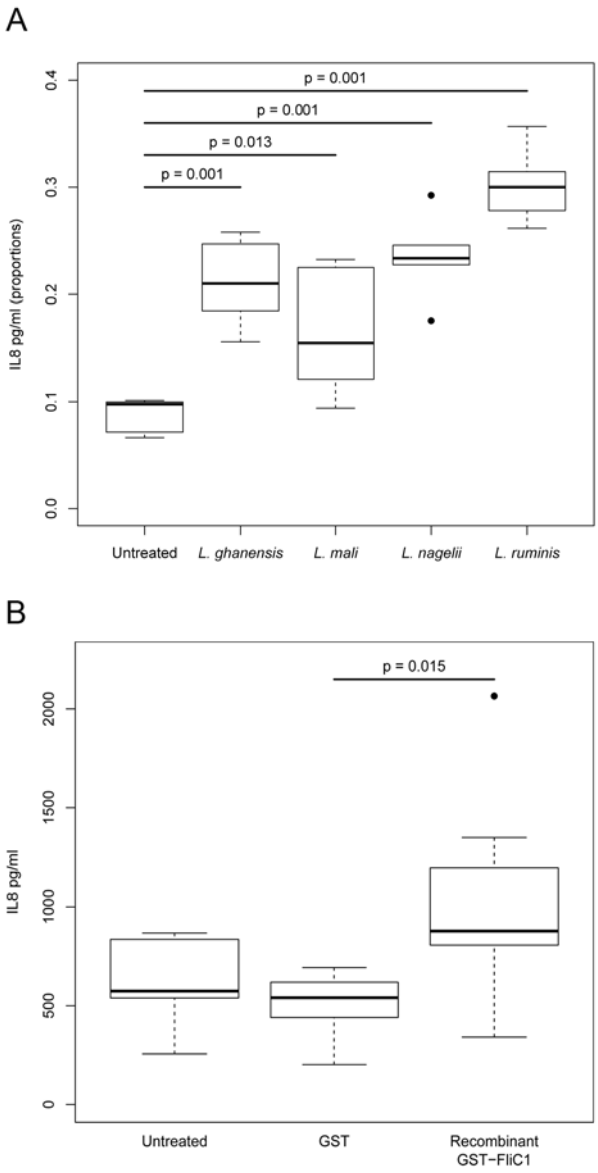


Figure S2.5: Characterization of the immune responses induced by native and recombinant *Lactobacillus* flagellin proteins in epithelial cell lines.

A: HT-29 cells secrete IL8 in response to the native flagellin proteins of various *Lactobacillus* species as indicated. Flagellin was added at a final concentration of 0.1 μg/ml. Boxplots show the median values and interquartile range based on data from six experimental replicates. A one-tailed Mann-Whitney U test was applied to calculate statistical significance. The data upon which this graph is based were converted to proportions as described in Materials and Methods.

B: T84 cells secrete IL8 in response to the recombinant GST-ATCC25644 flagellin protein. Graphs show median values and interquartile ranges. For "untreated" and "GST", n = 5. For the recombinant protein, n = 7. A one-tailed Mann-Whitney U test was applied to calculate statistical significance.

Figure S2.6: Failure of altered growth media to impart motility to *L. ruminis* ATCC25644 on semi-solid agar plates. A, B: Standard (A) or half-strength (B) MRS (Oxoid) inoculated with ATCC25644. C, D: MRS prepared from first principles with alternative phosphate sources, β-glycerophosphate (C) or K2HPO4 (D) and inoculated with ATCC25644. E, F: MRS prepared from first principles with alternative protein sources, Bactocasitone (E) and Bactopeptone (F) and inoculated with ATCC25644. G, H: MRS prepared from first principles with alternative carbohydrate sources, P95 (G) and Synergy I (H) and inoculated with ATCC25644. I, J: MRS prepared from first principles with uracil (0.005 g/500 ml) inoculated with ATCC25644 (I) or ATCC27782 (J). These photographs are of motility plates that were inoculated on different days. *L. ruminis* ATCC25644 is not motile under any of these conditions. *L. ruminis* ATCC27782 (J) is representative of a motile culture on semi-solid agar, and can be seen to cover the entire plate surface.

1

Table S2.1: Origin and phylogeny of motile *Lactobacillus* species described to date.

Year of isolation	Species	Type strain	Source	Clade [A]	Reference to species motility
1970	*L. mali*	DSM20444	Wine must, cider	*L. salivarius*	(Carr & Davies, 1970)
1973	*L. ruminis*	ATCC27780	Mammalian faeces, Bovine rumen	*L. salivarius*	(Al Jassim, 2003; Heilig *et al.*, 2002; Sharpe *et al.*, 1973; Yin & Zheng, 2005)
1982	*L. agilis*	DSM20509	Municipal sewage	*L. salivarius*	(Weiss *et al.*, 1981)
2000	*L. nagelii*	ATCC700692	Partly fermented grape juice	*L. salivarius*	(Edwards *et al.*, 2000)
2005	*L. satsumensis*	DSM16230	Mashes of shochu	*L. salivarius*	(Endo & Okada, 2005)
2006	*L. vini*	DSM20605	Fermenting grape musts	*L. salivarius*	(Rodas *et al.*, 2006)
2007	*L. ghanensis*	DSM18630	Fermenting cocoa	*L. salivarius*	(Nielsen *et al.*, 2007)
2008	*L. capillatus*	DSM19910	Fermented stinky tofu brine	*L. salivarius*	(Chao *et al.*, 2008)
2008	*L. uvarum*	DSM19971	Grape musts	*L. salivarius*	(Manes-Lazaro *et al.*, 2008)
2009	*L. oeni*	DSM19972	Bobal grape wines	*L. salivarius*	(Manes-Lazaro *et al.*, 2009a)
2009	*L. aquaticus*	DSM21051	Korean freshwater pond	*L. salivarius*	(Manes-Lazaro *et al.*, 2009b)
2009	*L. sucicola*	DSM21376	Sap of oak tree	*L. salivarius*	(Irisawa & Okada, 2009; Manes-Lazaro *et al.*, 2009b)

A. According to phylogeny of Felis *et al*., 2007.

Table S2.2: *Lactobacillus* strains and species used in this study.

Species	Strain	Origin	Motility [A]	Motility genes [B]
L. ruminis	ATCC 25644	Human	-	+
L. ruminis	ATCC 27780^T	Bovine rumen	+	+
L. ruminis	ATCC 27781	Bovine rumen	+	+
L. ruminis	ATCC 27782	Bovine rumen	+	+
L. ruminis	L5	Human	-	+
L. ruminis	S21	Human	-	+
L. ruminis	S23	Human	-	+
L. ruminis	S36	Human	-	+
L. ruminis	S38	Human	-	+
L. ghanensis	L489^T	Cocoa bean fermentations	+	+
L. mali	DSM20444^T	Apple juice from cider press	+	+
L. nagelii	DSM13675^T	Partially fermented wine	+	+

A, B: Data from this study.

Table S2.3: Closest homologs* of *L. ruminis* motility proteins.

Annotation	Locus Tag	Top non *L. ruminis* BLAST hit*	BLASTp top hit, excluding motile lactobacilli, ATCC27782 Query	% Identity ATCC27782
MCP	LRC_16170	Methyl-accepting chemotaxis sensory transducer	*Enterococcus casseliflavus EC30*, ZP_05646024.1	37%
motA	LRC_16150	MotA	*Enterococcus casseliflavus* ATCC12755, ZP_08144203.1	52%
motB	LRC_16140	MotB	*Enterococcus casseliflavus* EC20, ZP_05655638.1	45%
Hypothetical protein	LRC_16130	-	-	-
flgB	LRC_16120	FlgB	*Enterococcus saccharolyticus* 30_1, ZP_09112718.1	56%
flgC	LRC_16110	FlgC	*Enterococcus casseliflavus* ATCC12755, ZP_08144200.1	65%
fliE	LRC_16100	Predicted protein	*Enterococcus gallinarum* EG2, ZP_05650941.1	51%
fliF	LRC_16090	FliF	*Enterococcus gallinarum* EG2, ZP_05650942.1	44%
fliG	LRC_16080	FliG	*Enterococcus gallinarum EG2*, ZP_05650943.1	69%
fliH	LRC_16070	Predicted protein	*Enterococcus casseliflavus* EC30, ZP_05646033.1	20%
fliI	LRC_16060	Flagellum specific ATP synthase	*Enterococcus casseliflavus* ATCC12755, ZP_08144195.1	77%
fliJ	LRC_16050	Putative flagellar protein	*Enterococcus casseliflavus* ATCC12755, ZP_08144194.1	44%
fliK	LRC_16040	Hypothetical protein	*Enterococcus saccharolyticus* 30_1, ZP_09112710.1	44%
flgD	LRC_16030	Predicted protein	*Enterococcus gallinarum* EG2, ZP_05650948.1	43%
Flagellar operon protein	LRC_16020	Predicted protein	*Enterococcus casseliflavus* EC30, ZP_05646038.1	58%
flgE	LRC_16010	Flagellar hook protein	*Enterococcus casseliflavus* EC30, ZP_05646039.1	56%
flbD	LRC_16000	Putative flagellar protein	*Carnobacterium* sp. 17-4, YP_004375661.1	51%
fliL	LRC_15990	Hypothetical protein	*Enterococcus casseliflavus* ATCC12755, ZP_08144188.1	45%
fliO	LRC_15980	FliZ	*Enterococcus casseliflavus* ATCC12755, ZP_08144187.1	40%
fliP	LRC_15970	FliP	*Enterococcus casseliflavus* ATCC12755, ZP_08144186.1	60%
fliQ	LRC_15960	Flagellar biosynthesis protein	*Enterococcus gallinarum EG2, ZP_05650955.1*	69%
fliR	LRC_15950	FliR	*Enterococcus casseliflavus* EC30, ZP_05646045.1	42%
flhB	LRC_15940	FlhB	*Enterococcus casseliflavus* EC20, ZP_05655659.1	50%
flhA	LRC_15930	FlhA	*Enterococcus casseliflavus* EC20, ZP_05655660.1	66%
fliA	LRC_15920	RNA polymerase sigma factor *whiG*	*Enterococcus casseliflavus* EC30, ZP_05646048.1	65%
flgF	LRC_15910	Flagellar hook-basal body complex protein	*Enterococcus casseliflavus* EC30, ZP_05646049.1	54%
flgG	LRC_15900	FlgG	*Enterococcus casseliflavus* EC30, ZP_05646050.1	41%
ftsE-like	LRC_15890	FtsE	*Enterococcus casseliflavus* EC20, ZP_05655664.1	51%
MCP	LRC_15880	MCP	*Enterococcus gallinarum* EG2, ZP_05650963.1	38%
cheW	LRC_15870	Predicted protein	*Enterococcus casseliflavus* EC30, ZP_05646053.1	46%
cheD	LRC_15860	CheD	*Enterococcus casseliflavus* EC30, ZP_05646054.1	59%
cheB	LRC_15850	Chemotaxis response regulator protein-glutamate methylesterase	*Enterococcus casseliflavus* EC30, ZP_05646055.1	53%
cheR	LRC_15840	Chemotaxis protein methyltransferase, CheR	*Enterococcus casseliflavus ATCC*12755, ZP_08144173.1	60%
cheA	LRC_15830	Chemotaxis histidine kinase	*Enterococcus gallinarum* EG2,	58%

			ZP_05650968.1	
cheC	LRC_15820	Hypothetical protein	*Enterococcus saccharolyticus* 30_1, ZP_09112688.1	58%
cheY	LRC_15810	Sporulation initiation phosphotransferase F	*Enterococcus casseliflavus* ATCC12755, ZP_08144169.1	83%
cheW	LRC_15800	Hypothetical protein	*Enterococcus saccharolyticus* 30_1, ZP_09112686.1	56%
fliM	LRC_15790	FliM	*Enterococcus casseliflavus* ATCC12755, ZP_08144167.1	63%
fliY	LRC_15780	FliY	*Enterococcus gallinarum* ATCC12755, ZP_08144166.1	55%
flgM	LRC_15770	Hypothetical protein	*Enterococcus casseliflavus* ATCC12755. ZP_08144165.1	41%
flgN	LRC_15760	Predicted protein	*Enterococcus casseliflavus* EC30, ZP_05646064.1	46%
flgK	LRC_15750	FlgK	*Enterococcus casseliflavus* EC20, ZP_05655678.1	51%
flgL	LRC_15740	Flagellin	*Enterococcus casseliflavus* EC20, ZP_05655679.1	51%
Hypothetical protein	LRC_15730	Hypothetical protein PECL_1514**	*Pediococcus claussenii* ATCC BAA-344, AEV95735.1	30%
Hypothetical protein	LRC_15720	DNA-directed RNA polymerase specialized sigma subunit, sigma24 family protein	*Lactobacillus brevis subsp. gravesensis ATCC27305*, ZP_03938825.1	28%
Hypothetical protein	LRC_15710	Enniatin synthetase**	*Fusarium equiseti*, Q00869.2	32%
*fliC*1	LRC_15700	Flagellin-domain containing protein	*Caldicellulosiruptor kristjanssonii* 177R1B. YP_004026781.1	62%
Glycosyl-transferase	LRC_15690	Glycosyltransferase family 2	*Clostridium sp.* DL-VIII, *ZP_09206896.1*	33%
*fliC*2	LRC_15680	Flagellin-domain containing protein	*Mahella australiensis* 50-1 BON, YP_004463585.1	61%
Glycosyl-transferase	LRC_15670	Hypothetical protein ANACOL_00882	*Anaerotruncus colihominis* DSM 17241, ZP_02441601.1	32%
flaG	LRC_15660	Predicted protein	*Enterococcus casseliflavus* EC30, ZP_05646068.1	34%
fliD	LRC_15650	Flagellar hook-associated protein	*Enterococcus casseliflavus* EC30, ZP_05646069.1	43%
Hypothetical protein	LRC_15640	Acetyltransferase 1 -ike**	*Ectocarpus siliculosus*, CBN77876.1	33%
fliS	LRC_15630	FliS	*Enterococcus casseliflavus* EC30, ZP_05646071.1	44%

* Homologs in motile *Lactobacillus* genomes, *L. ruminis*, *L. mali* and *L. acidipiscis*, were excluded from this table.

** BLAST hit had high E-value (>1.0).

Table S2.4: Expression analysis of flagellum biogenesis and chemotaxis genes in *L. ruminis* derived by type I microarray.

	Motile growth phase		Non-motile growth phase	
Relative fold change in:	ATCC27782	qRT-PCR	ATCC27782	qRT-PCR
MCP	1.3		1.7	
motA	**6.6**		2.2	
motB	**21.6**		4.2	
flgB	**16.2**		4.3	
flgC	**12.1**		3.5	
fliE	**12.9**		2.5	
fliF	**14.1**		2.7	
fliG	**13.2**		3.1	
fliH	**13.7**		2.0	
fliI	**12.9**		3.2	
fliJ	**12.5**		3.3	
fliK	**17.1**		4.6	
flgD	**12.6**		4.4	
Flagellar operon protein	**9.8**		4.5	
flgE	**17.7**		2.8	
flbD	**9.0**		2.2	
fliL	**17.7**		2.3	
fliZ	**22.3**		2.1	
fliP	**41.6**		2.6	
fliQ	**26.9**		1.9	
fliR	**23.3**		2.4	
flhB	**17.9**		2.2	
flhA	**16.5**		3.0	
fliA	**14.5**		2.2	
flgF	**17.1**		3.1	
flgG	**14.9**		2.8	
ftsE-like	**7.2**		1.9	
MCP	**34.6**		4.5	
cheW	**27.9**		2.8	
cheD	**32.8**		3.1	
cheB	**21.8**		3.4	
cheR	**25.0**		2.1	
cheA	**29.8**		4.0	
cheC	**24.8**		2.6	
cheY	**38.2**		3.4	
cheW	**35.6**		3.3	
fliM	**19.5**	**160010.44**	3.3	**28487.54**
fliY	**15.2**		4.7	
flgM	**16.9**		2.8	
flgN	**34.2**		2.7	
flgK	**21.3**		3.7	
flgL	**28.5**		2.2	
Hypothetical protein, (LRC_15730)	**(20.1)**	**(477.59)**	0.5	**(12.726)**
Hypothetical protein, (LRC_15720)	1.5		2.4	
Hypothetical protein, (LRC_15710)	0.4		0.6	
fliC[‡]	**90.7**	**770.49**	**49.6**	**327.9**
fliC1 LRC_15700	ND	**0**	ND	ND
fliC2 LRC_15680	ND	**1**[†]	ND	ND
Possible glycosyl transferase	**31.0**		3.5	
flaG	**25.9**		4.3	
fliD	**27.4**		3.4	
Hypothetical protein, (LRC_15640)	**44.8**		3.6	
fliS	**12.7**		3.1	
Sigma 70 like ECF sigma factor (LRC_04420/ANHS_51c)	**(800.1)**		**(12.5)**	

Values tabulated are expression ratios of ATCC27782 relative to ATCC25644. Values in parentheses represent relative fold differences in expression of ATCC25644 relative to ATCC27782. Relative data for qRT-PCR are also shown. Genes with at least 5 fold relative difference in expression and P-values < $1.0x10^{-4}$ are in bold-face. [†] Normalized to expression level of *fliC*2, ATCC27782. [‡] Microarray probes for *fliC* could not distinguish between *fliC*1 and *fliC*2, although qRT-PCR shows only *fliC*2 is expressed. ND = No data.

Table S2.5: Primers used in this study.

Primer Name	Primer Sequence
B-Actin_F	5’-ATTGGCAATGAGCGGTTC-3’
B-Actin_R	5’-TGAAGGTAGTTTCGTGGATGC-3’;
TLR5_F	5’- TCTCCACAGTCACCAAACCA-3’;
TLR5_R	5’-AAGCTGGGCAACTATAAGGTCA-3’;
ANHS_515c FliC1_SmaI_F	5’-GCGCGCCCCGGGTGCGTATTAACACAAACGTCGC-3’
ANHS_515c FliC1_XhoI_R	5’-GCGCGCCTCGAGTTAGCCTTGAAGCAAGTTCA-3’
*groEL*_F	5’-CGACGATTACGAATGACGGC-3’
*groEL*_R	5’-AGTTGCCGTTGTCGTTCCGT-3’
*fliM*_F	5’-CGATGCGCCTAAAGTCAGAG-3’
*fliM*_R	5’-CACATTCGTCCTGAGCTGAG-3’
LRC_15730_F	5’-CATGCTCGTTGAAGCTGACC-3’
LRC_15730_R	5’-CACGTAGCTCAAAGGAGCGA-3’
fliC 1_LRC_15700-F	5’-AAATCGGTGGTTTGACACAG-3’
fliC 1_LRC_15700-R:	5’-GAACCACGATCTGTTTCT-3’;
*fliC2*_LRC_15680-F:	5’-AAATCGGTGGTTTGACACAA-3’;
fliC 2_LRC_15680-R	5’-CCACGATCCTCTGCC-3’;
*era*_F:	5’-TAAACCGCAAAGCCGTCTGG-3’;
*era*_R:	5’-ATGAAATCATCGCCGCGGCC-3’
SP1_ANHS_518	5’-GATTAGGCGCTTGGTCTGTT-3’;
SP2_ ANHS_518	5'-CATTTGGTCTGAAGTTCAGG-3'
SP3_ ANHS_518	5'-CATGTCGCATCAAATGCTTG-3'
SP1_flagellin	5’-AACGTTGCATCTGTCGATCC-3’
SP2_flagellin	5'-GATGGAGCCACGATCTGTTT-3'
SP3_flagellin:	5'-CACCGATTTGTGACTTCATC-3';
SP1_AHS51c:	5’-CGTCGTTGTCTGATTGTAGC-3’;
SP2_ANHS_51c:	5 '-GCCTCTTGAAGCAAATCGTC-3';
SP3_ANHS_51c:	5'-CCTAATAACCAACGGCTTGT-3'
1054561:1055463F	5’-CGAAAGGTTCCTTCAGTTGC-3’
1054561:1055463R	5’-GCAGCCAGGTACAGCTCATA-3’
653884:654800F:	5’-TTTCATCGTTATGCCAGCAG-3’
653884:654800R:	5’-CACGTTTCACCTGATCGAAAT-3’
626252:627921F:	5’-AATCACACGTCCGACAAACA-3’
626252:627921R	5’-TTGCTTCAATCCAAGGCTCT-3’
380111:380820F	5’-GGAAGGAAAACGTGTTTCAAAT-3’
380111:380820R:	5’-CAATTCCACGCATTGTTACG-3’
1654861:1656335F:	5’-GCGTTGTGAGACATTGGTTG-3’
1654861:1656335R	5’-TCCTCTTGGCACAGCTTCTC-3’
1338430:1339559F	5’-GATTCCAGCGTGGAAATGAT-3’
1338430:1339559R	5’-TGCACTCTTCATCGGTTTTG-3’
644444:645201F:	5’-ACGAACTGTCGGAAAACGAC-3’
644444:645201R:	5’-CGACATCTTATCGCCAGGTT-3’
27F	5’-AGAGTTTGATCMTGGCTCAG-3’
1492R	5’-TACGGCACCTTGTTACGACTT-3’

Table S2.6: Uniprot and NCBI accession numbers for motility proteins.

	MotA	MotB	FlgB	FlgC	FliE	FliF	FliG	FliH
A. tumefaciens	Q44456	Q7CT75/A9CK71	Q44335	Q44336	Q44337	Q7D1A2	Q44458	
A. metalliredigens	A6TK10/A6TRP8	A6TK11/A6TRP7	A6TRR2	A6TRR1	A6TRR0	A6TRQ9/A6TKW7	A6TRQ8/A6TKW8	A6TRQ7
A. thermophilum	B9MR69	B9MR70	B9MM28	B9MM27	B9MM26	B9MM25	B9MM24	B9MM23
B. amyloliquefaciens	A7Z3Y2	A7Z3Y1	A7Z4N8	A7Z4N9	A7Z4P0	A7Z4P1	A7Z4P2	A7Z4P3
B. cereus	Q817M5/Q81FG5	Q817M6/Q81FG4	Q81FE9	Q81FE8	Q81FE7	Q81FE6	Q81FE5	Q81FE4
B. thuringiensis	Q6HD12/Q6HKS4	Q6HD13/Q6HKS3	Q6HKQ8	Q6HKQ7	Q6HKQ6	Q6HKQ5	Q6HKQ4	Q6HKQ3
B. licheniformis	Q65KI9	Q65KJ0	Q65JN1	Q65JN0	Q65JM9	Q65JM8	Q65JM7	Q65JM6
B. pumilus	A8FCH6	A8FCH5	A8FD82	A8FD83	A8FD84	A8FD85	A8FD86	A8FD87
B. clausii	Q5WGI7	Q5WGI6	Q5WFQ3	Q5WFQ4	Q5WFQ5	Q5WFQ6	Q5WFQ7	Q5WFQ8
B. subtilis	P2861	P28612/P39064	P24500	P24501	P24502	P23447	P23448	P23449
Carnobacterium sp. AT7	A8X5W8	A8X5W7	A8X5Y5	A8X5Y4	A8X5Y2	A8X5Y1	A8X5Y0	A8X5X9
C. acetobutylicum	Q97I10/Q97M94	Q97I11/Q97M93	Q97H48	Q97H49	Q97H50	Q97H51	Q97H52	Q97H53
C. botulinum	C3KZ03/C3L1E0	C3L1D9	C3L1T9	C3L1T8	C3L1T7	C3L1F0	C3L1E9	C3L1E8
C. cellulolyticum	B8I571	B8I570	B8I3Q7	B8I3Q6	B8I3Q5	B8I3Q4	B8I3Q3	B8I3Q2
C. kluyveri	A5N3S8	A5N3S7	A5N7B1	A5N7B2	A5N7B3	A5N7B4	A5N7B5	A5N7B6
C. novyi	A0Q033	A0Q034	A0Q020	A0Q021	A0Q022	A0Q023	A0Q024	A0Q025
C. phytofermentans	A9KNF0	A9KNE9	A9KNG4	A9KNG3	A9KNG2	A9KNG1	A9KNG0	A9KNF9
C. tetani	Q893Y9	Q893Z0	Q893X8	Q893X9	NT02CT1784	Q893Y0	Q893Y1	Q893Y2
D. audaxviator	B1I5I2	B1I5I3	B1I5F4	B1I5F3	B1I5F2	B1I5F1	B1I5F0	B1I5E9
D. acetoxidans	C8W1C4	C8W1C5	C8W1F4	C8W1F5	C8W1F6	C8W1F7	C8W1F8	C8W1F9
D. reducens	A4J1F0/A4J7A0	A4J1E9/A4J799	A4J768	A4J767	A4J766	A4J765	A4J764	A4J763
E. casseliflavus	ZP_05646027.1	ZP_05646026	ZP 05646028	ZP_05646029	ZP_05646030	ZP_05646031	ZP_05646032.1	ZP_05646033.1
E. gallinarum	ZP_05650938.1	ZP_05650937	ZP_05650939.1	ZP_05650940	ZP_05650941	ZP_05650942	ZP_05650943.1	ZP_05650944.1
E. coli	P09348	P0AF06	P0ABW9	P0ABX2	P0A8T5	P25798	P0ABZ1	P31068
Geobacillus sp. strain WCH70	C5D4W8	C5D4W9	C5D8W1	C5D8W2	C5D8W3	C5D8W4	C5D8W5	C5D8W6
Geobacillus sp. strain Y412MC61	C9RXX9	C9RXX8	C9S078	C9S079	C9S080	C9S081	C9S082	C9S083
H. modesticaldum	B0TCQ0	B0TCQ1	B0TH99	B0THA0	B0THA1	B0THA2	B0THA3	B0THA4
L. innocua	Q92DW8	Q92DW7	Q92DX4	Q7AP21	Q92DX2	Q92DX1	Q92DX0	Q92DT9
L. monocytogenes	Q722L2	Q722L1	Q722I7	Q722I6	Q722I5	Q722I4	Q722I3	Q722I2
L. welshimeri	A0AGE0	A0AGE1	A0AGG5	A0AGG6	A0AGG7	A0AGG8	A0AGG9	A0AGH0
Paenibacillus sp. strain JDR-2	C6CRS7	C6CRS6	C6D2X6	C6D2X5	C6D2X4	C6D2X3	C6D2X2	C6D2X1
P. aeruginosa	Q9HXL1	O87127/Q9HXL2	Q9I4Q2	Q9I4Q1	Q51462	Q51463	Q51464	A6V987
R. sphaeroides	Q3IY87/Q3J1D0	Q3J1D1/Q3IZ42/Q3J439/Q3IYA1	Q3J1S6	Q3J1S7/Q3IY73	Q3IY74/Q3J1V8	Q3IY91/Q3J1V7	Q3J1V6	Q3IY92/Q3J1V5
S. typhimurium	P55891	P55892	P16437	P0A1I7	P26462	P15928	P0A1J9	P15934
T. lettingae	A8F892	A8F893	A8F3B7	A8F3B8	A8F3B9	A8F8G8	A8F8G9	A8F8H0
T. petrophila	A5IJB1	A5IJB0	A5IMK9	A5IMK8	A5IMK7	A5IKK0	A5IKK1	A5IKK2
Thermotoga sp. strain RQ2	B1L8D0	B1L8C9	B1LBW1	B1LBW0	B1LBV9	B1L9T2	B1L9T3	
Y. enterocolitica	A1JT92	A1JT90	A1JT24	A1JT20	A1JSX2	A1JSX5	A1JSX9	A1JSV2

	FliI	FliJ	FliK	FlgD	FOP	FlgE	FlbD	FliL	FliOZ
A. tumefaciens	O34171			A9CK64		A9CK68		Q7D186	
A. metalliredigens	A6TRQ6	A6TRQ5	A6TRQ3	A6TRQ2	A6TRQ1	A6TRQ0	A6TRP9	A6TRP6/A6TKW9	
A. thermophilum	B9MM22	B9MM21	B9MM19	B9MM18	B9MM17	B9MM16	B9MM15	B9MM14	B9MM10
B. amyloliquefaciens	A7Z4P4	A7Z4P5	A7Z4P7	A7Z4P8			A7Z4Q0	A7Z4Q1	A7Z4Q5
B. cereus	Q81FE3			Q81FE2		Q81FE1			
B. thuringiensis	Q6HKQ2			Q6HKP9		Q6HKP8			
B. licheniformis	Q65JM5	Q65JM4	Q65JM2	Q65JM1			Q65JL9	Q65JL8	Q65JL4
B. pumilus	A8FD88	A8FD89	A8FD91	A8FD92			A8FD94	A8FD95	A8FD99
B. clausii	Q5WFQ9	Q5WFR0	Q5WFR2	Q5WFR3			Q5WFR5	Q5WFR6	Q3V821
B. subtilis	P23445	P20487	P23451	P23455			C0H412	P23452	P35536
Carnobacterium sp. AT7	A8X5X8	A8U5X7	A8X5X4	A8X5X3	A8X5X2	A8U5X1	A8X5W9	A8X5W6	A8X5W5
C. acetobutylicum	Q97H54	Q97H55	Q97H56	Q97H57	Q97H58	Q97H59	Q97H60	Q97H61	Q97H62
C. botulinum	C3L1E7	C3L1E6	C3L1E5	C3L1E4	C3L1E3	C3L1E2	C3L1E1	C3L1D8	C3L1D7
C. cellulolyticum	B8I3Q1	B8I3Q0	B8I3P8	B8I3P7	B8I3P6	B8I3P5	B8I3P4		B8I3N9
C. kluyveri	A5N7B7	A5N7B8	A5N7B9	A5N7C0		A5N7C1	A5N7C2	A5N7C3	A5N7C4
C. novyi	A0Q026	A0Q027	A0Q028	A0Q029	A0Q030	A0Q031	A0Q032	A0Q035	A0Q036
C. phytofermentans	A9KNF8	A9KNF7	A9KNF5	A9KNF4	A9KNF3	A9KNF2	A9KNF1		A9KNE4
C. tetani	Q893Y3	Q893Y4	Q893Y5	Q893Y6	Q893Y7	Q893Y8		Q893Z1	Q893Z2
D. audaxviator	B1I5E8	B1I5E7	B1I5E6	B1I5E5	B1I5E4	B1I5E3		B1I5E2	B1I5K3
D. acetoxidans	C8W1G0	C8W1G1	C8W1G2	C8W1G3	C8W1G4	C8W1G5		C8W1G6	C8W1G8
D. reducens	A4J762	A4J761	A4J760	A4J759	A4J758	A4J757		A4J756	A4J754
E. casseliflavus	ZP_05646034	ZP_05646035	ZP_05646036.1	ZP_05646037.1	ZP_05646038.1	C9AX79	ZP_05646040.1	ZP_05646041	ZP_05646042.1
E. gallinarum	ZP_05650945	ZP_05650946	ZP_05650947.1	ZP_05650948.1	ZP_05650949.1	ZP_05650950.1	ZP_05650951.1	ZP_05650952	ZP_05650953.1
E. coli	P52612	P52613	P52614	P75936		P75937		P0ABX8	P52627/P22586
Geobacillus sp. strain WCH70	C5D8W7	C5D8W8	C5D8X0	C5D8X1			C5D8X3	C5D8X4	C5D999
Geobacillus sp. strain Y412MC61	C9S084	C9S085	C9S087	C9S088	D3E6T7		C9S090	C9S091	C9S095
H. modesticaldum	B0THA5	B0THA6	B0THA8	B0THA9	B0THB0	B0THB1	B0THB2	B0THB3	B0THB9
L. innocua	Q92DT8			Q92DV8		Q7AP22			
L. monocytogenes	Q722I1			Q722K1		Q722K0			
L. welshimeri	A0AGH1			A0AGF1		A0AGF2			
Paenibacillus sp. strain JDR-2	C6D2X0	C6D2W9	C6D2W7	C6D2W6	C6D2W5		C6D2W3		C6D2V7
P. aeruginosa	Q9I4N1	Q9I4N0	Q02JS5(PSEAB)	Q9I4Q0		Q9I4P9		Q9HTV7/Q9I3Q5	Q51467
R. sphaeroides	Q3J1V4	Q3J1V3	Q3J1V2	Q3IY67/Q3J1S8		Q3J1S9/Q3IYA0		Q3J1V1	Q3J1X8
S. typhimurium	P26465	P0A1K1	P26416	P0A1I9		P0A1J1		P26417	P0A210/P0A1L1
T. lettingae	A8F8H1	A8F7A1	A8F888	A8F889		A8F890	A8F891	A8F894	A8F4W0
T. petrophila	A5IKK3	A5IJR1	A5IJB5	A5IJB4		A5IJB3	A5IJB2	A5IJA9	A5IJ87
Thermotoga sp. strain RQ2	B1L9T4	B1L8Y3	B1L8D4	B1L8D3		B1L8D2	B1L8D1	B1L8C8	B1L808
Y. enterocolitica	A1JSV5	A1JSV8	A1JSW1	A1JT17		A1JT15		A1JSW4	A1JSP0/A1JSX5

	FliP	FliQ	FlhB	FlhA	FliA	FlgF	FlgG	CheB
A. tumefaciens	Q44344	A9CK63	Q7D181	Q7D165		Q7D183	Q44338	O85128
A. metalliredigens	A6TRP1	A6TRP0	A6TRN8	A6TRN7/A6TKX3	A6TRM8	A6TK82	A6TK82	A6TRN3
A. thermophilum	B9MM09	B9MM08	B9MM06	B9MM05	B9MLZ6	B9MMV9	B9MMV8	B9MQG4/B9MMY5/B9MPE1/B9MMK8
B. amyloliquefaciens	A7Z4Q6	A7Z4Q7	A7Z4Q9	A7Z4R0	A7Z4R8	A7Z9L3	A7Z4P9/A7Z9L2	A7Z4R3
B. cereus	Q81FC7	Q81FC6	Q81FC4			Q81FC3		
B. thuringiensis	Q3V861	Q3V860	Q3V858	Q6HKN6		Q6HKN5		
B. licheniformis	Q65JL3	Q65JL2	Q65JL0	Q65JK9	Q65JK1	Q65E24	Q65JM0/Q65E25	Q65JK6
B. pumilus	A8FDA0	A8FDA1	A8FDA3	A8FDA4	A8FDB2	A8FI75	A8FD93/A8FI74	A8FDA7
B. clausii	Q3V822	Q3V823	Q3V825	Q5WFS0	Q5WFS6	Q5WB89	Q5WFR4/Q5WB90	Q5WFS3
B. subtilis	P35528	P35535	P35538	P35620	P10726	P39752	P23446/P39753	Q05522
Carnobacterium sp. AT7	A8X5W4	A8X5W2	A8X5W0	A8X5V9	A8X5V7	A8X5V6	A8X5V5	A8X612
C. acetobutylicum	Q97H63	Q97H64	Q97H65	Q97H66	Q97H70	Q97H73	Q97H74	Q97GZ3
C. botulinum	C3L1D6	C3L1D5	C3L1D4	C3L1D3	C3L1C9	C3L1C6	C3L1C5	C3L219
C. cellulolyticum	B8I3N8	B8I3N7	B8I3N5	B8I3N4	B8I3M5	B8I067	B8I066	B8I4Z3/B8I3N1
C. kluyveri	A5N7C5	A5N7C6	A5N7C7	A5N7C8	A5N7D1	A5N7D3	A5N7D4	A5N5P6/A5MZ47
C. novyi	A0Q037	A0Q038	A0Q039	A0Q040	A0Q044	A0Q047	A0Q048	A0PZY3
C. phytofermentans	A9KNE3	A9KNE2	A9KNE0	A9KND9	A9KNC9	A9KRZ9	A9KS00	A9KND5
C. tetani	Q893Z3	P0CF98	Q893Z4	Q893Z5	Q893Z9	Q894A1	Q894A2	Q893S6
D. audaxviator	B1I5K2	B1I5K0	B1I5J8	B1I5J7	B1I5J3	B1I5J1		B1I5H9
D. acetoxidans	C8W1G9	C8W1H0	C8W1H2	C8W1H3	C8W1H7	C8W1H9	C8W1I0	C8W0W6/C8VX95/C8W1D1/C8W1C7
D. reducens	A4J753	A4J752	A4J750	A4J749	A4J745	A4J744	A4J743	A4J796
E. casseliflavus	ZP_05646043.1	ZP_05646044	ZP_05646046.1	ZP_05646047	ZP_05646048	ZP_05646049.1	ZP_05646050.1	ZP_05646055.1
E. gallinarum	ZP_05650954.1	ZP_05650955	ZP_05650957.1	ZP_05650958	ZP_05650959	ZP_05650960.1	ZP_05650961.1	ZP_05650966.1
E. coli	P0AC05	P0AC07	P76299	P76298	P0AEM6	P75938	P0ABX5	P07330
Geobacillus sp. strain WCH70	C5D9A0	C5D9A1	C5D9A3	C5D9A4	C5D9B2	C5D968	C5D8X2/C5D967	C5D9A7
Geobacillus sp. strain Y412MC61	C9S096	C9S097	C9S099	C9S0A0	C9S0A8	C9RWK7	C9S089/C9RWK6	C9S0A3
H. modesticaldum	B0THC0	B0THC1	B0THC3	B0THC4	B0THD2	B0THM0	B0THL9	B0TG88/B0THC7
L. innocua	Q92DX7	Q92DX6	Q92DX4	Q92DX3		Q92DX1		
L. monocytogenes	Q722L8	Q3V885	Q3V883	Q722L7		Q722L5		
L. welshimeri	A0AGD1	A0AGD2	A0AGD4	A0AGD5		A0AGD7		
Paenibacillus sp. strain JDR-2	C6D2V6	C6D2V5	C6D2V3	C6D2V2	C6D2X4	C6D5C1	C6D2W4/C6D5C0	C6D2X9
P. aeruginosa	Q51468	Q9I3Q4	Q9I3Q2	Q9I3P9	P29248	Q9I4P8	Q9I4P7	Q9I6V9/Q9HXT8/O87125
R. sphaeroides	Q3J1X7/Q3IY94	Q3J1X6/Q3IY75	Q3IY82/Q3J1X4	Q3IY84/Q3J1Y1	Q3J1Y3	Q3J1T0	Q3J1T1	Q3J653/Q3J1W3
S. typhimurium	P54700	P0A1L5	P40727	P40729	P0A2E8	P16323	P0A1J3	P04042
T. lettingae	A8F4V9	A8F4V8	A8F4V6	A8F4V5	A8F4X9	A8F457	A8F458	A8F579
T. petrophila	A5IJ88	A5IJ89	A5IIM4	A5IIM5	A5IIN1	A5IM40	A5IM41	A5IK11
Thermotoga sp. strain RQ2	B1L809	B1L810	B1LC25	B1LC26	B1LC32	B1LB52	B1LB51	B1L982
Y. enterocolitica	A1JSX7	A1JSY2	A1JT48	A1JT43	A1JSP4	A1JT11	A1JT09	A1JT63

	CheR	CheA	CheY	FliM	FliNY
A. tumefaciens	Q7D1A6	A9CK76	Q7D1A5/Q7D1A8	Q44457	Q57259
A. metalliredigens	A6TRF7	A6TRN2	A6TRP3/A6TKX2	A6TRP5/A6TKX0	A6TKX1
A. thermophilum	B9MQJ4/B9MPE4/B9MMK4	B9MM01/B9MMK3/B9MPE2/B9MKK5	B9MM11	B9MM13	B9MM12
B. amyloliquefaciens	A7Z624	A7Z4R4	A7Z4Q4	A7Z4Q2	A7Z4Q3
B. cereus	Q81FF8/Q81H19	Q7BYD5	Q81FG3	Q81FD0	Q81FD1/Q81FC9
B. thuringiensis	Q6HKR7/Q6HMG6	Q6HKS1	Q6HKS2	Q6HKN9	Q6HKP0/Q6HKN8
B. licheniformis	Q65I27	Q65JK5	Q65JL5	Q65JL7	Q65JL6
B. pumilus	A8FEK6	A8FDA8	A8FD98	A8FD96	A8FD97
B. clausii	Q5WGS9	Q5WEX3	Q5WFR9	Q5WFR7	Q5WFR8
B. subtilis	P31105	P29072	P24072	P23453	P24073
Carnobacterium sp. AT7	A8U610		A8U606	A8X604	
C. acetobutylicum	Q97GZ4/Q97MS0	Q97GZ5/Q97MS3	Q97GZ7	Q97GZ9	Q97H00
C. botulinum	C3L218	C3L217	C3L215	C3L213	C3L212
C. cellulolyticum	B8I2T5	B8I3N0/B8I4Z7	B8I3P0	B8I3P2	B8I3P1
C. kluyveri	A5MZ46	A5MZ45/A5N784	A5N783	A5MZ41	A5MZ40
C. novyi	A0PZY4	A0PZY5	A0PZY7	A0PZY9	A0PZZ0
C. phytofermentans	A9KLR5	A9KND4	A9KNE5	A9KNE7	A9KNE6
C. tetani	Q893S7	Q893S8	Q893T0	Q893T2	Q893T3
D. audaxviator	B1I5I9	B1I5I0	B1I5I7	B1I5G5	B1I5G4/B1I5E1
D. acetoxidans	C8W1I2/C8W3H5/C8VX96/C8W0W5	C8W1C6/C8W0W2	C8W1I4	C8W1I5	C8W1I6
D. reducens	A4J741/A4J3Z9	A4J797	A4J739	A4J738	
E. casseliflavus	ZP_05646056.1	ZP_05646057.1	ZP_05646059	C9AXA1	ZP_05646062.1
E. gallinarum	ZP_05650967.1	ZP_05650968.1	ZP_05650970	C9A3Y8	ZP_05650973.1
E. coli	P07364	P07363	P0AE67	P06974	P15070
Geobacillus sp. strain WCH70	C5D3E2	C5D9A8	C5D998	C5D996	C5D997
Geobacillus sp. strain Y412MC61	C9S068	C9S0A4	C9S094	C9S092	C9S093
H. modesticaldum	B0TAG3	B0THC8/B0TBE2/B0TG87/B0TC94	B0THB8	B0THB4	B0THB7
L. innocua	Q92DX0	Q92DW2	P0A4H6	Q92DV5	Q92DV6/Q92DV4/Q92DW1
L. monocytogenes	Q722L4	Q722K5	Q722K6	Q722J8	Q722J9/Q722J7/Q722K4
L. welshimeri	A0AGD8	A0AGE7	A0AGE6	A0AGF4	A0AGF3/A0AGF5/A0AGE8
Paenibacillus sp. strain JDR-2	C6CUC8/C6D7E2/C6CXS4	C6D2X8	C6D2V8	C6D2W0	C6D2V9
P. aeruginosa	O87131/Q9I6V7	O87124	Q51455	Q51465	Q51466
R. sphaeroides	Q3J1W2/Q3J654/Q3J3N9	Q3J1W8/Q3J3N7/Q3J657	Q3J3P0/Q3J3N6/Q3J3N0/Q3IX24/Q3J658/Q3J1W7	Q3J1V0/Q3J1J3	Q3J1X9
S. typhimurium	P07801	P09384	P0A2D5	P26418	P26419
T. lettingae	A8F5Z5	A8F4W3	A8F4W1	A8F895	A8F896
T. petrophila	A5IJV7	A5IJ84	A5IJ86	A5IJA8	A5IJA7
Thermotoga sp. strain RQ2	B1L929	B1L805	B1L807	B1L8C7	B1L8C6
Y. enterocolitica	A1JT66	A1JT85	A1JT60	A1JSW7	A1JSX0

	FlgM	FlgN	FlgK	FlgL	FliC	FlaG	FliD	FliS
A. tumefaciens			A9CK67	A9CK66	Q7D189/Q7D179/			
A. metalliredigens	A6TL82	A6TL83	A6TL84	A6TL85	A6TLF2/A6TL89	A6TLE1/A6TP26	A6TLE2	A6TLE3
A. thermophilum	B9MKB3	B9MKB2	B9MKB1	B9MKB0	B9MKA0	B9MK90	B9MK89	B9MK88
B. amyloliquefaciens	A7Z9B2	A7Z9B1	A7Z9B0	A7Z9A9	A7Z9A5		A7Z9A4	A7Z9A3
B. cereus			Q81FF4	Q81FF3	Q81FD3/Q81FD4/Q81FD5/Q81FD6		Q81FF2	Q81FF1
B. thuringiensis			Q6HKR3	Q6HKR2	Q6HKP2/Q6HKP3		Q6HKR1	Q6HKR0
B. licheniformis	Q65EB1	Q65EB2	Q65EB3	Q65EB4	Q65EB8	Q65EB9	Q65EC0	Q65EC1
B. pumilus	A8FHX9	A8FHX8	A8FHX7	A8FHX6	A8FC67/A8FC66/A8FC68/A8FC69/A8F9D2	A8FHX2	A8FHX1	A8FHX0
B. clausii	Q5WDE5	Q5WDE6	Q5WDE7	Q5WDE8	Q5WBM8	Q5WDF1	Q5WDF2	Q5WDF3
B. subtilis	P39809	P39808	P39810	P96501	P02968	P39737	P39738	P39739
Carnobacterium sp. AT7	A8X601	A8X600	ZP_02184101	A8X5Z6	A8U5Z1	A8X5Z0	A8X5Y9	A8X5Y6
C. acetobutylicum	Q97H01	Q97H02/Q97H02	Q97H03	Q97H04	Q97IK6/Q97H04/Q97IT5/Q97H46/O69136	Q97H07	Q97H10	Q97H09
C. botulinum	C3L211	C3L210/C3L210	C3L209	C3L208	C3L1W8/C3L1U0/C3KZ58/C3L200/C3L1Z9	C3L205	C3L202	C3L203
C. cellulolyticum	B8I4C6	B8I4C7/B8I4E1	B8I4C8/B8I4C9	B8I4D0	B8I4D4/B8I4D5/B8I4D6	B8I4D8	B8I4D9	B8I4E0
C. kluyveri	A5MZ39	A5MZ38/A5MZ38	A5MZ37	A5MZ35	A5MZ27	A5MZ32	A5MZ29	A5MZ30
C. novyi	A0PZZ1	A0PZZ2	A0PZZ3	A0PZZ4	A0Q018/A0Q019/A0Q003/A0Q002	A0PZZ7	A0Q000	A0PZZ9
C. phytofermentans	A9KSP6		A9KSP9/A9KSP8	A9KSQ0	A9KL48/A9KSQ8	A9KSQ4	A9KSQ5	A9KSQ6
C. tetani	NT02CT1841	Q893T4/Q893T4	Q893T5	Q893T6	Q893W6/Q893X7/Q893U2	Q893T8	Q893X0	Q893T9
D. audaxviator	B1I5H8	B1I5H7/B1I5H7	B1I5H6	B1I5H5	B1I5G9	B1I5H0	B1I5H1	B1I5H2
D. acetoxidans	C8W1E4	C8W1E3/C8W1E3	C8W1E2	C8W1E1	C8W1D2	C8W1D8	C8W1D6	C8W1D7
D. reducens	A4J795		A4J793	A4J792	A4J786/A4J790	A4J770	A4J775	A4J772
E. casseliflavus	ZP_05646063	ZP_05646064.1	ZP_05646065		EEV29400	ZP_05646068.1	ZP_05646069	ZP_05646071.1
E. gallinarum	ZP_05650974	ZP_05650975.1	ZP_05650976		ZP_05650978	ZP_05650979.1	ZP_05650980	ZP_05650982.1
E. coli	P0AEM4	P43533	P33235	P29744	P04949		P24216	P26608
Geobacillus sp. strain WCH70	C5D7X0	C5D7T9	C5D7T8	C5D7T7	C5D7T2/C5D7U3	C5D7S	C5D7S5	C5D7S4/C5D866
Geobacillus sp. strain Y412MC61	C9RVR9	C9RVR8	C9RVR7	C9RVR6	C9RVE6/C9RVS2	C9RVC6	C9RVC5	C9RVC4/C9RVS3
H. modesticaldum	B0TH48	B0TH47	B0TH44	B0TH43	B0TH01	B0TH34	B0TH38	B0TH36/B0TH00
L. innocua			Q92DX9	Q92DX8	Q92DW3		Q92DX7	Q92DX6
L. monocytogenes			Q722J2	Q722J1	Q722K7		Q722J0	Q722I9
L. welshimeri			A0AGG0	A0AGG1	A0AGE5		A0AGG2	A0AGG3
Paenibacillus sp. strain JDR-2	C6D595	C6D594	C6D593	C6D592	C6D588	C6D4W0	C6D4V9	C6D4V8
P. aeruginosa	Q9HYP5		Q9I4P3	Q9I4P2	P72151	O33420/A3L9X7/P72152/Q9I4N7	Q9K3C5	Q9I4N6
R. sphaeroides	Q3J1X7		Q3J1T5	Q3J1T6	Q3J1U0		Q3J1T9	Q3J1X0
S. typhimurium	P26477	P0A1J7	P0A1J5	P16326	P52616/P06179		P16328	P26609
T. lettingae	A8F506		A8F504	A8F503	A8F3U3	A8F3K8	A8F3K7	A8F8R4
T. petrophila	A5IKY5		A5IKY7	A5IKY8	A5IJ26	A5IN55	A5IN56	A5IN07
Thermotoga sp. strain RQ2	B1LA66		B1LA68	B1LA69	Q4FF54	Q4FF29	B1LCW8	B1LCR6
Y. enterocolitica	A1JT31	A1JT33	A1JSZ7	A1JSZ5	A1JSQ5/A1JSQ9/A1JSQ1		A1JSR2	A1JSR5

Chapter 3

Notes:

E. rectale, R. hominis, R. inulinivorans and R. intestinalis motility gene annotations were also inspected by P. O. Sheridan.

MetaPhlAn analysis was run by S. Coghlan.

BowTie 2 was run by M. J. Claesson.

Heat plots were created by I. B. Jeffrey.

E. rectale and *R. inulinivorans* cultures for flagellin isolation were prepared by J. Martin, who also recovered flagellin from these species on one occasion.

This chapter is intended for submission as a research article to PLoSOne.

Table of Contents

Chapter 3

3.7 Supporting Information 236

Pro-inflammatory flagellin proteins of dominant motile commensal bacteria are variably abundant in the intestinal microbiome of elderly humans.

B. A. Neville[1]; P. O. Sheridan[2]; H. M. B. Harris[1]; S. Coghlan[1]; I. B. Jeffery [1]; M. J. Claesson[1]; R. P. Ross[3]; K. P. Scott[2]; P. W. O'Toole[1]*

[1] Department of Microbiology, University College Cork, Cork, Ireland.

[2] Rowett Institute of Nutrition and Health, University of Aberdeen, Bucksburn, Aberdeen AB21 9SB, United Kingdom.

[3] Teagasc Moorepark Food Research Centre, Fermoy, Co. Cork, Ireland.

*To whom correspondence should be addressed. E-mail: pwotoole@ucc.ie

Tel: +353-21-4903997. Fax: +353-21-4903101

Abstract.

Some *Eubacterium* and *Roseburia* species are among the most abundant motile bacteria present in the intestinal microbiota of healthy adults. These flagellate species contribute "cell motility" category genes to the intestinal microbiome and flagellin proteins to the intestinal proteome.

We predicted and annotated the motility genes of six *Eubacterium* and *Roseburia* species that occur in the human intestinal microbiota and examined their respective locus organization by comparative genomics. Motility gene order was generally conserved across these loci. Five of these species harboured multiple genes for predicted flagellins.

Flagellin proteins were isolated from *R. inulinivorans* strain A2-194 and from *E. rectale* strains A1-86 and M104/1. The amino-termini sequences of the *R. inulinivorans* and *E. rectale* A1-86 proteins were almost identical. These protein preparations stimulated secretion of interleukin-8 (IL-8) from human intestinal epithelial cell lines, suggesting that these flagellins were pro-inflammatory. Flagellins from the other four species were predicted to be pro-inflammatory on the basis of alignment to the consensus sequence of pro-inflammatory flagellins from the β- and γ- proteobacteria. Many *fliC* genes were deduced to be under the control of σ^{28}. The relatedness of flagellin proteins was determined by phylogenetic analysis.

Genes for flagellum biogenesis pathways from target *Eubacterium* and *Roseburia* species were variably abundant in shotgun metagenomes from 27 elderly individuals, suggesting that the current depth of coverage used for metagenomic sequencing (3.13-4.79 Gb total sequence in our study)

insufficiently captures the functional diversity of genomes present at less than ~1% relative abundance.

E. rectale and *R. inulinivorans* thus appear to synthesize complex flagella composed of flagellin proteins that stimulate IL-8 production. A greater depth of sequencing, improved evenness of sequencing and improved metagenome assembly from short reads will be required to facilitate *in silico* analyses of complete complex biochemical pathways for target species from shotgun metagenomes.

3.1 Introduction.

The mammalian colon is one of the most densely populated microbial ecosystems known (Arumugam *et al.*, 2011). The microorganisms that occupy this niche, which are collectively known as the colonic microbiota, can influence the health and well-being of the host by effecting physiological and immune functions (Mazmanian *et al.*, 2005; Rakoff-Nahoum *et al.*, 2004; Round & Mazmanian, 2010; Round *et al.*, 2011; Stappenbeck *et al.*, 2002; Swanson *et al.*, 2011). In particular, microbial metabolites, structural molecules and released cellular components are potential antigens and microbe-associated molecular patterns (MAMPs) that may stimulate the immune system (Kawai & Akira, 2011). The collection of genomes from the members of a microbial community is known as a microbiome. The genes and functions encoded by the intestinal microbiome therefore govern which bacteria and food-derived immunomodulatory molecules are likely to be present in the intestine.

Flagellin is the major structural protein in the flagella filaments of motile bacteria (Yonekura *et al.*, 2005). Flagellins and the genes encoding them, are variably abundant in the intestines (Erridge *et al.*, 2010; Kolmeder *et al.*, 2012;

Turnbaugh *et al.*, 2009) and the "cell motility" category is apparently a low-abundance microbial function in this niche (Kurokawa *et al.*, 2007; Qin *et al.*, 2011). However, motile bacteria bear significant immunostimulatory potential because humans and other animals harbour cell-surface and cytoplasmic pattern recognition receptors which respond to extra- and intra- cellular flagellin molecules respectively (Carvalho *et al.*, 2012; Gewirtz *et al.*, 2001; Hayashi *et al.*, 2001).

Particular motile *Eubacterium* and *Roseburia* species are among the most abundant bacterial species in the human intestinal microbiota (Ahmed *et al.*, 2007; Aminov *et al.*, 2006; Claesson *et al.*, 2011; Qin *et al.*, 2011; Walker *et al.*, 2011). These commensals are also notable as producers of the short chain fatty acid, butyrate, in the gut (Duncan *et al.*, 2002; Duncan *et al.*, 2007). To date, the genetic basis for flagellum biogenesis among these *Eubacterium* and *Roseburia* species has not been formally characterised, nor has the potential immune response to their flagellin proteins been established. However, it is known that heat-killed *Eubacterium rectale* cells can induce nuclear factor-κB (NF-κB) by signalling through TLR2 and TLR5 (Erridge *et al.*, 2010). Conditioned media from *Roseburia* cultures significantly stimulated and enhanced NF-κB activation in HT-29 and Caco-2 cells, while conditioned medium from *E. rectale* had an inhibitory effect on NF-κB activation (Lakhdari *et al.*, 2011). The authors of this study suggested that the immunomodulatory properties were attributed (without experimental confirmation) to flagellin and also to butyrate production (which was shown to be positively correlated with NF-κB activity in TNFα treated cell lines) by these strains (Lakhdari *et al.*, 2011). Furthermore, flagellin proteins from members of *Clostridium* cluster XIV, which may include some of the

species examined here, have been circumstantially implicated in the development of Crohn's disease and murine colitis (Duck *et al.*, 2007; Lodes *et al.*, 2004).

The genera *Roseburia* and *Eubacterium* are members of the phylum *Firmicutes* (Euzeby, 2010). While the genus *Eubacterium* is large and heterogeneous, the genus *Roseburia* is small and accordingly more homogeneous (Duncan *et al.*, 2006; Wade, 2006). The reclassification of *Eubacterium* species to other genera is quite common (Euzeby, 1997). Indeed, *E. rectale* could be more appropriately classified as a *Roseburia* species on the basis of 16S rRNA gene analyses and phenotypic properties (Duncan *et al.*, 2006), but to date its classification and nomenclature have not been revised.

In this study, we describe the genetic basis for flagellum biogenesis in six of the motile *Eubacterium* and *Roseburia* species commonly isolated from the human gastrointestinal (GI) tract. We performed genome annotation and comparative genomics, focussing on the motility loci within the genomes of these species. The pro-inflammatory potential of their flagellin proteins was predicted *in silico*, and was also experimentally tested for flagellin proteins isolated from *E. rectale* and *R. inulinivorans* strains. Finally, in order to evaluate whether genome coverage of the target species in the datasets from an intestinal metagenomics project (ELDERMET) was adequeate to identify complete biochemical pathways, we investigated the relationship between the proportional abundance of these *Eubacterium* and *Roseburia* species and our ability to detect genes for their encoded biochemical functions.

3.2 Results.

3.2.1 Improvement of genome annotation and comparative genomics of *Eubacterium* and *Roseburia* motility loci.

Initially the annotation of the genetic locus responsible for motility in each of these genomes was inspected, verified and improved as required (given that these annotations had previously been annotated by automated means only). Open reading frames (ORFs) that had not been detected by the automated annotation system were included in our improved annotation, while genes with potential frame-shifts or contig breaks were identified. The *fliJ* gene (ROSEINA2194_00946-00947) and the flagellar operon protein (FOP) (ROSEINA2194_00953-00954) genes in *R. inulinivorans*, *fliH* (ROSINTL182_07396-07395) in *R. intestinalis* and *fliF* (locus tag not assigned) in *R. hominis* apparently included frameshifts. Because these strains are motile (P. Sheridan, S. Duncan, personal communication), it is likely that these frameshifts are technical artefacts arising from sequencing or assembly errors. The primary motility locus was split over two contigs in the *R. intestinalis* genome assembly. The contig break occurred in the *flhA* gene.

The gene content and genetic organization of the largest motility loci of six *Eubacterium* and *Roseburia* species were then compared (Figure 3.1; Table S3.1). Three motility loci, *flgB-fliA*, *flgM-flgN/fliC* and *mbl-flgJ* were identified in *E. rectale*, *E. eligens* and the three *Roseburia* genomes examined. The *E. siraeum* V10Sc8a genome included two of these motility loci. In this *E. siraeum* genome, *flgF* and *flgG* were located at the *flgB-fliA* motility locus (Figure 3.1), and the genetic arrangement *mbl-flgF-flgG-flgJ* was not identified in this genome. The *flgB-fliA* locus of the *Clostridium* cluster XIVa species of interest

contained at least 34 contiguous genes and spanned 30.5 - 31.5 kb (Figure 3.1, panel A; Table S3.1). The corresponding motility locus of *E. siraeum* V10Sc8a, a member species of *Clostridium* cluster IV was smaller (~ 26.3 kb) and included fewer genes (29) overall with a slightly different arrangement.

The arrangement of genes from *flgB* to *flgE* is generally well conserved in the *Eubacterium* and *Roseburia* genomes studied (Figure 3.1, panel A). Except for the *E. rectale* and *R. hominis* genomes, a *flbD* gene was present immediately downstream of *flgE* in each genome. The *motAB* gene pair was followed by *fliLMY* in each genome. The arrangement of genes between *fliO* and *pilZ* was conserved in *E. rectale*, *E. eligens* and all of the *Roseburia* genomes examined. This locus was interrupted by a *fliA-flgF-flgG* gene translocation in *E. siraeum*. A *cheY*-like chemotaxis gene immediately preceded the *fliO-pilZ* gene cluster in each genome except *E rectale* A1-86.

A set of five contiguous chemotaxis genes organized as *cheBAWCD* were located immediately downstream of *pilZ* in *E. rectale, E. eligens* and all of the *Roseburia* genomes studied. The equivalent *E. siraeum* V10Sc8a motility locus only contained the last two of these five chemotaxis genes. The *fliA* gene was the most distal gene at this locus for all species of *Clostridium* cluster XIV examined. In the *E. siraeum* genome, *cheD* is the most distal gene of this motility cluster.

A single *flgM-flgN/fliC* motility locus occurs in four of the six genomes studied (Figure 3.1, Panel C; Table S3.1). In these genomes, up to thirteen genes are present at this locus which spans 8.0-13.6 kb. The organization of the seven genes between *flgM* and *csrA* is well conserved. The presence of two

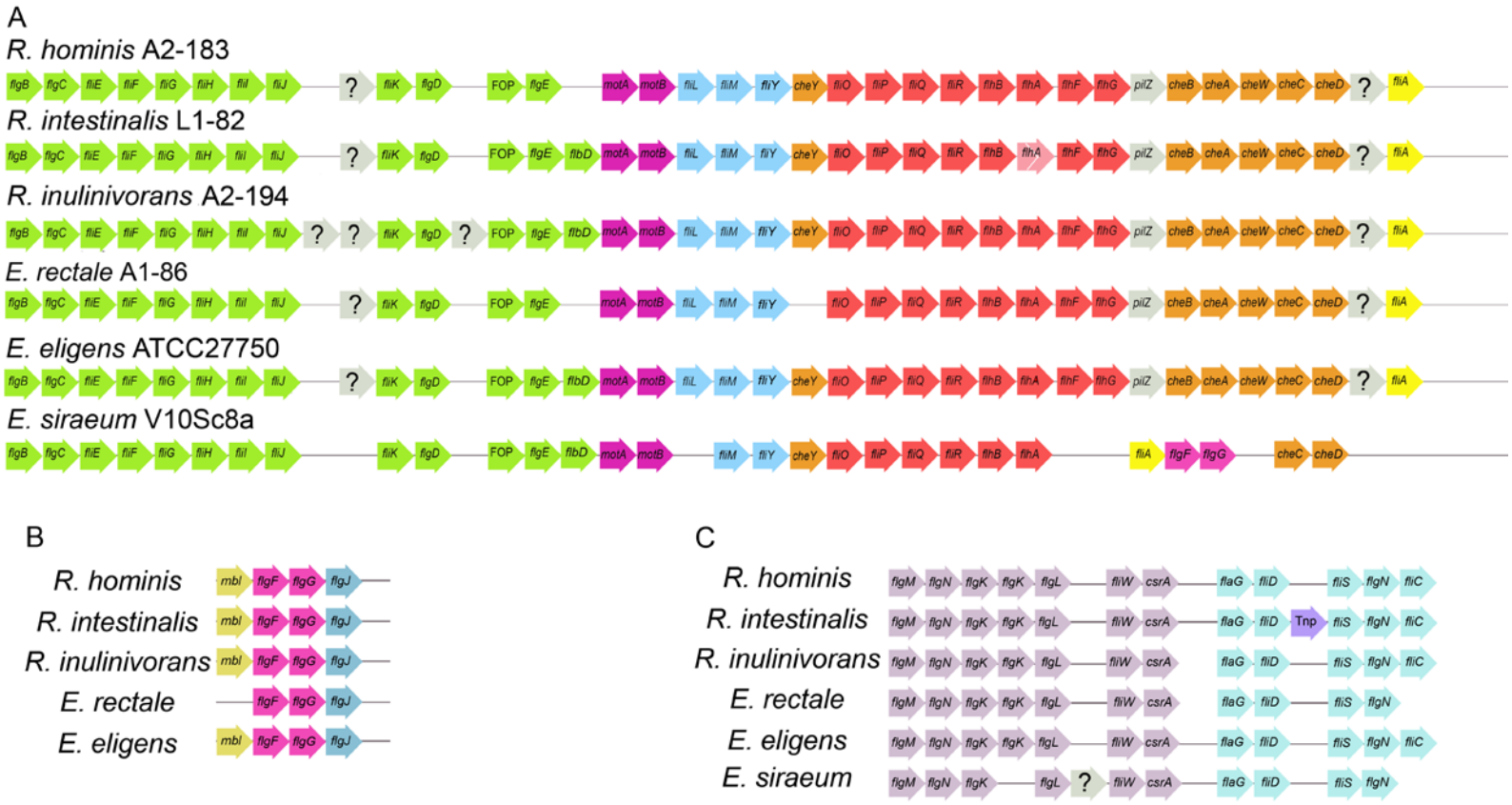


Figure 3.1: Gene order plot of three different major motility gene loci in *Eubacterium* and *Roseburia* genomes.

Genes are represented by labelled arrows. Genes that are found consecutively at a single locus (A - C) are indicated by a horizontal line. The distances between the genes at these loci were modified in this schematic diagram so that homologous genes from different genomes could be aligned. Hypothetical genes are indicated by grey arrows with ? symbols. A physical gap in the *R. intestinalis* genome assembly occurs in the *flhA* gene (Panel A, light red). A transposase gene (Tnp) is present between *fliD* and *fliS* in *R. intestinalis* (Panel C). The *flaG-flgN/fliC* gene cluster is not located immediately downstream of the *flgM-csrA* gene cluster in *R. inulinivorans* and *E. rectale* (Panel C). Colours were arbitrarily assigned to assist visual interpretation of gene rearrangements.

consecutive non-identical copies of *flgK* in five out of six genomes (single flgK in *E. siraeum*) is noteworthy. The *flaG-flgN/fliC* gene cluster is separated from the *flgM-csrA* gene cluster in *R. inulinivoran*s A2-194 and *E. rectale* A1-86. The genetic organisation of the *flaG-flgN/fliC* gene cluster is similar in all genomes, with the additional inclusion of a predicted transposase gene between *fliD* and *fliS* in *R. intestinalis* L1-82 and the absence of the flagellin gene from this locus in *E. rectale* A1-86, *E. rectale* M104/1 (FP929043.1; ERE_13960-ERE_13910) and *E. siraeum* V10Sc8a. Neither the separation of the *E. rectale* and *R. inulinivorans flgM-csrA* and *flaG-flgN/fliC* gene clusters from each other, nor the absence of flagellin genes from these genomic loci in *E. rectale* and *E. siraeum* were due to breaks in the respective draft genome assemblies.

The *mbl-flgJ* operon of four of the genomes studied included homologs of *flgF* and *flgG*, two genes which encode structural proteins of the flagellar rod and which were flanked by an MreB-like gene (*mbl*) to the 5′ end, and *flgJ*, (a muramidase), to the 3′ end. The *E. rectale* genome included a *flgF-flgG-flgJ* arrangement, but lacked an *mbl* homolog at this locus. As previously mentioned, *flgF* and *flgG* were found in the largest of the motility loci of the *E. siraeum* genome, beside the other genes encoding structural components of the basal body. The extent of sequence conservation across these motility loci was examined with Artemis Comparison Tool (ACT) plots. The motility loci of *E. rectale*, *E. eligens* and the three *Roseburia* species were similar. Although the genetic organisation of the equivalent *E. siraeum* motility loci were comparable to those of the other species studied, the primary sequence of this region was less conserved, reflected by the lower level of sequence relatedness visible in Figure S3.1.

3.2.2 Isolation, size determination and amino-terminal sequencing of the flagellin proteins of *E. rectale* and *R. inulinivorans.*

Separation of the flagellin proteins recovered from *E. rectale* A1-86 and M104/1 by SDS-PAGE revealed a single, major protein band at ~ 50 kDa (Figure 3.2). In contrast, three protein bands ranging in size from ~ 28 kDa to ~ 50 kDa were identified in the *R. inulinivorans* A2-194 flagellin preparation (Figure 3.2). The first ten residues at the amino-terminus of these candidate flagellin protein bands from *E. rectale* A1-86 and *R. inulinivorans* A2-194 (four bands in total) were sequenced and were found to be almost identical (Table S3.2). These sequences were compared to the translated *fliC* sequences from each genome.

Five *fliC* genes were annotated in the *E. rectale* A1-86 genome and the predicted molecular masses of these flagellin proteins were similar, ranging from ~ 47 to ~ 53 kDa (Table 3.1). Five proteins of such similar molecular weights would not have been separated under the SDS-PAGE conditions used here. The first ten residues of these four predicted flagellin proteins are identical, and matched the chemically determined amino-terminal sequence of the ~ 50 kDa protein band exactly. The flagellin protein encoded by the coding DNA sequence (CDS) EUR_28730 is similar in size (~ 50.78 kDa) to the other predicted flagellin proteins, but only four of its first ten residues match the chemically determined sequence, suggesting that it is not expressed to high levels in this strain under these growth conditions.

Four *fliC* genes were annotated in the genome of *E. rectale* M104/1. The first ten residues predicted for one of these flagellin proteins, (ERE_14720) were

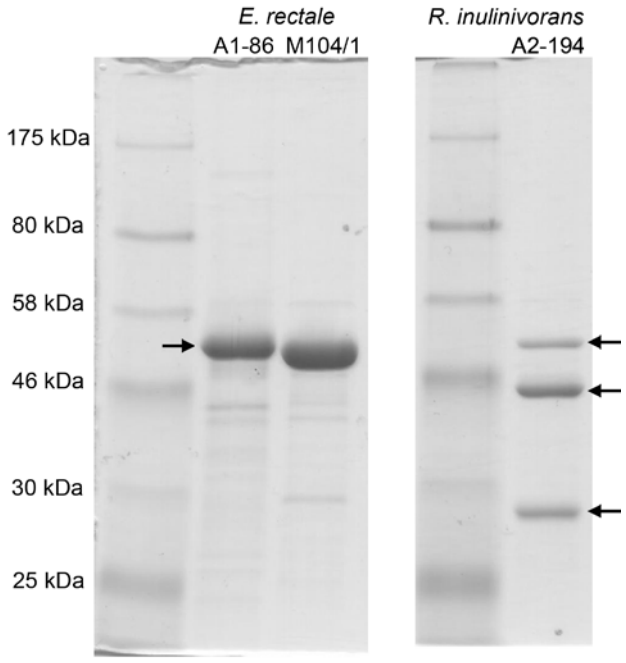


Figure 3.2: Flagellin proteins from *E. rectale* and *R. inulinivorans* separated on Coomassie stained SDS-PAGE gels. Arrows indicate the proteins for which amino-terminal sequence data is available. The broad-range, pre-stained protein marker used (P7708S) was purchased from New England Biolabs.

Table 3.1: Summary of the properties of *Eubacterium* and *Roseburia* flagellin proteins and their predicted promoter and ribosome binding site sequences.

Species	No. Flagellins	Locus Tag	Phylogenetic Clade (Fig. S3.2)	Accession	Size (aa)	Size (kDa)	Sequence of first ten residues	Predicted -35 sequence*	Predicted -10 sequence*	-35 -10 spacing (bp)	Predicted sigma factor*	-10 to start-codon spacing (bp)	Predicted RBS	RBS-Start codon spacing (bp)
R. hominis L1-83	3	RHOM_15820	(a)	AEN98270.1	506	54.48	MRINYNVSAS	taaa	gcgatat	9	28	261	AGGAGA	8
		RHOM_00820	(d)	AEN95291.1	275	30.62	MVVNHNMAAI	taaa	tcgatat	17	28	47	AAGAGG	9
		RHOM_00665	(e)	AEN95260.1	270	28.60	MVVQHNLTAM	taaa	ccgatat	16	28	136	AGGAGG	8
R. intestinalis L1	4 (5)	ROSINTL182_05247†	(a)	ZP_04742102.2	486	51.90	MRINYNVSAA	taga	ccgatat	15	28	78	AGAAGG	9
		ROSINTL182_08635†	(c)	ZP_04745261.1	539	56.13	MVVQHNMSAM	taaa	-	-	-	-	CGGAGG	14
		ROSINTL182_05608	(d)	ZP_04742436.1	275	30.55	MVVNHNMALI	taaa tttaca	tcgatat cataaa	17 9	28 43	47 24	AAGAGG	9
		ROSINTL182_07256	(e)	ZP_04743973.1	272	29.04	MVVQHNMTAM	taaa	ccgatat	16	28	149	AGGAGG	9
		ROSINTL182_09568	-	ZP_04746122.1	61	6.97	MTLIQNRLEY	taaa	-	-	-		-	-
R. inulinivorans A2-194	6	ROSEINA2194_00754	(a)	ZP_03752351.1	493	52.52	MRINNNMSAV	taag	acgatat	17	28	34	AGAAGG	10
		ROSEINA2194_01954	(d)	ZP_03753535.1	426	47.24	MQVLAHNLAA	taat	ccgataa	27	28	193	AGGAGA	6
		ROSEINA2194_00384†	(e)	ZP_03751985.1	270	28.77	MVVQHNMTAA	taaa attaca	ccgatat aataat	16 12	28 43	146 0	AGGAGG	8
		ROSEINA2194_00549†	(f)	ZP_03752147.1	389	42.06	MVVQHNMQAM	tttaca	aataat	18	43	142	CGGAGG	8
		ROSEINA2194_01473	(f)	ZP_03753062.1	392	42.26	MVVQHNLQAM	-	-	-	-	-	AGGAGG	8
		ROSEINA2194_02155	(f)	ZP_03753734.1	466	49.23	MVVQHNMQAM	tgaa	gcgataa	23	28	375	AGGAGG	8
E. eligens ATCC27750	3	EUBELI_00422	(c)	YP_002929886	497	52.32	MVVQHNMAAM	taaa	-	-	-	-	CGGAGG	8
		EUBELI_00241†	(e)	YP_002929724.1	270	28.93	MVVQHNLSAM	taaa	ccgatat	16	28	93	AGGAGG	8
		EUBELI_00264	(e)	YP_002929747.1	270	29.12	MVVQHNLSAM	ttaa taaaca	ccgataa aataat	16 13	28 43	92 52	AGGAGG	8
E. rectale A1-86	5	EUR_28730	(a)	CBK91820.1	476	50.78	MKINRNMSAV	taaa	tcgatat	17	28	69	AGGAAA	9
		EUR_04790	(f)	CBK89689.1	504	53.41	MVVQHNMQAA	tttcca	cataat	9	43	32	AGGAGG	8
		EUR_14430	(f)	CBK90534.1	480	50.22	MVVQHNMQAA	-	-	-	-	-	TGGAGG	8
		EUR_04300	(f)	CBK89645.1	476	50.10	MVVQHNMQAA	tttcca	cataat	9	43	33	AGGAGG	8
		EUR_14450	(f)	CBK90536.1	455	47.51	MVVQHNMQAA	tttacc	aataat	12	43	22	TGGAGG	8
E. rectale M104/1	4	ERE_01930	(a)	CBK92329.1	476	50.77	MKINRNMSAV	taaa	tcgatat	17	28	69	AGGAAA	9
		ERE_14590	(f)	CBK93435.1	458	48.29	MVVQHNMQAM	-	-	-	-	-	AGGAGG	8
		ERE_14720	(f)	CBK93446.1	504	53.41	MVVQHNMQAA	-	-	-	-	-	AGGAGG	8
		ERE_12290†	-	CBK93233.1	446	46.82	YRINRAADDA	-	-	-	-	-	-	-
E. siraeum V10Sc8a	1	ES1_07000†	(b)	CBL33805.1	530	55.81	MVVQHNLNAI	tttaca taaa	tataaa ccgatat	10 17	43 28	258 192	AGGAGG	17‡
E. siraeum DSM15702	1	EUBSIR_02119†	(b)	ZP_02423261.1	539	56.24	MVVQHNLNAI	tttaca taaa	caaaat ccgatat	11 17	43 28	258 192	AGGAGG	17‡
E. siraeum 70/3	1	EUS_23890†	(b)	CBK97362.1	547	56.97	MVVQHNLNAI	tttaca taaa	tataaa ccgatat	9 17	43 28	258 192	AGGAGG	17‡

* Sequences were compared to the -35 and -10 recognition sequences for *Butyrivibrio fibrisolvens* σ^{28} and σ^{43}, which are -35:TAAA (N16-17) -10:MCGATAa and -35:TTtACA (N19) -10:cATAAT respectively. The general bacterial consensus sequences for σ^{28} and σ^{43} are -35:TAAA (N15) -10:CCGATAT and -35:TTGACA (N15) -10:TATAAT respectively. † Predicted start positions were moved on the basis of alignment to amino-terminal sequences of *E. rectale* and *R. inulinivorans* flagellins. ‡ An alternative start codon exists three residues upstream of the predicted start position. Use of this alternative start codon would yield a distance of 8 bp between the predicted RBS and the start-codon.

.

identical to the amino termini of four of the predicted flagellin proteins of *E. rectale* A1-86. The predicted molecular weights of three of these proteins were similar to those predicted for the flagellins of *E. rectale* A1-86. The CDS ERE_12290 is proximally truncated by a break in the draft genome assembly, and was thus selectively excluded from further analyses.

Six *fliC* genes were annotated in the *R. inulinivorans* A2-194 genome. The predicted molecular masses of these candidate flagellin proteins ranged from ~ 29 kDa to ~ 53 kDa (Table 3.1). It appears that the translated product of CDS ROSEINA2194_00384 corresponds to the protein product at ~ 29 kDa in the SDS-PAGE gel. The products of CDSs ROSEINA2194_00549 and ROSEINA2194_01473 have predicted molecular masses of ~ 42 kDa. These may correspond to the protein product migrating at ~ 43 kDa on the SDS-PAGE gel. Indeed, the sequence of the flagellin product of CDS ROSEINA2194_00549 corresponds to this protein band, while the product of CDS ROSEINA2194_01473 differs only at residue 7.

Flagellin products of CDSs ROSEINA2194_01954, ROSEINA2194_02155 and ROSEINA2194_00754 have predicted molecular masses of ~ 47, ~49 and ~50 kDa respectively, and they may be present in the protein band of ~ 50 kDa on the SDS-PAGE gel.

3.2.3 *In silico* flagellin promoter analysis.

The nucleotide sequences upstream of the *fliC* genes in each genome of interest were inspected to identify potential promoter sequences and to infer which sigma factors might direct transcription from each promoter (Table 3.1). Promoters under the direction of either σ^{28} or σ^{43} were identified by comparison to the consensus sequences identified for these promoters in *Butyrivibrio*

fibrisolvens, and to the bacterial consensus sequences for promoters controlled by these sigma factors.

The outcomes of this promoter analysis are reported with reference to the clades in the phylogenetic tree based on flagellin proteins, shown in Figure S3.2. CDSs corresponding to the flagellins in clades A, D and E were under the presumptive control of σ^{28}, with the exception of CDSs ROSINTL182_05608 and EUBELI_00264 which were apparently also controlled by σ^{43}. Both σ^{28} and σ^{43} consensus sequences were identified for the CDSs encoding the *E. siraeum* flagellin proteins (clade B), but the σ^{28} sequences were closer than the σ^{43} sequences to the predicted start codons of these CDSs. Potential promoters could not be identified for every CDS with a corresponding protein in clade F. The CDSs for which promoters could be identified were mostly under the control of σ^{43}.

The inferred σ^{28} and σ^{43} promoters varied considerably in their distance from the predicted CDS start codons, (σ^{28} : range, 47-375 bp; mean = 139 bp. σ^{43}: range, 0-258 bp; mean = 108 bp). The unconventional spacing between the predicted -35 and -10 recognition sequences, and the lack of absolute conservation in the predicted recognition sequences, suggests that if the predicted σ^{28} promoters of ROSEINA2194_01954 and ROSEINA2194_02155 are functional, transcription from these promoters could be suboptimal. This could explain the low abundance of one or both of these flagellins in *R. inulinivorans* cultures. Promoter analysis in *E. rectale* M104/1 was hindered because the regions upstream of the target CDSs were often disrupted by gaps in the draft genome assembly. No potential σ^{28} or σ^{43} promoter sequences were identified

upstream of *fliC* CDS EUBELI_00422, ROSINTL182_09568 or ROSINTL182_08635.

3.2.4 *In silico* and *in vitro* analysis of the pro-inflammatory potential of flagellin proteins from *Eubacterium* and *Roseburia* species.

To predict if the *Eubacterium* and *Roseburia* flagellin proteins were likely to be pro-inflammatory, these proteins were aligned to a consensus sequence (11 residues long) derived from a region of the pro-inflammatory flagellins of the β- and γ- proteobacteria (Andersen-Nissen *et al.*, 2005; Smith *et al.*, 2003). Residues L87, R89, L93 and Q96 of the *Eubacterium* and *Roseburia* flagellin proteins inspected here were absolutely conserved with respect to the consensus sequence (Figure 3.3). These residues are critical for TLR5 signalling and flagellin polymerisation (Andersen-Nissen *et al.*, 2005; Smith *et al.*, 2003). Another residue, Q88, that is critical for signalling and polymerisation, is also completely conserved in each of the *Eubacterium* and *Roseburia* sequences with respect to the β- and γ- proteobacteria flagellin consensus sequence, except for the translated products of CDSs ROSINTL182_05608 and RHOM_00820, in which a Q88D substitution is evident. On the basis of their overall similarity to the consensus sequence, these proteins were predicted to have pro-inflammatory properties.

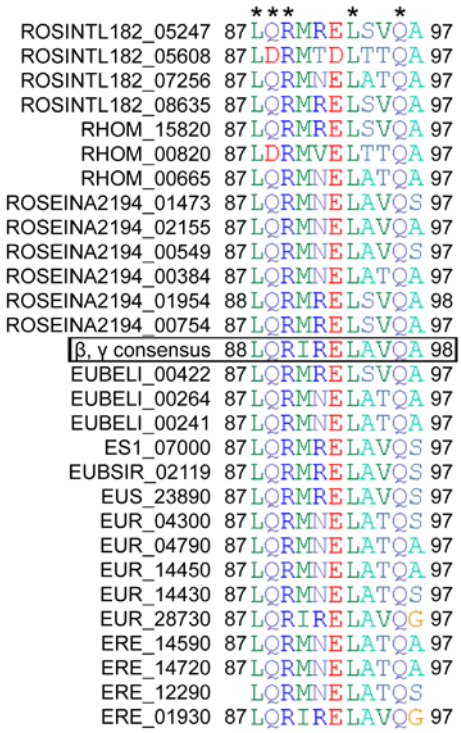


Figure 3.3: Mutliple alignment of the consensus region of the flagellin proteins of β and γ proteobacteria that is recognised via TLR5 with the corresponding regions of predicted flagellin proteins from the *Roseburia* and *Eubacterium* species studied.

Residues that are critical for TLR5 recognition are indicated with an asterisk. Alignment was performed with ClustalW in BioEdit. Flagellin proteins from the various species are labelled with a locus tag. A gap in the draft genome assembly meant that positional information could not be included for the sequence fragment of CDS ERE_12290 in this alignment.

Two human intestinal epithelial cell lines (IECs), T84 and HT-29, were exposed to the flagellin proteins isolated from *R. inulinivorans* A2-194 and *E. rectale* strains A1-86 and M104/1. Increased interleukin-8 (IL-8) secretion by the IECs in response to these flagellin preparations was taken as evidence of a pro-inflammatory response. Significantly more IL-8 was secreted from T84 and from HT-29 cells treated with each of the *Eubacterium* and *Roseburia* flagellin preparations than from the untreated control cells (one-tailed Mann-Whitney U test, $P \leq 0.01$, $n = 5$; $n = 6$ respectively) (Figure 3.4).

3.2.5 Identification of selected *Eubacterium* and *Roseburia* species in 27 individual metagenomes.

MetaPhlAn (Segata *et al.*, 2012) was used to determine the relative abundance of 5 of the 6 species of interest in a metagenome database derived from the faecal microbiotas of 27 elderly individuals (Claesson *et al.*, 2012). The relative abundance of *R. hominis* was not considered using this method because its genome was not included as part of the Integrated Microbial Genomes system, upon which the MetaPhlAn clade-specific marker database was based (Segata *et al.*, 2012). Metagenomes EM039 and EM173 were excluded from the MetaPhlAn analysis as they were prepared using alternative sequencing and assembly strategies, which meant that the MetaPhlAn results generated from these two metagenomes were not directly comparable to those from the other 25 metagenomes (Claesson *et al.*, 2012).

According to MetaPhlAn's read-based classification, 23 of the 25 metagenomes harboured at least one of the five species of interest at a relative abundance ≥ 0.5 % (Table S3.3). Twenty of the 25 metagenomes harboured at

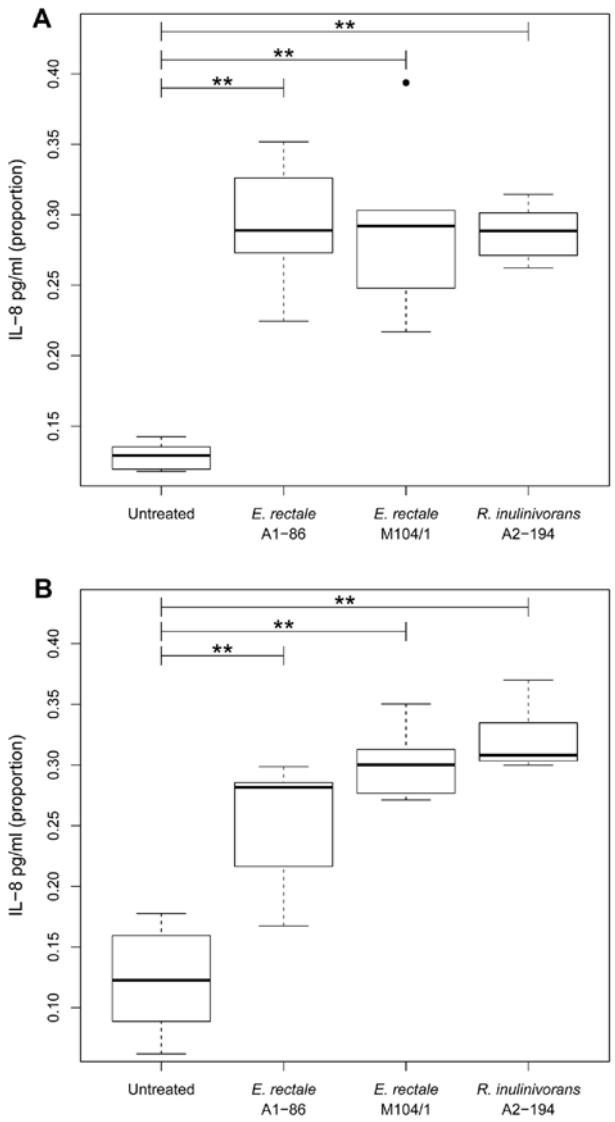


Figure 3.4: IL-8 secretion from T84 cells (A) and HT-29 cells (B) in response to flagellin preparations from *E. rectale* and *R. inulinivorans.*

Concentrations of IL-8 as determined by ELISA were converted to proportions (as described in materials and methods) for statistical analysis. Boxplots show the median value and interquartile range. Outliers are indicated by a black dot. Horizontal bars with the ** symbol indicate that significantly more IL-8 was secreted from the cells treated with flagellin preparations than from the untreated control cells, P-value < 0.01, one-tailed Mann-Whitney U test, n = 5 for T84 cells, n = 6 for HT-29 cells.

least one species of interest at a relative abundance of ≥ 1 %. The relative abundances of each species varied considerably across the metagenomes, and the range of *E. siraeum* relative abundance in particular, was quite large (0.01 % (EM191) - 31.59 % (EM305)). Five of the individuals harboured this species at a relative abundance > 3 %. Eight people harboured *E. siraeum* at a predicted relative abundance of ⩽ 0.1 %.

Recruitment plots of the whole genome sequences of the species of interest aligned to the metagenomes indicated that these genomes were incompletely represented in each of the individual metagenomes (data not shown). The sequencing coverage for each genome of interest was calculated as a function of metagenome sequencing depth, target genome size and the predicted relative abundance. The species of interest were often represented at less than 10 fold coverage in these metagenomes (Table S3.4). This level of genome coverage would probably be insufficient to represent the genomes of interest completely (Fleischmann *et al.*, 1995; Lander & Waterman, 1988).

3.2.6 Identification of *Eubacterium* and *Roseburia* motility genes in the faecal metagenomes of 27 elderly individuals.

The detection of motility CDSs from raw reads was a function of target CDS length and species relative abundance (Figure S3.3). The normalised number of raw reads that were mapped to each target CDS increased with both CDS length and the relative abundance of the species of interest in each metagenome. Thus, longer CDSs could be detected at lower species relative abundances than short CDSs (Figure S3.3).

In general, at a species relative abundance of ~0.1 % or greater, ~10 ($Log_{10}1$) reads (normalised value) were mapped to most of the target genes from

each species (Figure S3.3), and the target DNA sequence was considered as "present" in the database. At species relative abundance values greater than or equal to ~0.4 %, more than ~32 reads ($Log_{10}1.5$) (normalised value) mapped to each target CDS, strongly suggesting that the target DNA sequences were present in the database. In general, homology based methods could identify target genes only when the larger of these species abundance thresholds was exceeded (Table S3.5). However, motility CDSs were not always detected when a species occurred at a relative abundance ≥ 0.4 %. For example, the species *R. inulinivorans* was estimated at 1.41 % relative abundance in EM251 and the corresponding heat-plot suggests that many of the unassembled reads from this metagenome mapped to the target motility CDSs (Figure S3.3). However, no genes of the *flgB-fliA* motility locus were detected in the assembled metagenome database for this individual by either the homology and annotation or recruitment plot methods (Table S3.5, Data not shown). Similarly, metagenome EM326 appeared to harbour a complete set of motility genes for *E. eligens*, a species which occurred at 1.54 % relative abundance in this metagenome (Figure S3.3). However, a recruitment plot indicated that few genes at the *flgB-fliA* motility locus of this species were present in the assembled EM326 metagenome (Data not shown).

The heat-plots also show that the genomes of interest were sometimes incompletely represented in the raw read databases. For example, zero or very few reads mapped to the *E. rectale flgB-fliA* motility locus in metagenomes EM148, EM175, EM205 and EM232, even though *E. rectale* was determined to occur at high relative abundances (> 0.9 %) in these metagenomes. Similarly,

target *E. eligens* motility genes were non-uniformly detected in the metagenomes examined, even when this species occurred at high (>1 %) relative abundance.

Homology searches and gene context information were used to determine if motility genes of the *flgB-fliA* and *flaG-flgN/fliC* motility loci from the species of interest could be identified from assembled metagenomes. At least some of these *Eubacterium* and *Roseburia* motility genes of interest from the *flgB-fliA* or *flaG-flgN/fliC* motility loci were identified in 23 of the 27 assembled metagenomes (Table S3.5). *E. siraeum* motility CDSs were identified in 11 of these 23 metagenomes. Motility CDSs from two or more of the target species were detected in 11 of these 23 metagenomes. No single metagenome appeared to harbour complete motility gene sets for all the bacterial species (Table S3.5).

There was overall correspondence in the detection of *E. siraeum*, *R. intestinalis* and *R. inulinivorans* motility genes from raw and assembled reads (Figure S3.3, Table S3.5), though target motility CDSs could be detected at lower species relative abundances in the database of raw reads than in the database of assembled metagenomes according to the search criteria used. Our inability to detect the motility genes of species that are apparently present in the metagenome database could be a consequence of the incomplete representation of the genome of interest in the metagenome database arising from a non-uniform distribution of sequencing coverage across a target genome, or DNA degradation prior to metagenome library sequencing.

To evaluate the overall abundance of cell motility genes in these assembled metagenomes, the number of “cell motility” COGs (category N) associated with each metagenome was investigated (Table S3.6). The number of “cell motility” COGs represented by each assembled metagenome varied

considerably, ranging from 2 COGs (EM227) to 19 COGs (EM283). Accordingly, the proportion of COGs assigned to this functional category varied across the metagenomes, and ranged from 0.13 % (EM227) to 0.87 % (EM205, EM326) of total COGs assigned to any category per metagenome. Thus, the function of "cell motility" was not abundantly encoded in any of these assembled metagenomes.

3.2.7 Identification of *Eubacterium, Roseburia* flagellin genes and proteins in the faecal metagenomes of 27 elderly individuals.

The presence of flagellin proteins in each of the 27 metagenomes was evaluated with fragment recruitment plots (Figure S3.4) and also by BLAST searches. The recruitment plots revealed that the flagellin proteins of the species of interest were present in 8 of the 27 metagenomes. Two of the four full-length *R. intestinalis* flagellins (ROSINTL182_05608 and ROSINTL_07256) were only represented in metagenome EM268. Of the six *R. inulinivorans* flagellins, only the product of *fliC* CDS ROSEINA2194_00754 was identified, and was represented in two metagenomes, EM268 and EM175. Partial matches to *R. inulinivorans* flagellin proteins encoded by ROSEINA2194_00754 and ROSEINA2194_01954 were identified in metagenome EM173.

The protein product of *E. rectale* CDS ERE_01930 was the only *E. rectale* flagellin represented in the database, being identified in 5 metagenomes (EM039, EM205, EM251, EM268, EM219). The protein encoded by CDS ERE_01930 is 99 % similar to EUR_28730, and would explain why a non-identical, but highly similar homolog of EUR_28730 occurs in every metagenome that also encodes an identical match to ERE_01930. The *E. siraeum* 70/3 flagellin protein encoded by CDS EUS_23890 was present only in

metagenome EM039. Homologs of this flagellin which are 74 % and 88 % identical to EUS_23890 respectively from other *E. siraeum* strains were not identified in any of the metagenomes examined. However, a protein similar to the *E. siraeum* flagellin encoded by CDS ES1_07000 was identified in metagenome EM176. *E. eligens* flagellin proteins were not identified in any of the metagenomes by this method. Recruitment plots could not be constructed for metagenomes EM208, EM227, EM238 and EM275 because no informative alignment data were returned by the analysis, indicating that these flagellin proteins were not represented in the recruitment plots above the thresholds used (which is consistent with results presented above). When a flagellin protein of interest was detected at 100 % similarity by the recruitment plot method, the other flagellin proteins of this species were not also detected. Filtered tBLASTn searches ($\geq$ 90 % minimum identity, E-value $\leq 1.0 \times 10^{-8}$, $\geq$ 250 residues long) suggested that *Eubacterium* and *Roseburia* flagellins were represented in 8 metagenomes (EM039, EM175, EM204, EM205, EM209, EM219, EM251 and EM268). EM268 harboured sequences which aligned to five flagellins (ROSINTL182_07256, ROSINTL182_05608, ROSINTL182_05247, ROSEINA2194_00754 and one sequence that aligned to both ERE_01930 and EUR_28730). The equivalent *E. rectale* flagellin homologues from two different strains (ERE_01930 and EUR_28730) aligned to sequences in 5 metagenomes, (EM039, EM205, EM219, EM251, EM268). The *E. siraeum* flagellin EUS_23890 aligned only to EM039. Flagellin proteins ROSINTL182_05247, ROSEINA2194_00384 and ROSEINA2194_00754 aligned to metagenomes EM209, EM204 and EM175 respectively.

Sequences that could be assigned to COG1344, which represents "flagellin and related hook-associated proteins", were present in 23 of the 27 assembled metagenomes (Table S3.6). Because this analysis was performed on assembled metagenomes, it only indicates the presence or absence of the target COGs in the metagenome databases, and does not provide the overall abundance of particular COGs. Metagenomes EM148, EM204, EM227 and EM308 did not harbour any sequences that could be assigned to this COG category. This automated functional analysis therefore suggests that "flagellin and related hook-associated proteins" are variably represented in these metagenome databases.

3.3 Discussion.

Because of their production of flagella, the motile *Eubacterium* and *Roseburia* species have considerable immunostimulatory potential. While motility may be a colonisation factor for enteric *Roseburia* species (Scott *et al.*, 2011; Stanton & Savage, 1894), the expression of flagellin proteins that are recognised by human TLR5 nevertheless confers a pro-inflammatory capacity upon these species (Duck *et al.*, 2007). By *in silico* analysis, the flagellin proteins of the *Eubacterium* and *Roseburia* species studied here were all predicted to be pro-inflammatory, and this pro-inflammatory capacity was experimentally supported for the flagellin proteins isolated from strains of *E. rectale* and *R. inulinivorans*. These findings are consistent with those of previous studies, which demonstrated that whole cells and conditioned media from species of this phylogenetic cluster could activate NF-κB or expression from an NF-κB reporter construct (Erridge *et al.*, 2010; Lakhdari *et al.*, 2011). Although NF-κB is often activated in response to pathogenic infections, its activation is not necessarily undesirable and the pro-inflammatory flagellin

proteins characterised here could contribute favourably to gut health by promoting intestinal epithelial homeostasis and by preventing cell-death and disease (Rakoff-Nahoum *et al.*, 2004; Vijay-Kumar *et al.*, 2006; Wullaert *et al.*, 2011).

The arrangement of genes in operons and/or transcriptional units which reflect the order of their temporal expression is a common feature of bacterial flagellar systems which contributes to the efficient regulation of flagellum biogenesis (Kalir *et al.*, 2001; Zaslaver *et al.*, 2006). The genetic arrangement of motility genes in the *Eubacterium* and *Roseburia* genomes revealed here is consistent with that found in other motile species of the phylum *Firmicutes* (Forde, 2013). Gene order is known to become less conserved with increasing genetic distance between species (Tamames, 2001). Consistent with this, the genetic organisation of the major motility loci were very similar among the *Clostridium* cluster XIVa genomes investigated, but the *E. siraeum* motility locus was more unlike the others at a sequence level and with respect to gene content. This reflects its phylogenetic positioning in *Clostridium* cluster IV rather than in *Clostridium* cluster XIVa.

The *Eubacterium* and *Roseburia* motility genes were found at various loci throughout each genome, as is the case with several *Clostridium* and *Bacillus* species. The genes in the largest of the *Eubacterium* and *Roseburia* motility loci encode the structural and regulatory components of the basal-body and hook. These are expected to be transcribed early in the flagellum biogenesis pathway to anchor the flagellum in the cell membrane. The organisation of the genes for the structural, chaperone and regulatory functions involved in flagellar filament formation at another motility locus (*flgM-flgN/fliC*) may enable the efficient

regulation and timely expression of these genes. In support of this hypothesis, a similar gene arrangement occurs in a number of other bacterial lineages (Mukherjee *et al.*, 2011).

In four of the genomes studied, two genes encoding structural rod proteins, *flgF* and *flgG*, which transmit torque from the motor to the hook and filament, were found in a separate four gene operon. The *mbl* and *flgJ* genes were located immediately up- and down- stream of the *flgF- flgG* gene pair, respectively. The *mbl* gene encodes an MreB-like protein which has a role in determining cell morphology and polarity (Abhayawardhane & Stewart, 1995). The FlgJ protein is a rod-specific muramidase with peptidoglycan hydrolysing ability that is exploited during the construction of transmural flagellar structures (Nambu *et al.*, 1999). In some *Firmicutes* species (Forde, 2013), including *E. siraeum* V10Sc8a, *flgF* and *flgG* are found in an operon with the genes for other basal body and rod proteins. However, the *mbl-flgF-flgG-flgJ* genetic arrangement described here is also found in the genomes of several closely related *Butyrivibrio* and *Clostridium* species from *Clostridium* clusters XIV and III, and in *Alkaliphilus oremlandii* (family *Clostridiaceae*) and *Abiotrophia defectiva* (class *Bacilli*). The *E. rectale* FlgF and FlgG proteins are 54 % (154/282 aa) and 52 % (141/282 aa) similar to *Bacillus subtilis* subsp. *subtilis* FlhO (CAB05950.1) and FlhP (CAB05941.1) respectively, suggesting that these genes are homologous. The *mbl-flhO-flhP* gene arrangement occurs in *Bacillus*, *Geobacillus* and *Oceanobacillus* species. The functional and evolutionary significance of the *mbl - flgJ* genetic arrangement is presently unknown.

Flagellin expression is known to occur at higher levels in *R. inulinivorans* A2-194 when it is grown on starch rather than on glucose, inulin or fructan

substrates (Scott *et al.*, 2011). This nutritional control of motility gene expression implies that pleiotropic global regulators may direct motility gene transcription or translation in *Roseburia* species. Under nutrient rich conditions, CodY represses flagellin expression in *B. subtilis* (Bergara *et al.*, 2003). A *codY* homolog was identified immediately upstream of the *flgB-fliA* motility locus in the *E. rectale, E. eligens, R. hominis* and *R. intestinalis* genomes examined. (In *R. inulinivorans*, the CDS encoding the predicted *codY* homolog (ROSEINA2194_0938) is apparently fused to the 3′ end of a CDS encoding a protein with DNA topoisomerase I function). CsrA, a global regulator that inhibits flagellin gene expression in *B. subtilis* (Yakhnin *et al.*, 2007), but which is necessary for motility and flagellum biosynthesis in *E. coli* (Wei *et al.*, 2001), was also found at the *flgM-flgN/fliC* motility locus of all genomes examined. In other species, the activities of CodY and CsrA can be modulated by changes in intracellular guanosine tetraphosphate (ppGpp), guanosine nucleoside triphosphate (GTP) or branched chain amino-acid pools (Bergara *et al.*, 2003; Dalebroux & Swanson, 2012). Unfavourable environmental conditions such as nutrient limitation, induce a stringent response in some bacteria which leads to either motility gene expression or repression by altering intracellular concentrations of these effector molecules (Dalebroux & Swanson, 2012). Further experiments would be required to determine which, if any of these effector molecules, modulate motility gene transcription via CodY or CsrA in motile *Eubacterium* and *Roseburia* species during growth on different carbohydrate substrates.

In silico analysis of promoter consensus sequences suggested that the *fliC* genes in the *Eubacterium* and *Roseburia* genomes of interest were mostly under

the control of σ^{28}, although some σ^{43} dependent promoters were also identified. In *B. fibrisolvens,* transcription of one *fliC* gene is driven from two different promoters, yielding two transcripts with alternative transcription start-sites (Kalmokoff *et al.*, 2000). For the *Eubacterium* and *Roseburia fliC* genes with potentially more than one promoter, it is not yet clear if transcription proceeds from both. The presence of two promoters for a single *fliC* gene, one of which is under the presumptive control of a housekeeping sigma factor, suggests that there may be a requirement for constitutive *fliC* transcription at a basal level in these species. It also suggests that post-transcriptional or post-translational control mechanisms, such as those that have been described for other motile species (Yakhnin *et al.*, 2007), might additionally regulate flagellin expression in these species.

The motile *Eubacterium* and *Roseburia* species bear subterminal flagella (Duncan *et al.*, 2002; Stanton *et al.*, 2009) and the annotation of several flagellin proteins in the genomes of these *Eubacterium* and *Roseburia* species suggests that these bacteria might produce complex flagella in which the filament is composed of several different flagellin proteins. This inference is supported by the recovery of at least three flagellin proteins from *R. inulinivorans* cultures. It is possible that *E. rectale* also produces complex flagella, but the sizes and amino-terminal sequences of its flagellins were insufficiently unique to determine which of its flagellins were expressed. In contrast, only one flagellin gene was annotated in each of the genomes of three *E. siraeum* strains, so this species presumably produces flagella composed of a single flagellin protein. Gene gain by duplication or horizontal gene transfer could explain the

occurrence of multiple genes encoding flagellin in the genomes of these species of interest.

We attempted to identify motility CDSs of specific motile, enteric *Eubacterium* and *Roseburia* species from the raw read and assembled metagenome datasets generated by the ELDERMET project (Claesson *et al.*, 2012). Our heat-plots showed that the identification of motility CDSs from databases of unassembled reads was a function of both target gene length, gene context and target species relative abundance. Longer CDSs would, therefore, be detected at lower species relative abundances than shorter CDSs (Figure S3.3-A). At species relative abundances of ~ 0.1 %, unassembled reads mapped non-uniformly to the target motility loci (Figure S3.3), implying an uneven depth of sequencing coverage of the target genome at this level of species relative abundance. Therefore a greater depth of sequencing could potentially improve gene detection in these low abundance species by increasing the overall sequence coverage of these loci.

In metagenomes that were thought to include *E. rectale* at high (> 1 %) species relative abundances, the apparent absence of the *E. rectale flgB-fliA* motility locus was unexpected. Technical issues, such as DNA degradation or a DNA sequence composition which was refractory to sequencing might explain the lower than expected coverage of this region in databases of raw reads. Alternatively, the divergence or loss of this region in enteric *E. rectale* strains would also preclude the detection of these target motility genes by comparison to the reference genome of a laboratory strain.

We suspect that incomplete sequence coverage of the target bacterial genomes also imposed a limitation on our ability to identify specific genes or

pathways from the assembled metagenomes. The assembly status of the query genome and the metagenome database may also influence the outcome, because more fractured assemblies yield shorter alignments. Thus, even at the large sequencing depths (3317 to 4798 Mb) and metagenome contig lengths (2050 bp ≤ N50 ≤ 64999 bp) used here (Claesson *et al.*, 2012), these metagenomes appear to incompletely capture the total functional diversity encoded in these faecal microbial communities.

Consistent with earlier studies (Kurokawa *et al.*, 2007), our recruitment plot and COG analyses suggest that genes encoding cell motility functions occur at variable and low abundances in the human gut microbiome. Poor coverage of low abundance genomes is a known current limitation of metagenomics (Warnecke & Hugenholtz, 2007) and gene finding from assembled, but fragmented sequences is a recognised challenge for pathway reconstruction from metagenomes (De Filippo *et al.*, 2012). Our attempt to identify genes involved in bacterial motility from specific high-abundance target species from databases of raw reads and assembled metagenomes, highlights the need for a greater depth and evenness of sequencing or improved metagenome assembly from short reads to improve gene detection and pathway reconstruction.

In summary, we have confirmed the pro-inflammatory nature of the flagellins of some of the most abundant motile commensal bacteria in the human GI tract *in vitro* and we have investigated the potential regulation of these genes by *in silico* means. We also highlight the need for greater depth and evenness of sequencing in the preparation of metagenome databases to ensure that the genetic functionality encoded by an ecosystem is fully captured at species level.

3.4 Materials and Methods.

3.4.1 Culture conditions.

The three strains (*E. rectale* A1-86, *E. rectale* M104/1 and *R. inulinivorans* A2-194) were previously isolated from human faecal samples (Barcenilla *et al.*, 2000; Louis *et al.*, 2004). The growth medium used was anaerobic M2GSC (Miyazaki *et al.*, 1997). This medium was divided into 7.5 ml aliquots in Hungate tubes, sealed with butyl rubber septa (Bellco Glass) or 500 ml aliquots in 1 litre laboratory bottles (Duran Group), with specially modified airtight caps. All cultures were inoculated using the anaerobic methods described by (Bryant, 1972) and were incubated anaerobically at 37 °C without agitation. In brief, carbon dioxide gas was diffused through the growth medium before it was sealed in an airtight vessel. Carbon dioxide was pumped into the overnight cultures and into the fresh medium to maintain the anaerobic conditions during inoculation.

In order to obtain sufficient quantities of flagellin protein, large batches of bacterial culture were grown anaerobically: Two overnight 7.5 ml cultures of M2GSC broths were used to inoculate each single anaerobic bottle containing 500 ml of M2GSC. Duplicate bottles were prepared for each strain. These subcultures were incubated for 16-18 hr before harvesting the flagellin proteins using methods outlined previously (Neville *et al.*, 2012).

3.4.2 SDS-PAGE, staining, quantification and amino-terminal sequencing of flagellin proteins.

Flagellin proteins were electrophoresed on 10 % SDS-PAGE gels and were visualized by staining with Coomassie blue stain followed by destaining with “destain solution” (methanol: acetic acid: water, 454: 92 : 454).

Proteins separated by electrophoresis were transferred to Immobilon membrane for amino-terminal sequencing. Transfer of proteins was performed at 40 mA for 50 mins in transfer buffer (1x CAPS, 100 ml methanol, 800 ml water). The membrane was stained and destained post-transfer to visualize the proteins. The protein bands of interest were excised from the membrane and the first ten residues of each protein band were amino-terminally sequenced by AltaBioscience, Birmingham, UK.

Proteins were quantified using the Pierce BCA protein assay (ThermoScientific) according to the microplate procedure outlined by the manufacturer.

3.4.3 Stimulation of intestinal epithelial cells and IL-8 ELISA.

HT-29 and T84 cells were routinely cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % foetal bovine serum and 1 % penicillin/streptomycin antibiotics (stock concentrations: 10,000 U penicillin and 10 mg streptomycin/ml) and were incubated at 37 °C under conditions of 5 % CO_2. IECs were seeded at a density of 2 x 10^4 cells/well of a sterile 96 well plate. After seeding, IECs were allowed to adhere overnight before flagellin treatment.

Flagellin proteins were added to each well to a final concentration of 0.1 µg/well. Exposure of the IECs to flagellin proteins took place for 12 hours. Supernatants were subsequently recovered. The IL-8 concentration in these supernatants was measured with the IL-8 ELISA Duo kit (R&D systems) according to the manufacturer's instructions. Experimental replicates were performed on different days. For statistical analysis, the raw IL-8 values were converted to proportions by dividing the IL-8 concentration for each treatment in

a single experiment by the sum of the IL-8 concentrations for all of the treatments from the same experiment. A one-tailed Mann-Whitney U test was performed on the transformed values.

3.4.4 Genome annotation and improvement, comparative genomics, metagenome assembly.

Draft and complete genome sequences were downloaded from the nucleotide database on the National Center for Biotechnology Information website (Table S3.7). Several of these genomes had previously been annotated by automated procedures. These auto-annotations of motility genes at the major motility loci in the *E. rectale* A1-86 and *R. inulinivorans* A2-194 genomes were inspected. The motility gene arrangements in the other genomes of interest were examined with respect to the major motility loci of the *E. rectale* and *R. inulinivorans* genomes. Additional open reading frames that were not previously identified in the auto-annotation of these draft genomes were inferred on the basis of their genetic neighbourhood and BLASTp similarity to characterized homologues. The CDSs that represented fragments of genes that apparently included frameshift mutations were merged. Start positions of genes encoding flagellin proteins were adjusted to correspond to the amino-terminal sequence derived for the flagellin proteins that were recovered from *E. rectale* and *R. inulinivorans*.

Assembled metagenomes representing the intestinal microbiomes of 27 elderly Irish individuals from one of three community settings (community, rehabilitation and long-stay) were generated previously (Claesson *et al.*, 2012) and each included on average 4.6 Gb of sequence information. Twenty-five of these metagenomes were constructed from libraries of 91 bp paired-end Illumina

reads with an insert size of 350 bp. Two of these metagenomes (EM039 and EM173) were assembled using two different types of sequencing technologies, specifically paired-end Illumina reads that were 101 bp in length with a 500 bp insert size in combination with 551,726 and 665,164 454 Titanium sequencing reads for EM039 and EM173, respectively.

3.4.5 Analyses of presence or absence, relative abundance and extent of genome coverage of *Eubacterium* and *Roseburia* species of interest in metagenomes.

MetaPhlAn 1.6.0 (Segata *et al.*, 2012) was used to infer the relative abundances of the target species in the 27 metagenomes. The "MetaPhlAn script" and the "BowTie2 database of the MetaPhlAn markers" were downloaded from http://huttenhower.sph.harvard.edu/metaphlan. Unfiltered paired-end reads were combined in a FASTQ file which was converted to FASTA format using FASTQ-to-FASTA (FASTX-Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/commandline.html). The output file was subjected to MetaPhlAn analysis using default parameters.

The estimated coverage of each target genome in each of the 27 metagenomes was calculated as a function of the metagenome size, the average size of the target species' genome and the MetaPhlAn-predicted relative abundance of the species of interest according to the following formula: ((Metagenome size (Mb) x Rel. Abundance (%))/(Target genome size (Mb)) (Warnecke & Hugenholtz, 2007).

3.4.6 Identification and annotation of motility proteins of *Eubacterium* and *Roseburia* species in metagenomes.

Two approaches, based on either raw sequencing reads or reads assembled into contigs, were adopted for the identification of motility genes from the target species of interest in the ELDERMET metagenomes. Bowtie 2 (Langmead & Salzberg, 2012) was used with default settings (end-to-end read alignment, --sensitive -D 15 –R 2 –N 0 –L 22 –i S,1,1.25) to map raw sequencing reads from each metagenome to the *Eubacterium* and *Roseburia* ORFs and CDSs of interest. The number of mapped reads was normalised according to the following calculation: (No. mapped reads) x (Mean sequencing depth/Sequencing depth per metagenome). The mean sequencing depth was taken as 4.79×10^9 bases per metagenome. The total sequencing depth for each metagenome was reported as part of the supporting information accompanying an earlier publication (Claesson *et al.*, 2012).

Heat-plots were created with the "Heatplot" function as part of the Made4 package (Culhane *et al.*, 2005) for R. These plots were based on the normalised number of mapped reads per gene per metagenome, the MetaPhlAn (Segata *et al.*, 2012) derived species relative abundance values and target CDS lengths (bp). For metagenomes EM039 and EM173, species relative abundance values were inferred by ranking all the metagenomes in order of increasing total number of normalised mapped sequencing reads. Target CDSs were considered as present at a minimum threshold of ~10 normalised reads mapped per gene ($Log_{10}1$).

A selection of 177 *Eubacterium* and *Roseburia* motility proteins (excluding flagellins) which represented the *flgB-fliA* and *flgM-flgN/fliC* motility loci of eight different species (*E. cellulosolvens, E. eligens, E. rectale, E.*

siraeum, E. yurii subsp. *margaretiae, R. hominis, R. intestinalis, R. inulinivorans*) were used as tBLASTn queries to search the database of assembled metagenomes for contigs which likely harboured motility genes from the species of interest. The metagenome contigs that yielded alignments which were ≥ 90 % identical to the query proteins were retrieved from the database. These contigs were viewed and all potential ORFs were called using Artemis (Rutherford *et al.*, 2000). These ORFs were annotated on the basis of BLASTp homology to proteins in the non-redundant protein database (NR) available from NCBI, and also by a general inspection of their genetic neighbourhood. The motility genes of a target species were considered to be present in a target metagenome if the best BLASTp hits for at least half of the motility CDSs on each contig occurred with identity ≥ 90 % to homologs from only one of the target species.

3.4.7 COG category analysis.

The 27 assembled metagenomes (Claesson *et al.*, 2012) are publically available on the MG-RAST website (Meyer *et al.*, 2008). COG classifications were determined via MG-RAST for each metagenome using default parameters (≥ 60 % identity, ≥ 15 aa alignment length, E-value ≤ 1 x 10^{-5}). Data were viewed in tabular output format and were filtered at "level 2" to limit results to "cell motility" COGs. The proportion of COGs assigned to this category was expressed as a percentage of total COGs (total number of COGs returned before filtering).

3.4.8 Generation of recruitment plots.

Recruitment plots were constructed using PROmer 3.07 (Delcher *et al.*, 2002) to align the query sequences to the database of assembled metagenomes.

Query sequences were typically complete or draft genome sequences, genomic fragments representing a motility locus of interest or a multi-fasta file representing genes of interest. The PROmer delta output file was filtered using mummerplot 3.5 (part of the MUMmer package) (Kurtz *et al.*, 2004). The plots were generated with a range of 80 – 100 % similarity represented on the Y axis.

3.4.9 Comparative genomics.

Nucleotide and amino-acid alignments were performed with MUSCLE (Edgar, 2004) or ClustalW in BioEdit. Artemis Comparison Tool was used to view the conservation and arrangement of large genome segments across species (Carver *et al.*, 2008). The comparison files were generated in tabular format using tBLASTx (Altschul *et al.*, 1990). A minimum identity threshold of 30 % was imposed on the alignments for visualisation purposes.

3.4.10 Phylogenetic analysis.

Phylogenies constructed from protein sequences were first aligned using MUSCLE (Edgar, 2004). A rooted flagellin protein phylogenetic tree was constructed using PHyML 3.0 (Guindon *et al.*, 2010) with the LG substitution matrix. Modelgenerator (Keane *et al.*, 2006) was used to choose the most appropriate substitution model. Alignment columns that included gaps were removed before constructing the maximum likelihood tree.

3.4.11 Promoter sequence analysis.

The nucleotide sequences upstream of the genes encoding flagellin proteins were inspected to identify potential sigma factor consensus sequences and ribosome binding sites (RBS). The promoter sequences of the housekeeping sigma factor (σ^{43}) (-35: TTtACA; -10: cATAAT) and the flagellar sigma factor (σ^{28}) (-35: TAAA; -10: MCGATAa) of *Butyrivibrio fibrisolvens* (another motile

species of *Clostridium* cluster XIVa) were used as reference sequences (Kalmokoff *et al.*, 2000). Ribosome binding sites were expected to occur within 20 bp of the predicted start-codon (Chen *et al.*, 1994), and to conform to the sequence AGGAGG.

3.5 Acknowledgements.

The authors would like to thank B. M. Forde, J. Martin and S. Rampelli for advice and technical assistance. This work was supported by a Principal Investigator Award from Science Foundation Ireland to PWOT. BAN received an Embark studentship from the Irish Research Council for Science Engineering and Technology.

3.6 References.

Abhayawardhane, Y. & Stewart, G. C. (1995). *Bacillus subtilis* possesses a second determinant with extensive sequence similarity to the *Escherichia coli mreB* morphogene. *J Bacteriol* **177**, 765-773.

Ahmed, S., Macfarlane, G. T., Fite, A., McBain, A. J., Gilbert, P. & Macfarlane, S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* **73**, 7435-7442.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Aminov, R. I., Walker, A. W., Duncan, S. H., Harmsen, H. J., Welling, G. W. & Flint, H. J. (2006). Molecular diversity, cultivation, and improved detection by fluorescent *in situ* hybridization of a dominant group of human gut bacteria related to *Roseburia* spp. or *Eubacterium rectale*. *Appl Environ Microbiol* **72**, 6371-6376.

Andersen-Nissen, E., Smith, K. D., Strobe, K. L., Barrett, S. L., Cookson, B. T., Logan, S. M. & Aderem, A. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc Natl Acad Sci U S A* **102**, 9247-9252.

Arumugam, M., Raes, J., Pelletier, E. & other authors (2011). Enterotypes of the human gut microbiome. *Nature* **473**, 174-180.

Barcenilla, A., Pryde, S. E., Martin, J. C., Duncan, S. H., Stewart, C. S., Henderson, C. & Flint, H. J. (2000). Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* **66**, 1654-1661.

Bergara, F., Ibarra, C., Iwamasa, J., Patarroyo, J. C., Aguilera, R. & Marquez-Magana, L. M. (2003). CodY is a nutritional repressor of flagellar gene expression in *Bacillus subtilis*. *J Bacteriol* **185**, 3118-3126.

Bryant, M. P. (1972). Commentary on the Hungate technique for culture of anaerobic bacteria. *Am J Clin Nutr* **25**, 1324-1328.

Carvalho, F. A., Nalbantoglu, I., Aitken, J. D., Uchiyama, R., Su, Y., Doho, G. H., Vijay-Kumar, M. & Gewirtz, A. T. (2012). Cytosolic flagellin receptor NLRC4 protects mice against mucosal and systemic challenges. *Mucosal Immunol* **5**, 288-298.

Carver, T., Berriman, M., Tivey, A., Patel, C., Bohme, U., Barrell, B. G., Parkhill, J. & Rajandream, M. A. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672-2676.

Chen, H., Bjerknes, M., Kumar, R. & Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the

translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res* **22**, 4953-4957.

Claesson, M. J., Cusack, S., O'Sullivan, O. & other authors (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4586-4591.

Claesson, M. J., Jeffery, I. B., Conde, S. & other authors (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184.

Culhane, A. C., Thioulouse, J., Perriere, G. & Higgins, D. G. (2005). MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**, 2789-2790.

Dalebroux, Z. D. & Swanson, M. S. (2012). ppGpp: magic beyond RNA polymerase. *Nat Rev Microbiol* **10**, 203-212.

De Filippo, C., Ramazzotti, M., Fontana, P. & Cavalieri, D. (2012). Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. *Brief Bioinform* **13**, 696-710.

Delcher, A., Phillippy, A., Carlton, J. & Salzberg, S. (2002). Fast algorithms for large-scale genome alignment and comparision. *Nucleic Acids Res* **30**, 2478-2483.

Duck, L. W., Walter, M. R., Novak, J., Kelly, D., Tomasi, M., Cong, Y. & Elson, C. O. (2007). Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis* **13**, 1191-1201.

Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. (2002). *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int J Syst Evol Microbiol* **52**, 1615-1620.

Duncan, S. H., Aminov, R. I., Scott, K. P., Louis, P., Stanton, T. B. & Flint, H. J. (2006). Proposal of *Roseburia faecis* sp. nov., *Roseburia hominis* sp. nov. and *Roseburia inulinivorans* sp. nov., based on isolates from human faeces. *Int J Syst Evol Microbiol* **56**, 2437-2441.

Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A. M., Flint, H. J. & Lobley, G. E. (2007). Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl Environ Microbiol* **73**, 1073-1078.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Erridge, C., Duncan, S. H., Bereswill, S. & Heimesaat, M. M. (2010). The induction of colitis and ileitis in mice is associated with marked increases in intestinal concentrations of stimulants of TLRs 2, 4, and 5. *PloS one* **5**, e9125.

Euzeby, J. (2010). *Lachnospiraceae.* http://www.bacterio.cict.fr/bacdico/ll/lachnospiraceae.html

Euzeby, J. P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* **47**, 590-592.

Fleischmann, R. D., Adams, M. D., White, O. & other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Forde, B. M. (2013). Genomics of commensal lactobacilli. In *Microbiology*, pp. 290. Cork: University College Cork.

Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J. & Madara, J. L. (2001). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* **167**, 1882-1885.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Hayashi, F., Smith, K. D., Ozinsky, A. & other authors (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099-1103.

Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M. G. & Alon, U. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**, 2080-2083.

Kalmokoff, M. L., Allard, S., Austin, J. W., Whitford, M. F., Hefford, M. A. & Teather, R. M. (2000). Biochemical and genetic characterization of the flagellar filaments from the rumen anaerobe *Butyrivibrio fibrisolvens* OR77. *Anaerobe* **6**, 93-109.

Kawai, T. & Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34**, 637-650.

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McLnerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* **6**, 29.

Kolmeder, C. A., de Been, M., Nikkila, J. & other authors (2012). Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PloS one* **7**, e29913.

Kurokawa, K., Itoh, T., Kuwahara, T. & other authors (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12.

Lakhdari, O., Tap, J., Beguet-Crespel, F. & other authors (2011). Identification of NF-kappaB modulation capabilities within human intestinal commensal bacteria. *J Biomed Biotechnol* **2011**, 282356.

Lander, E. S. & Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231-239.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359.

Lodes, M. J., Cong, Y., Elson, C. O., Mohamath, R., Landers, C. J., Targan, S. R., Fort, M. & Hershberg, R. M. (2004). Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest* **113**, 1296-1306.

Louis, P., Duncan, S. H., McCrae, S. I., Millar, J., Jackson, M. S. & Flint, H. J. (2004). Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* **186**, 2099-2106.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107-118.

Meyer, F., Paarmann, D., D'Souza, M. & other authors (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*.

Miyazaki, K., Martin, J. C., Marinsek-Logar, R. & Flint, H. J. (1997). Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp. *brevis*) B(1)4. *Anaerobe* **3**, 373-381.

Mukherjee, S., Yakhnin, H., Kysela, D., Sokoloski, J., Babitzke, P. & Kearns, D. B. (2011). CsrA-FliW interaction governs flagellin homeostasis and a checkpoint on flagellar morphogenesis in *Bacillus subtilis*. *Mol Microbiol* **82**, 447-461.

Nambu, T., Minamino, T., Macnab, R. M. & Kutsukake, K. (1999). Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in *Salmonella typhimurium*. *J Bacteriol* **181**, 1555-1561.

Neville, B. A., Forde, B. M., Claesson, M. J., Darby, T., Coghlan, A., Nally, K., Ross, R. P. & O'Toole, P. W. (2012). Characterization of pro-inflammatory

flagellin proteins produced by *Lactobacillus ruminis* and related motile lactobacilli. *PLoS One* **7**, e40592.

Qin, J., Li, R., Raes, J. & other authors (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229-241.

Round, J. L. & Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A* **107**, 12204-12209.

Round, J. L., Lee, S. M., Li, J., Tran, G., Jabri, B., Chatila, T. A. & Mazmanian, S. K. (2011). The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332**, 974-977.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Scott, K. P., Martin, J. C., Chassard, C. & other authors (2011). Substrate-driven gene expression in *Roseburia inulinivorans*: importance of inducible enzymes in the utilization of inulin and starch. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4672-4679.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O. & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**, 811-814.

Smith, K. D., Andersen-Nissen, E., Hayashi, F., Strobe, K., Bergman, M. A., Barrett, S. L., Cookson, B. T. & Aderem, A. (2003). Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat Immunol* **4**, 1247-1253.

Stanton, T. B. & Savage, D. C. (1984). Motility as a factor in bowel colonization by *Roseburia cecicola*, an obligately anaerobic bacterium from the mouse caecum. *J Gen Microbiol* **130**, 173-183.

Stanton, T. B., Duncan, S. H. & Flint, H. J. (2009). Genus XVI. *Roseburia* Stanton and Savage 1983a, 626. In *Bergey's manual of systematic bacteriology*, pp. 954-956. Edited by P. Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K. H. Schleifer & W. B. Whitman. New York: Springer

Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A* **99**, 15451-15455.

Swanson, P. A., 2nd, Kumar, A., Samarin, S., Vijay-Kumar, M., Kundu, K., Murthy, N., Hansen, J., Nusrat, A. & Neish, A. S. (2011). Enteric commensal bacteria potentiate epithelial restitution via reactive oxygen species-mediated inactivation of focal adhesion kinase phosphatases. *Proc Natl Acad Sci U S A* **108**, 8803-8808.

Tamames, J. (2001). Evolution of gene order conservation in prokaryotes. *Genome Biol* **2**, 00020.00021-00020.00011.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T. & other authors (2009). A core gut microbiome in obese and lean twins. *Nature* **457**, 480-484.

Vijay-Kumar, M., Wu, H., Jones, R., Grant, G., Babbin, B., King, T. P., Kelly, D., Gewirtz, A. T. & Neish, A. S. (2006). Flagellin suppresses epithelial apoptosis and limits disease during enteric infection. *Am J Pathol* **169**, 1686-1700.

Wade, W. G. (2006). The genus *Eubacterium* and related genera. In *The Prokaryotes*, pp. 823-835. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer & E. Stackebrandt. New York: Springer.

Walker, A. W., Ince, J., Duncan, S. H. & other authors (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* **5**, 220-230.

Warnecke, F. & Hugenholtz, P. (2007). Building on basic metagenomics with complementary technologies. *Genome Biol* **8**.

Wei, B. L., Brun-Zinkernagel, A. M., Simecka, J. W., Pruss, B. M., Babitzke, P. & Romeo, T. (2001). Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol Microbiol* **40**, 245-256.

Wullaert, A., Bonnet, M. C. & Pasparakis, M. (2011). NF-κB in the regulation of epithelial homeostasis and inflammation. *Cell Res* **21**, 146-158.

Yakhnin, H., Pandit, P., Petty, T. J., Baker, C. S., Romeo, T. & Babitzke, P. (2007). CsrA of *Bacillus subtilis* regulates translation initiation of the gene encoding the flagellin protein (hag) by blocking ribosome binding. *Mol Microbiol* **64**, 1605-1620.

Yonekura, K., Maki-Yonekura, S. & Namba, K. (2005). Building the atomic model for the bacterial flagellar filament by electron cryomicroscopy and image analysis. *Structure* **13**, 407-412.

Zaslaver, A., Mayo, A., Ronen, M. & Alon, U. (2006). Optimal gene partition into operons correlates with gene functional order. *Phys Biol* **3**, 183-189.

3.7 Supporting Information.

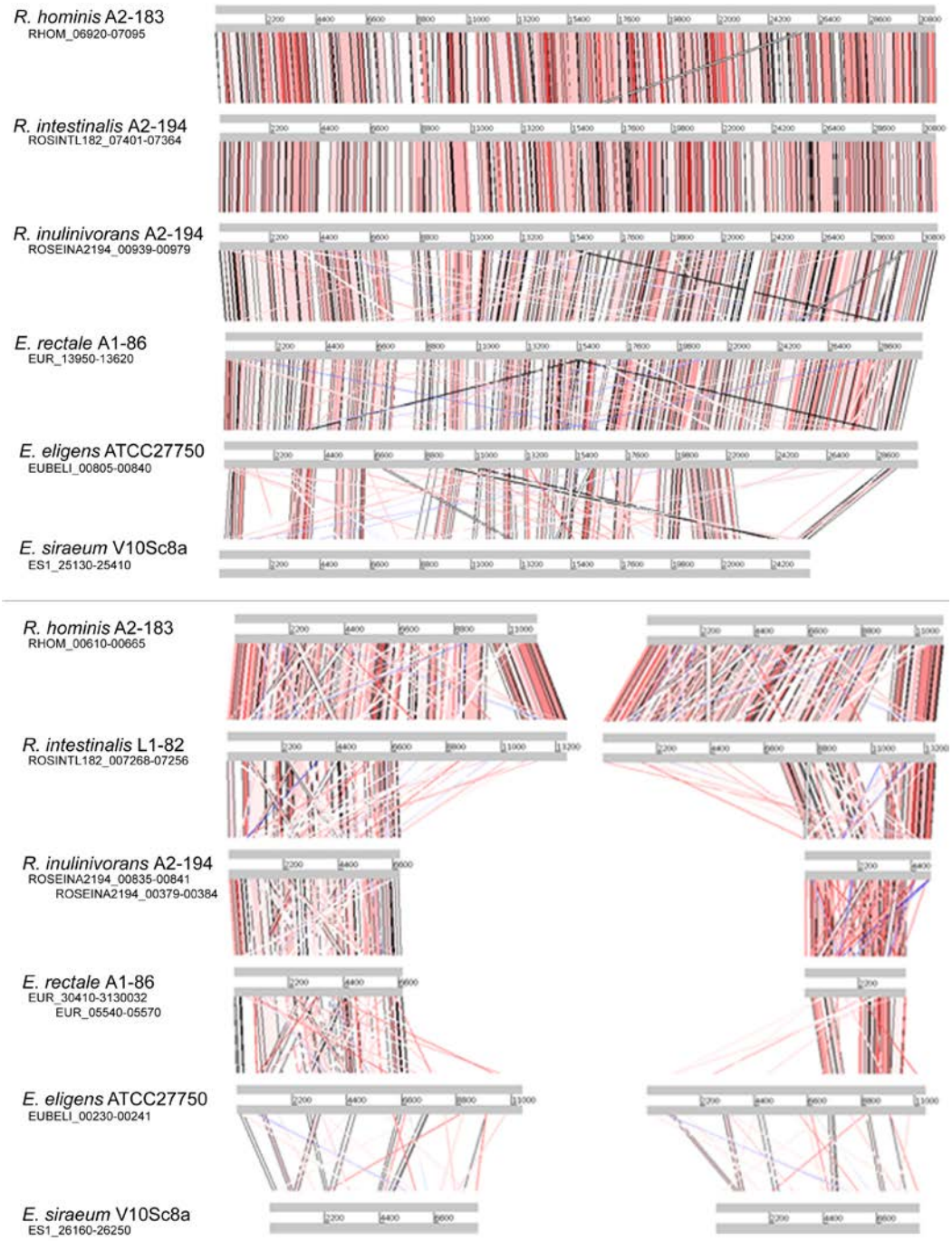


Figure S3.1: ACT alignments of *flgB-fliA* (top) and *flgM-flgN/fliC* (bottom) motility loci.

Locus tags indicate which genomic region is represented. A minimum threshold of 30 % identity was imposed on the alignments. Alignments involving *E. rectale* and *R. inulinivorans flgM-csrA* and *flaG-flgN/fliC* are on the bottom left and right, respectively.

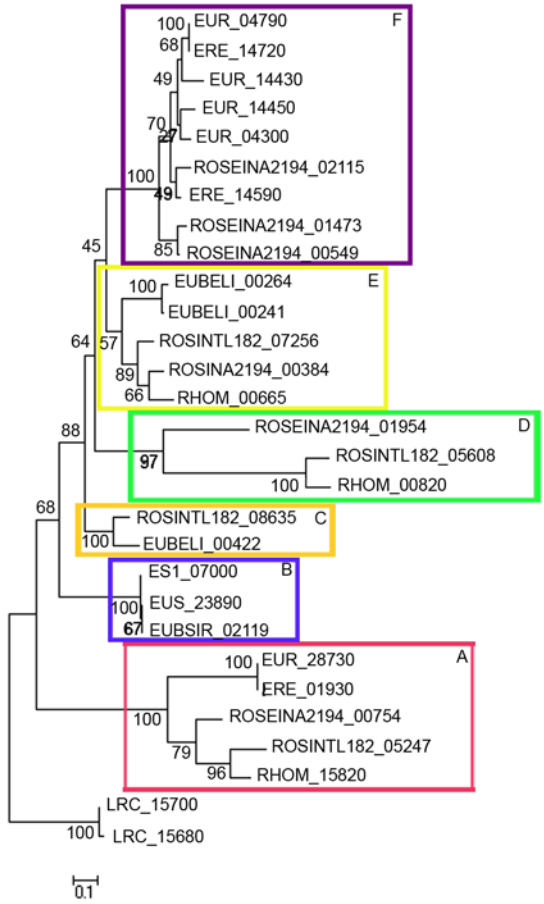
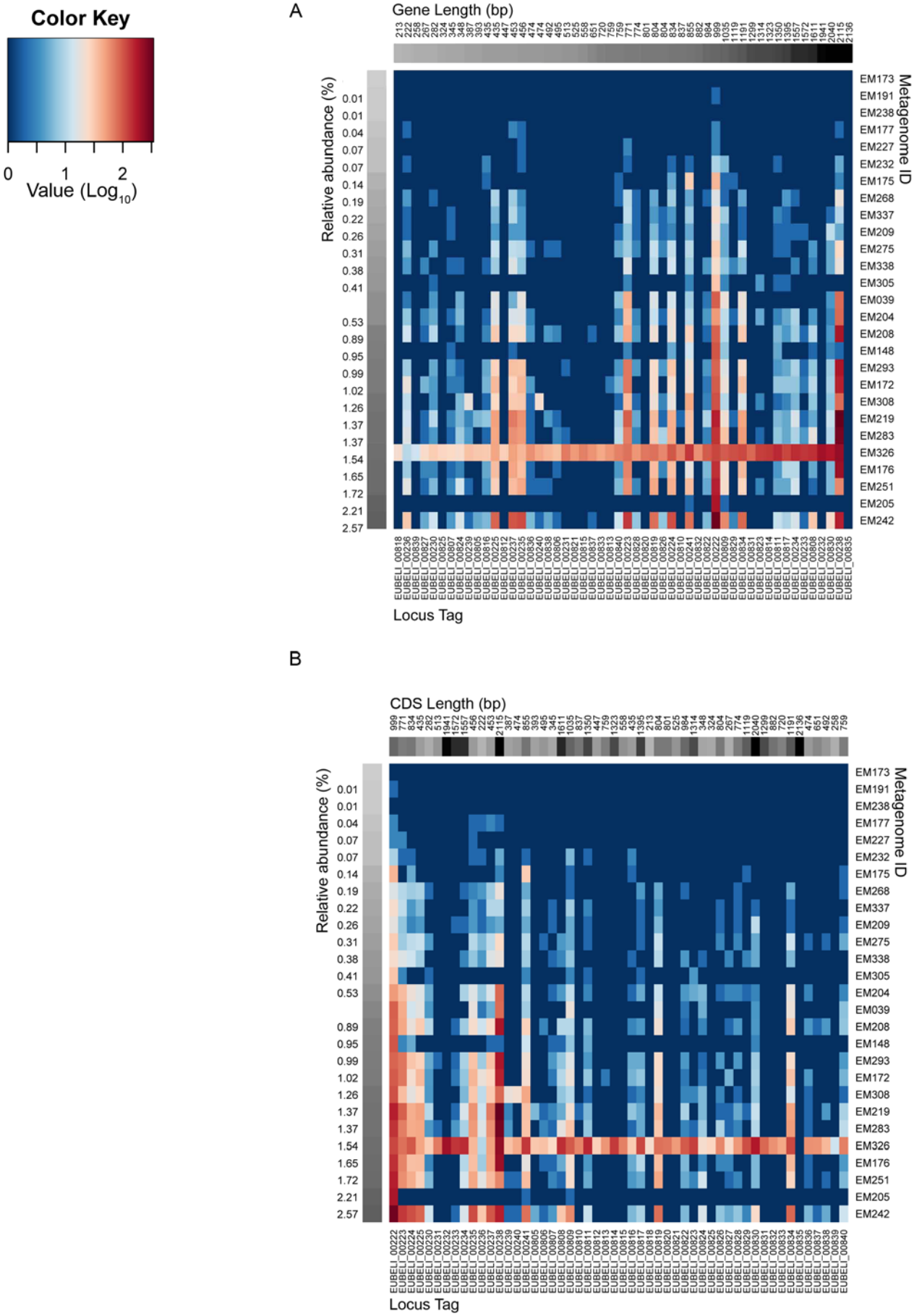


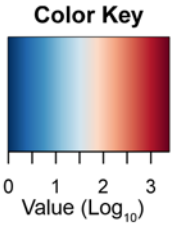
Figure S3.2: Phylogenetic tree of flagellin proteins.

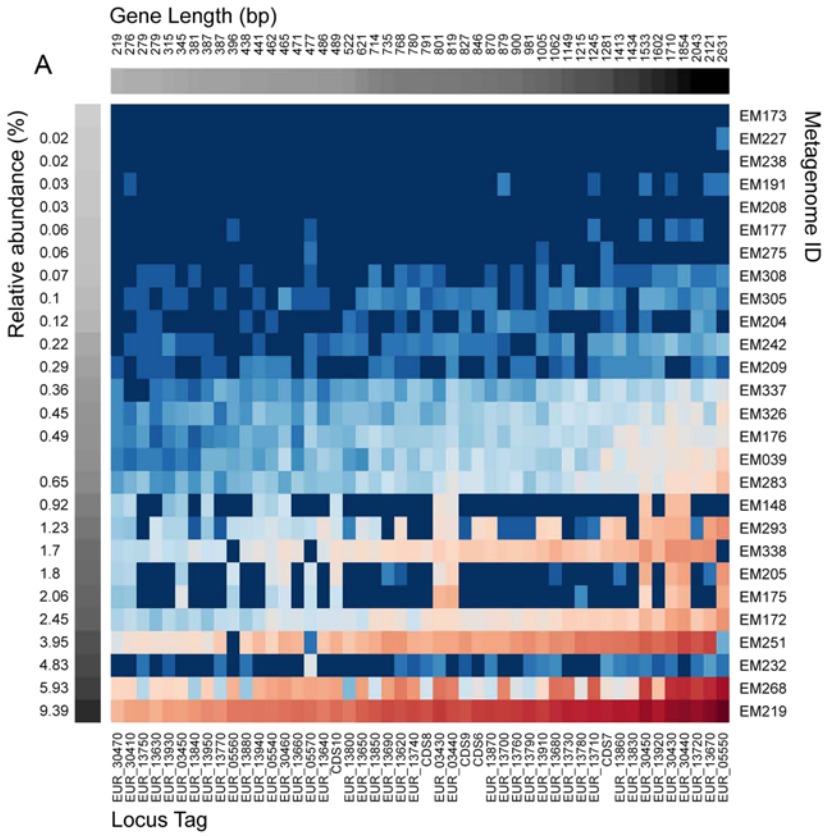
The flagellin tree was constructed from flagellin protein sequences using PHYML with model LG. Numbers at each node are bootstrap values. Locus tags were used to label flagellin proteins. Strongly supported clades (bootstrap ≥ 55) are surrounded by coloured boxes and are labelled with a letter A - F.

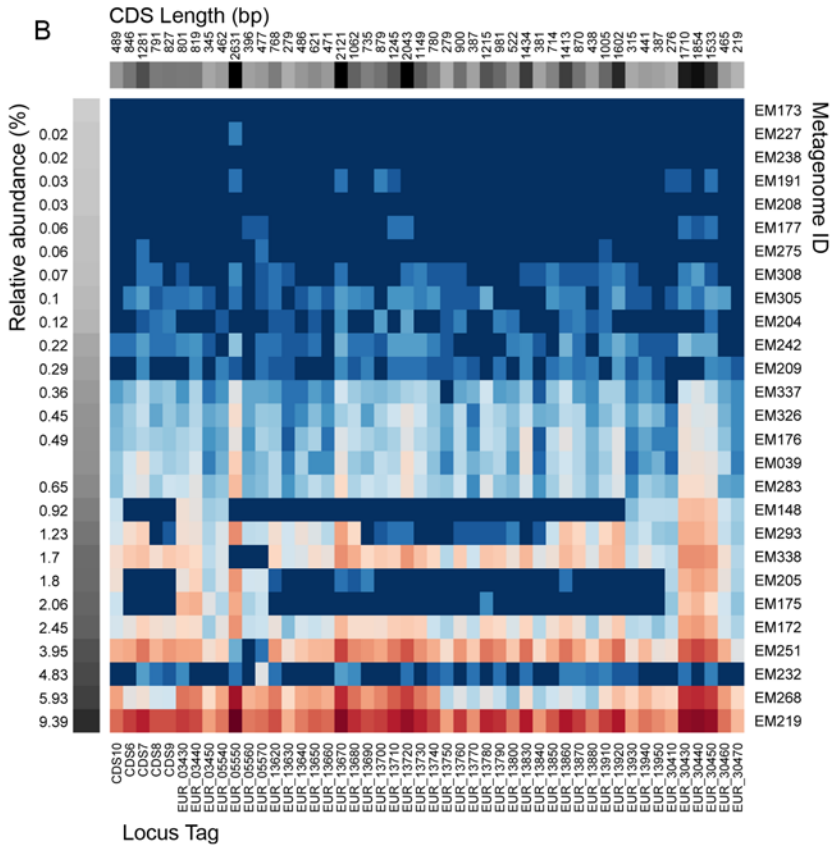
E. eligens


Color Key
0
1
2
Value (Log10)
A
Gene Length (bp)
Relative abundance (%)
Metagenome ID
Locus Tag
B
CDS Length (bp)
Relative abundance (%)
Metagenome ID
Locus Tag

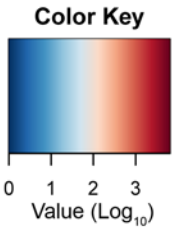
E. rectale

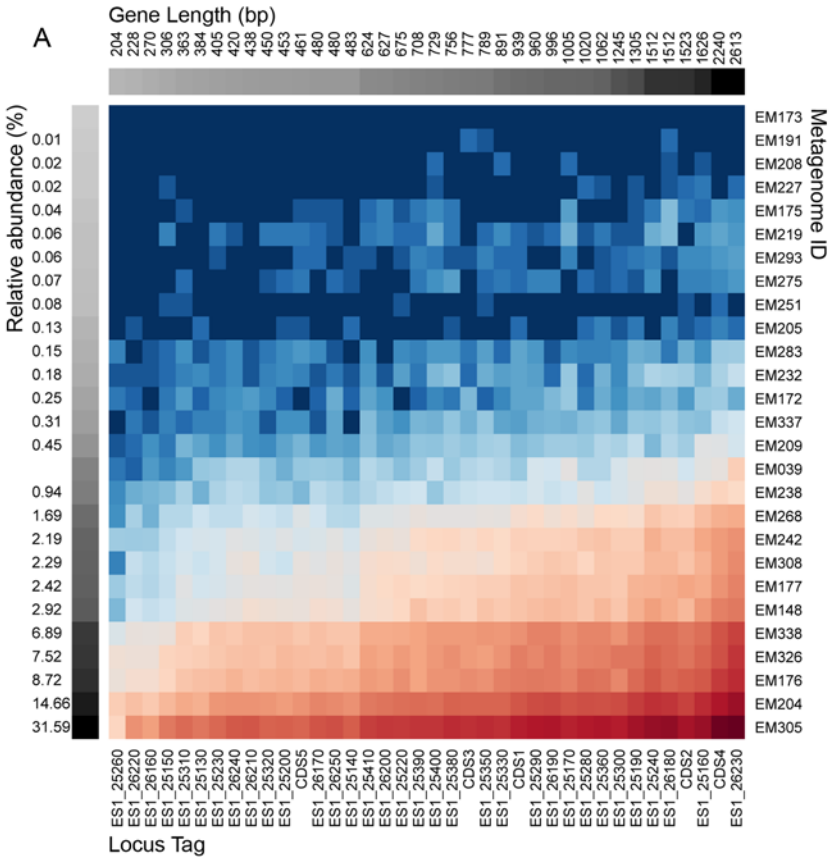

Color Key
0
1
2
3
Value (Log10)

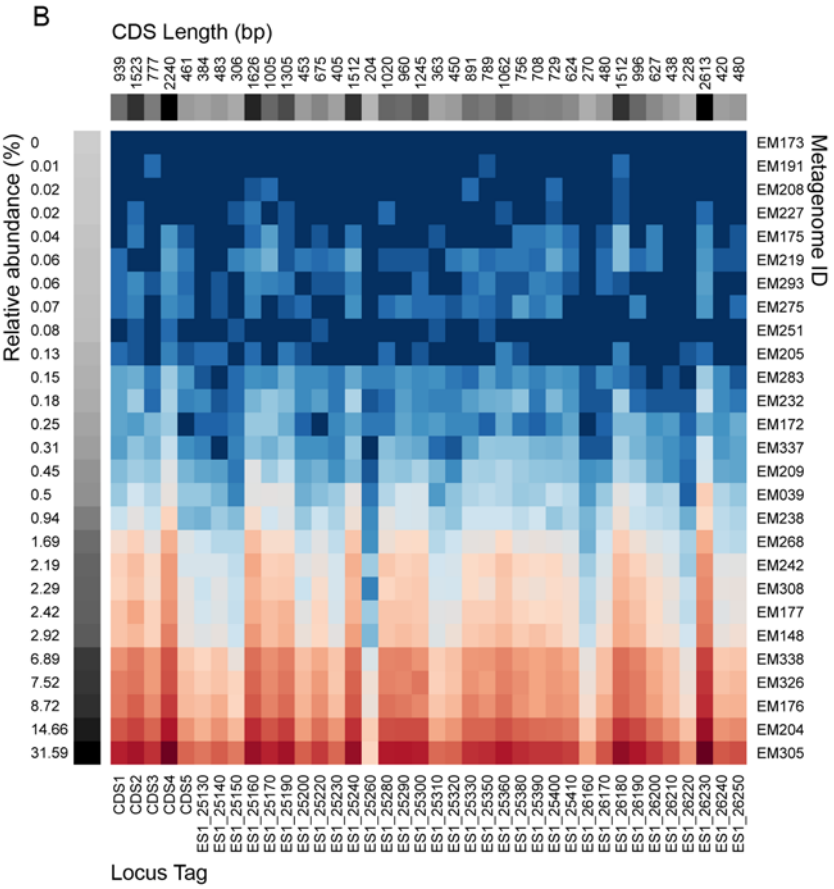

A
Gene Length (bp)
Relative abundance (%)
Metagenome ID
Locus Tag


B
CDS Length (bp)
Relative abundance (%)
Metagenome ID
Locus Tag

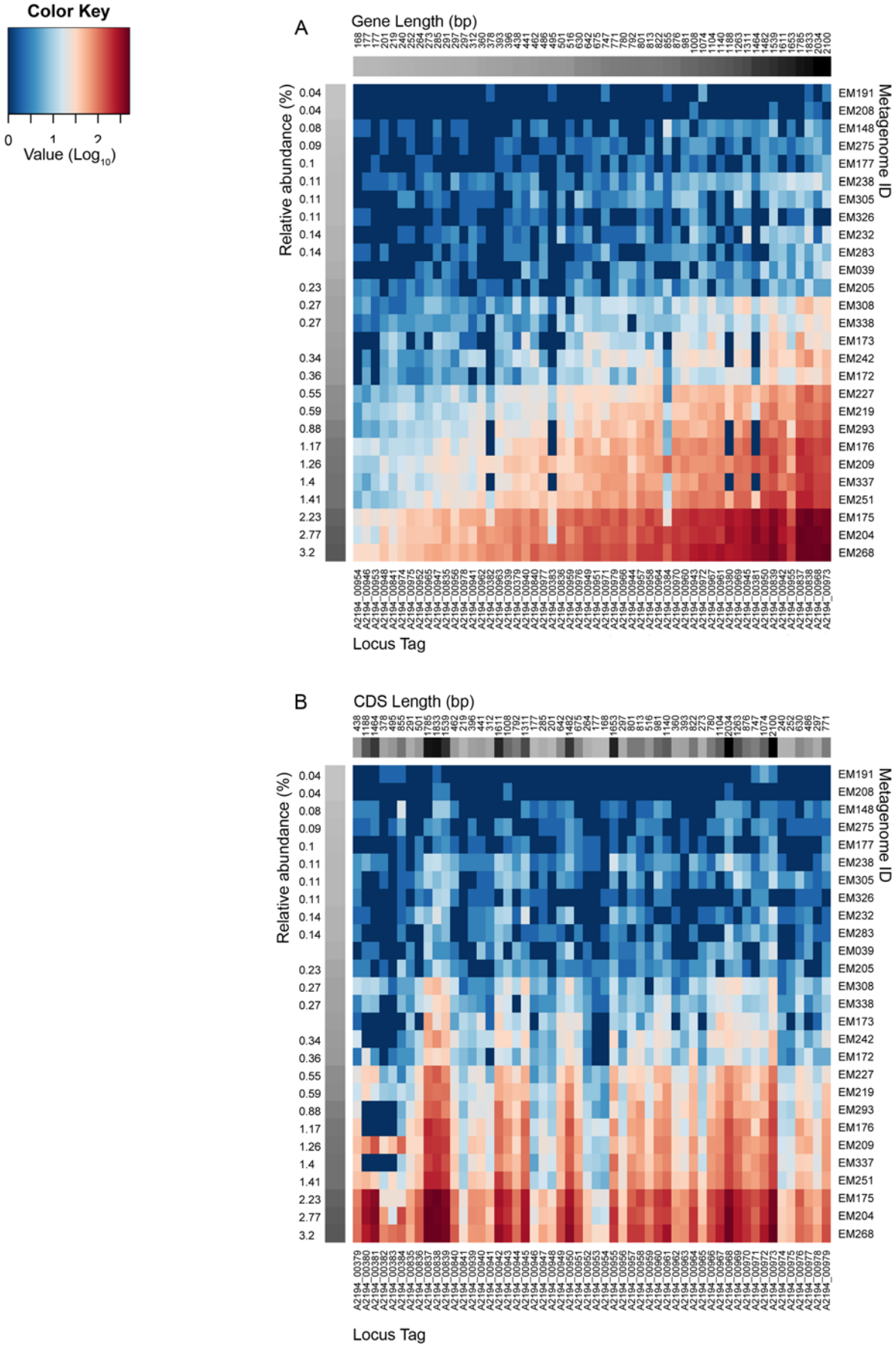
E. siraeum


Color Key
0
1
2
3
Value (Log10)


A
Gene Length (bp)
204 228 270 306 363 384 405 420 438 450 453 461 480 480 483 624 627 675 708 729 756 777 789 891 939 960 996 1005 1020 1062 1245 1305 1512 1512 1523 1626 2240 2613
Relative abundance (%)
0.01 0.02 0.02 0.04 0.06 0.06 0.07 0.08 0.13 0.15 0.18 0.25 0.31 0.45 0.94 1.69 2.19 2.29 2.42 2.92 6.89 7.52 8.72 14.66 31.59
EM173 EM191 EM208 EM227 EM175 EM219 EM293 EM275 EM251 EM205 EM283 EM232 EM172 EM337 EM209 EM039 EM238 EM268 EM242 EM308 EM177 EM148 EM338 EM326 EM176 EM204 EM305
Metagenome ID
ES1_25260 ES1_26220 ES1_26160 ES1_25150 ES1_25310 ES1_25130 ES1_25230 ES1_26240 ES1_26210 ES1_25320 ES1_25200 CDS5 ES1_26170 ES1_26250 ES1_25140 ES1_25410 ES1_26200 ES1_25220 ES1_25390 ES1_25400 ES1_25380 CDS3 ES1_25350 ES1_25330 CDS1 ES1_25290 ES1_26190 ES1_25170 ES1_25280 ES1_25360 ES1_25300 ES1_25190 ES1_25240 ES1_26180 CDS2 ES1_25160 CDS4 ES1_26230
Locus Tag


B
CDS Length (bp)
939 1523 777 2240 461 384 483 306 1626 1005 1305 453 675 405 1512 204 1020 960 1245 363 450 891 789 1062 756 708 729 624 270 480 1512 996 627 438 228 2613 420 480
Relative abundance (%)
0 0.01 0.02 0.02 0.04 0.06 0.06 0.07 0.08 0.13 0.15 0.18 0.25 0.31 0.45 0.5 0.94 1.69 2.19 2.29 2.42 2.92 6.89 7.52 8.72 14.66 31.59
EM173 EM191 EM208 EM227 EM175 EM219 EM293 EM275 EM251 EM205 EM283 EM232 EM172 EM337 EM209 EM039 EM238 EM268 EM242 EM308 EM177 EM148 EM338 EM326 EM176 EM204 EM305
Metagenome ID
CDS1 CDS2 CDS3 CDS4 CDS5 ES1_25130 ES1_25140 ES1_25150 ES1_25160 ES1_25170 ES1_25190 ES1_25200 ES1_25220 ES1_25230 ES1_25240 ES1_25260 ES1_25280 ES1_25290 ES1_25300 ES1_25310 ES1_25320 ES1_25330 ES1_25350 ES1_25360 ES1_25380 ES1_25390 ES1_25400 ES1_25410 ES1_26160 ES1_26170 ES1_26180 ES1_26190 ES1_26200 ES1_26210 ES1_26220 ES1_26230 ES1_26240 ES1_26250
Locus Tag

R. inulinivorans



R. intestinalis

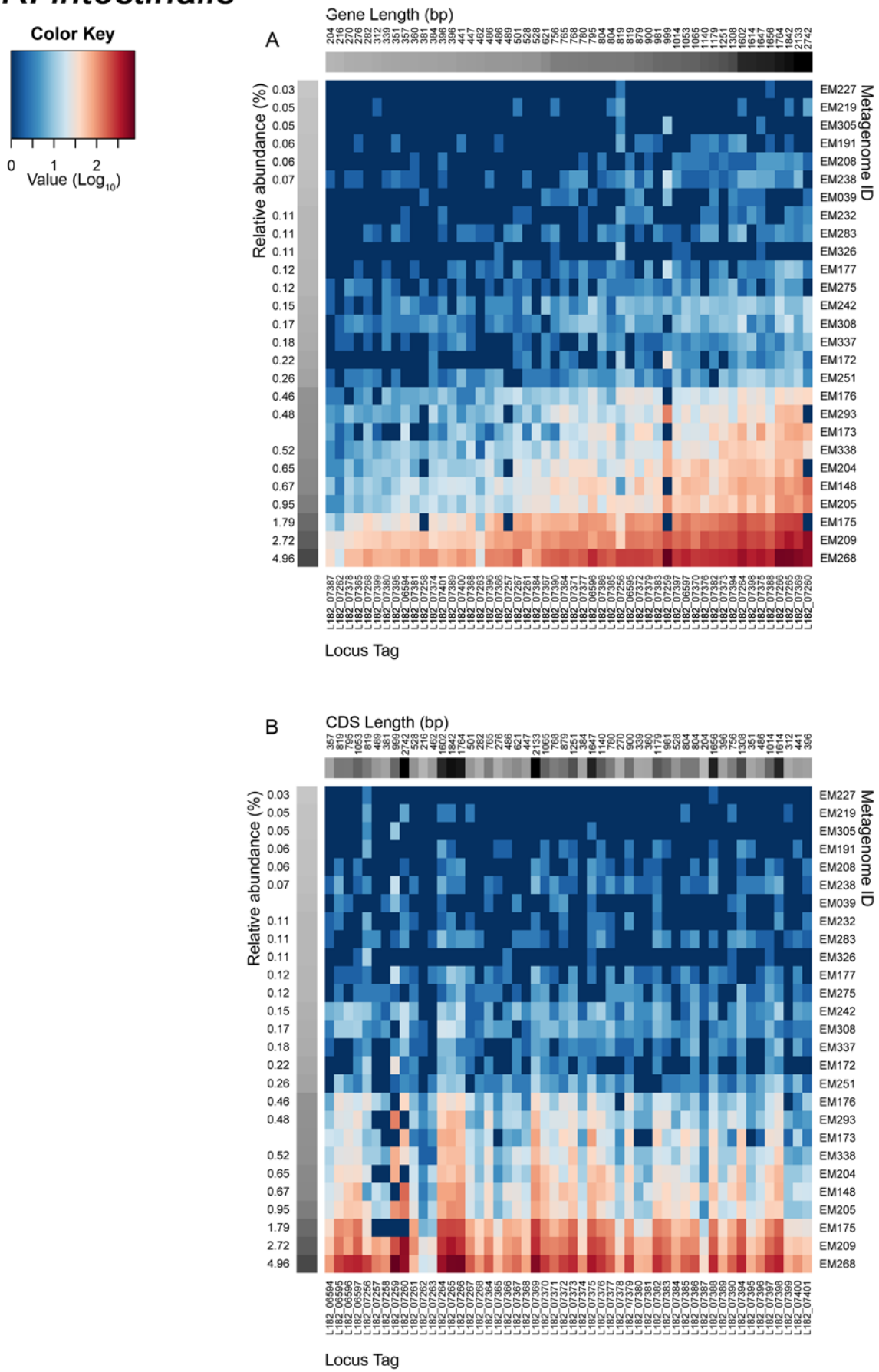


Figure S3.3: Heat-plots showing the relationship between the normalized number of reads mapped to target motility CDSs as a function of CDS length and target species relative abundance.

Heat-plots labelled “A” show that the normalized number of reads that mapped to each target gene increases with increasing CDS length and species relative abundance. Heat-plots labelled “B” show that the normalized number of reads that mapped to target CDSs varied depending on gene context. For each species, heat-plots A and B present the same data, but differ due to alternative arrangements of the CDSs on the X-axis. In heat-plots labelled “A”, CDSs are arranged according to increasing length, while in heat-plots labelled “B”, motility loci were organized by motility locus/gene context. CDSs without a locus tag were grouped together and not with the other CDSs of their respective motility loci (heat-plots B).

A


Similarity (%)
80
85
90
95
100
EUBELI_00422
EUBELI_00241
EUBELI_00264
EUBSIR_02119
EUS_23890
ES1_07000
ERE_14590
ERE_14720
ERE_12290
ERE_01930
EUR_04300
EUR_04790
EUR_14450
EUR_14430
EUR_28730
ROSEINA2194_00549
ROSEINA2194_00754
ROSEINA2194_01473
ROSEINA2194_01954
ROSEINA2194_02155
ROSEINA2194_00384
RHOM_00665
RHOM_00820
RHOM_15820
ROSINTL182_05608
ROSINTL182_08635
ROSINTL182_07256
ROSINTL182_05247
EM251
EM204
EM175
EM039
EM268
EM205
EM176
EM148
EM283
EM209
EM177
EM172

B


Similarity (%)
80
85
90
95
100
EUBELI_00422
EUBELI_00241
EUBELI_00264
EUBSIR_02119
EUS_23890
ES1_07000
ERE_14590
ERE_14720
ERE_12290
ERE_01930
EUR_04300
EUR_04790
EUR_14450
EUR_14430
EUR_28730
ROSEINA2194_00549
ROSEINA2194_00754
ROSEINA2194_01473
ROSEINA2194_01954
ROSEINA2194_02155
ROSEINA2194_00384
RHOM_00665
RHOM_00820
RHOM_15820
ROSINTL182_05608
ROSINTL182_08635
ROSINTL182_07256
ROSINTL182_05247
EM293
EM173
EM326
EM219
EM308
EM191
EM337
EM232
EM242
EM338
EM305

Figure S3.4: Recruitment plots demonstrating the presence or absence of the flagellin proteins of interest in 27 metagenomes..

A: Community dwelling individuals. B: Individuals from rehabilitation (EM219-EM238) and long-stay (EM173-EM308) community settings. Each plot shows matches with 80-100 % similarity to the query flagellin sequence, which are labelled with locus-tags. Matches in red are in the same orientation as the query sequence. Matches in blue are inverted relative to the query sequence. No matches were detected for metagenomes from four long-stay individuals (EM208, EM227, EM238 and EM275), so plots could not be constructed.

Table S3.1: Locus tags for motility loci from genomes of interest.

Protein	*E. rectale* A1-86	*E. eligens* ATCC27750	*R. intestinalis* L1-82	*R. hominis* A2-183	*R. inulinivorans* A2-194	*E. siraeum* V10Sc8a
Locus A						
FlgB	EUR_13950	EUBELI_00805	ROSINTL182_07401	RHOM_06920	ROSEINA2194_00939	ES1_25130
FlgC	EUR_13940	EUBELI_00806	ROSINTL182_07400	RHOM_06925	ROSEINA2194_00940	ES1_25140
FliE	EUR_13930	EUBELI_00807	ROSINTL182_07399	RHOM_06930	ROSEINA2194_00941	ES1_25150
FliF	EUR_13920	EUBELI_00808	ROSINTL182_07398	Present	ROSEINA2194_00942	ES1_25160
FliG	EUR_13910	EUBELI_00809	ROSINTL182_07397	RHOM_06945	ROSEINA2194_00943	ES1_25170
FliH	Present	EUBELI_00810	ROSINTL182_07396-07395	RHOM_06950	ROSEINA2194_00944	Present
FliI	Present	EUBELI_00811	ROSINTL182_07394	RHOM_06955	ROSEINA2194_00945	ES1_25190
FliJ	EUR_13880	EUBELI_00812	ROSINTL182_07393	RHOM_06960	ROSINA2194_00946-00947	ES1_25200
Hypothetical	Absent	Absent	Absent	Absent	ROSEINA2194_00948	Absent
Hypothetical	EUR_13870	EUBELI_00813	ROSINTL182_07392	RHOM_06965	ROSEINA2194_00949	Absent
FliK	EUR_13860	EUBELI_00814	ROSINTL182_07391	RHOM_06970	ROSEINA2194_00950	Present
FlgD	EUR_13850	EUBELI_00815	ROSINTL182_07390	RHOM_06975	ROSEINA2194_00951	ES1_25220
Hypothetical	Absent	Absent	Absent	Absent	ROSEINA2194_00952	Absent
FOP	EUR_13840	EUBELI_00816	ROSINTL182_07389	RHOM_06980	ROSEINA2194_00953-00954	ES1_25230
FlgE	EUR_13830	EUBELI_00817	ROSINTL182_07388	RHOM_06985	ROSEINA2194_00955	ES1_25240
FlbD	Absent	EUBELI_00818	ROSINTL182_07387	Absent	ROSEINA2194_00956	ES1_25260
MotA	Present	EUBELI_00819	ROSINTL182_07386	RHOM_06990	ROSEINA2194_00957	Present
MotB	Present	EUBELI_00820	ROSINTL182_07385	RHOM_06995	ROSEINA2194_00958	ES1_25280
FliL	EUR_13800	EUBELI_00821	ROSINTL182_07384	RHOM_07000	ROSEINA2194_00959	Absent
FliM	EUR_13790	EUBELI_00822	ROSINTL182_07383	RHOM_07005	ROSEINA2194_00960	ES1_25290
FliN	EUR_13780	EUBELI_00823	ROSINTL182_07382	RHOM_07010	ROSEINA2194_00961	ES1_25300
CheY-like	Absent	EUBELI_00824	ROSINTL182_07381	RHOM_07015	ROSEINA2194_00962	ES1_25310
FliO	EUR_13770	EUBELI_00825	ROSINTL182_07380	RHOM_07020	ROSEINA2194_00963	ES1_25320
FliP	EUR_13760	EUBELI_00826	ROSINTL182_07379	RHOM_07025	ROSEINA2194_00964	ES1_25330
FliQ	EUR_13750	EUBELI_00827	ROSINTL182_07378	RHOM_07030	ROSEINA2194_00965	Present
FliR	EUR_13740	EUBELI_00828	ROSINTL182_07377	RHOM_07035	ROSEINA2194_00966	ES1_25350
FlhB	EUR_13730	EUBELI_00829	ROSINTL182_07376	RHOM_07040	ROSEINA2194_00967	ES1_25360
FlhA	EUR_13720	EUBELI_00830	ROSINTL182_07375-07374	RHOM_07045	ROSEINA2194_00968	Present
FlhF	EUR_13710	EUBELI_00831	ROSINTL182_07373	RHOM_07050	ROSEINA2194_00969	Absent
FlhG	EUR_13700	EUBELI_00832	ROSINTL182_07372	RHOM_07055	ROSEINA2194_00970	Absent
PilZ	EUR_13690	EUBELI_00833	ROSINTL182_07371	RHOM_07060	ROSEINA2194_00971	Absent
CheB	EUR_13680	EUBELI_00834	ROSINTL182_07370	RHOM_07065	ROSEINA2194_00972	Absent
CheA	EUR_13670	EUBELI_00835	ROSINTL182_07369	RHOM_07070	ROSEINA2194_00973	Absent
CheW	EUR_13660	EUBELI_00836	ROSINTL182_07368	RHOM_07075	ROSEINA2194_00974	Absent
CheC	EUR_13650	EUBELI_00837	ROSINTL182_07367	RHOM_07080	ROSEINA2194_00976	ES1_25410
CheD	EUR_13640	EUBELI_00838	ROSINTL182_07366	RHOM_07085	ROSEINA2194_00977	ES1_25420*
Hypo	EUR_13630	EUBELI_00839	ROSINTL182_07365	RHOM_07090	ROSEINA2194_00978	Absent
FliA	EUR_13620	EUBELI_00840	ROSINTL182_07364	RHOM_07095	ROSEINA2194_00979	ES1_25380
Accession No.	FP929042.1	NC_012788.1	ABYJ02000109.1 ABYJ02000108.1	CP003040	ACFY01000039.1	FP929059.1
TOTAL LENGTH (nt)	30520	31143	31347	31496	31444	26329
TOTAL CDS	34	36	36	36	38	29

Protein	*E. rectale* A1-86	*E. eligens* ATCC27750	*R. intestinalis* L1-82	*R. hominis* A2-183	*R. inulinivorans* A2-194	*E. siraeum* V10Sc8a
Locus B						
MBL	Absent	EUBELI_00222	ROSINTL182_06597	RHOM_12465	ROSEINA2194_03009	-
FlgF	EUR_03430	EUBELI_00223	ROSINTL182_06596	RHOM_12460	ROSEINA2194_03007-03008	ES1_25390
FlgG	EUR_03440	EUBELI_00224	ROSINTL182_06595	RHOM_12455	ROSEINA2194_03006	ES1_25400
FlgJ	EUR_03450	EUBELI_00225	ROSINTL182_06594	RHOM_12450	ROSEINA2194_03005	-
Accession No.	FP929042.1	NC_012788.1	NZ_ABYJ02000069.1	NC_015977	ACFY01000115	
TOTAL LENGTH (nt)	1984	3195	3220	3158	3119	
TOTAL CDS	3	4	4	4	5 (4)	
Locus C						
FlaG	EUR_05540	EUBELI_00237	ROSINTL182_07261	RHOM_00645	ROSEINA2194_00379	Absent
FliD	EUR_05550	EUBELI_00238	ROSINTL182_07260	RHOM_00650	ROSEINA2194_00380-00381	ES1_26230
Transposase	Absent	Absent	ROSINTL182_07259	Absent	Absent	Absent
FliS	EUR_05560	EUBELI_00239	ROSINTL182_07258	RHOM_00655	ROSEINA2194_00382	ES1_26240
FlgN	EUR_05570	EUBELI_00240	ROSINTL182_07257	RHOM_00660	ROSEINA2194_00383	ES1_26250
Flagellin	Absent	EUBELI_00241	ROSINTL182_07256	RHOM_00665	ROSEINA2194_00384	Absent
Accession No.	FP929042.1	NC_012788.1	ABYJ02000104.1	CP003040	ACFY01000000	FP929059.1
TOTAL LENGTH (nt)	6764 & 4152	11440	13616	12130	6871 & 5165	8351
TOTAL CDS	7 & 4	12	13	12	7 & 5	10

* Locus tag inferred based on its position immediately after ES1_25410.

Table S3.2: Amino-terminal sequences of *E. rectale* A1-86 and *R. inulinivorans* A2-194 flagellin proteins.

Flagellin source & Size	1	2	3	4	5	6	7	8	9	10
E. rectale ~ 52 kDa	M	V	V	Q	H	N	M	Q	A	A
R. inulinivorans ~ 51 kDa	M	V	V	Q	H	N	M	Q	A	A
R. inulinivorans ~ 43 kDa	M	V	V	Q	H	N	M	Q	A	A/M
R. inulinivorans ~ 29 kDa	M	V/Q	V/P	Q	H	N	N/M	T/M	A/T	A

Table S3.3: Relative abundance (%) of each target species in 25 of the shotgun metagenomes of interest, as calculated by MetaPhlAn.

Metagenome	Location	*E. eligens*	*E. rectale*	*E. siraeum*	*R. intestinalis*	*R. inulinivorans*
EM148	Community	0.95	0.92	2.92	0.67	0.08
EM172	Community	1.02	2.45	0.25	0.22	0.36
EM175	Community	0.14	2.06	0.04	1.79	2.23
EM176	Community	1.65	0.49	8.72	0.46	1.17
EM177	Community	0.04	0.06	2.42	0.12	0.10
EM204	Community	0.53	0.12	14.66	0.65	2.77
EM205	Community	2.21	1.80	0.13	0.95	0.23
EM209	Community	0.26	0.29	0.45	2.72	1.26
EM251	Community	1.72	3.95	0.08	0.26	1.41
EM268	Community	0.19	5.93	1.69	4.96	3.20
EM283	Community	1.37	0.65	0.15	0.11	0.14
EM219	Rehabilitation	1.37	9.39	0.06	0.05	0.59
EM232	Rehabilitation	0.07	4.83	0.18	0.11	0.14
EM305	Rehabilitation	0.41	0.10	31.59	0.05	0.11
EM326	Rehabilitation	1.54	0.45	7.52	0.11	0.11
EM337	Rehabilitation	0.22	0.36	0.31	0.18	1.40
EM338	Rehabilitation	0.38	1.70	6.89	0.52	0.27
EM191	Longstay	0.01	0.03	0.01	0.06	0.04
EM208	Longstay	0.89	0.03	0.02	0.06	0.04
EM227	Longstay	0.07	0.02	0.02	0.03	0.55
EM238	Longstay	0.01	0.02	0.94	0.07	0.11
EM242	Longstay	2.57	0.22	2.19	0.15	0.34
EM275	Longstay	0.31	0.06	0.07	0.12	0.09
EM293	Longstay	0.99	1.23	0.06	0.48	0.88
EM308	Longstay	1.26	0.07	2.29	0.17	0.27

Note: Due to different assembly procedures, EM039 and EM173 were excluded from MetaPhlAn analysis.

Table S3.4: Estimated target genome coverage in each metagenome.

Metagenome	Species	Rel. Abundance target species (%)	Estimated coverage
EM148	*E. eligens*	0.95	16.45
4795.11 Mb	*E. rectale*	0.92	12.62
	E. siraeum	2.92	49.53
	R. intestinalis	0.67	7.52
	R. inulinivorans	0.08	0.95
EM172	*E. eligens*	1.02	16.47
4472.60 Mb	*E. rectale*	2.45	31.34
	E. siraeum	0.25	3.96
	R. intestinalis	0.22	2.30
	R. inulinivorans	0.36	3.98
EM175	*E. eligens*	0.14	2.41
4772.91 Mb	*E. rectale*	2.06	28.12
	E. siraeum	0.04	0.68
	R. intestinalis	1.79	20.01
	R. inulinivorans	2.23	26.28
EM176	*E. eligens*	1.65	28.48
4780.81 Mb	*E. rectale*	0.49	6.70
	E. siraeum	8.72	147.48
	R. intestinalis	0.46	5.15
	R. inulinivorans	1.17	13.81
EM177	*E. eligens*	0.04	0.69
4776.79 Mb	*E. rectale*	0.06	0.82
	E. siraeum	2.42	40.90
	R. intestinalis	0.12	1.34
	R. inulinivorans	0.1	1.18
EM204	*E. eligens*	0.53	9.14
4779.02 Mb	*E. rectale*	0.12	1.64
	E. siraeum	14.66	247.86
	R. intestinalis	0.65	7.27
	R. inulinivorans	2.77	32.69
EM205	*E. eligens*	2.21	38.19
4786.83 Mb	*E. rectale*	1.8	24.64
	E. siraeum	0.13	2.20
	R. intestinalis	0.95	10.65
	R. inulinivorans	0.23	2.72
EM209	*E. eligens*	0.26	4.49
4788.57 Mb	*E. rectale*	0.29	3.97
	E. siraeum	0.45	7.62
	R. intestinalis	2.72	30.50
	R. inulinivorans	1.26	14.90
EM251	*E. eligens*	1.72	29.72
4786.59 Mb	*E. rectale*	3.95	54.07
	E. siraeum	0.08	1.35
	R. intestinalis	0.26	2.91
	R. inulinivorans	1.41	16.66
EM268	*E. eligens*	0.19	3.29
4789.76 Mb	*E. rectale*	5.93	81.23
	E. siraeum	1.69	28.64

	R. intestinalis	4.96	55.64
	R. inulinivorans	3.2	37.84
EM283	*E. eligens*	1.37	23.68
4788.08 Mb	*E. rectale*	0.65	8.90
	E. siraeum	0.15	2.54
	R. intestinalis	0.11	1.23
	R. inulinivorans	0.14	1.66
EM219	*E. eligens*	1.37	23.72
4796.21 Mb	*E. rectale*	9.39	128.80
	E. siraeum	0.06	1.02
	R. intestinalis	0.05	0.56
	R. inulinivorans	0.59	6.99
EM232	*E. eligens*	0.07	1.21
4794.20 Mb	*E. rectale*	4.83	66.22
	E. siraeum	0.18	3.05
	R. intestinalis	0.11	1.24
	R. inulinivorans	0.14	1.66
EM305	*E. eligens*	0.41	7.09
4793.44 Mb	*E. rectale*	0.1	1.37
	E. siraeum	31.59	535.70
	R. intestinalis	0.05	0.56
	R. inulinivorans	0.11	1.30
EM326	*E. eligens*	1.54	26.66
4795.95 Mb	*E. rectale*	0.45	6.17
	E. siraeum	7.52	127.59
	R. intestinalis	0.11	1.24
	R. inulinivorans	0.11	1.30
EM337	*E. eligens*	0.22	3.81
4798.09 Mb	*E. rectale*	0.36	4.94
	E. siraeum	0.31	5.26
	R. intestinalis	0.18	2.02
	R. inulinivorans	1.4	16.59
EM338	*E. eligens*	0.38	6.58
4793.68 Mb	*E. rectale*	1.7	23.31
	E. siraeum	6.89	116.85
	R. intestinalis	0.52	5.84
	R. inulinivorans	0.27	3.20
EM191	*E. eligens*	0.01	0.17
4789.62 Mb	*E. rectale*	0.03	0.41
	E. siraeum	0.01	0.17
	R. intestinalis	0.06	0.67
	R. inulinivorans	0.04	0.47
EM208	*E. eligens*	0.89	15.39
4790.74 Mb	*E. rectale*	0.03	0.41
	E. siraeum	0.02	0.34
	R. intestinalis	0.06	0.67
	R. inulinivorans	0.04	0.47
EM227	*E. eligens*	0.02	0.35
4794.62 Mb	*E. rectale*	0.02	0.27
	E. siraeum	0.02	0.34
	R. intestinalis	0.03	0.34
	R. inulinivorans	0.55	6.51
EM238	*E. eligens*	0.01	0.17
4793.43 Mb	*E. rectale*	0.02	0.27

	E. siraeum	0.94	15.94
	R. intestinalis	0.07	0.79
	R. inulinivorans	0.11	1.30
EM242	*E. eligens*	2.57	44.50
4796.50 Mb	*E. rectale*	0.22	3.02
	E. siraeum	2.19	37.16
	R. intestinalis	0.15	1.68
	R. inulinivorans	0.34	4.03
EM275	*E. eligens*	0.31	5.37
4796.45 Mb	*E. rectale*	0.06	0.82
	E. siraeum	0.07	1.19
	R. intestinalis	0.12	1.35
	R. inulinivorans	0.09	1.07
EM293	*E. eligens*	0.99	17.14
4795.73 Mb	*E. rectale*	1.23	16.87
	E. siraeum	0.06	1.02
	R. intestinalis	0.48	5.39
	R. inulinivorans	0.88	10.42
EM308	*E. eligens*	1.26	21.81
4794.98 Mb	*E. rectale*	0.07	0.96
	E. siraeum	2.29	38.85
	R. intestinalis	0.17	1.91
	R. inulinivorans	0.27	3.20

Note: Average target genome sizes in Mb are as follows: *E. eligens* 2.27, *E. rectale*, 3.5, *E. siraeum* 2.83, *R. intestinalis*, 4.27, *R. inulinivorans*, 4.05.

Table S3.5: Summary of the number of ORFs per metagenome identified as a motility gene or gene fragment from a species of interest.

Metagenome ID	*E. eligens*	*E. rectale*	*E. siraeum*	*R. hominis*	*R. intestinalis*	*R. inulinivorans*
EM039	17	11	37	-	-	-
EM148	-	-	41	-	-	-
EM172	2	2	-	-	-	-
EM175	-	-	-	-	20	41
EM176	-	-	37	-	-	-
EM177	-	-	34	-	-	-
EM204	-	-	39	-	-	52
EM205	-	10	-	-	1	1
EM209	-	-	-	-	42	28
EM251	-	44	-	-	-	-
EM268	-	23	42	-	45	44
EM283	5	-	-	-	-	-
EM219	-	37	-	-	-	-
EM232	-	-	-	37	-	-
EM305	-	-	34	-	-	-
EM326	9	-	38	-	-	-
EM337	-	-	-	1	-	22
EM338	-	39	35	-	-	-
EM173	-	-	-	-	11	5
EM191	-	-	-	1	-	-
EM208	-	-	-	-	-	-
EM227	-	-	-	-	-	-
EM238	-	-	-	-	-	-
EM242	-	-	28	-	-	-
EM275	-	-	-	-	-	-
EM293	1	-	-	-	-	-
EM308	-	-	34	-	-	-

Table S3.6: "Cell Motility" COG category analysis for assembled metagenomes.

Metagenome	Total COGs represented	No. different "Cell Motility" COGS represented	Cell motility as % of total COGs	No. sequences in COG1344
EM039	1893	13	0.69	12
EM148	1431	5	0.35	-
EM172	1808	12	0.66	14
EM175	1754	9	0.51	7
EM176	1731	14	0.81	29
EM177	2026	12	0.59	4
EM204	1674	7	0.42	-
EM205	1836	16	0.87	23
EM209	1765	9	0.51	11
EM251	1520	11	0.72	7
EM268	1671	10	0.60	8
EM283	2235	19	0.85	16
EM219	1834	9	0.49	4
EM232	1764	7	0.40	3
EM305	1839	14	0.76	8
EM326	1717	15	0.87	7
EM337	1972	16	0.81	8
EM338	1921	15	0.78	11
EM173	1777	9	0.51	7
EM191	1770	13	0.73	3
EM208	1535	9	0.59	2
EM227	1554	2	0.13	-
EM238	1661	3	0.18	1
EM242	1545	9	0.58	12
EM275	1817	7	0.39	1
EM293	1503	5	0.33	3
EM308	1820	8	0.44	-

Table S3.7: Strains and Genomes used in this study.

Species	Strain	Isolated from:	Genome accession No.	No. Contigs	Contig N50 (nt)	No. Scaffolds	Scaffold N50 (nt)	Size (nt)	Genome Publication Reference (PMID)	Strain Publication Reference (PMID)
E. eligens	ATCC27750	Human faeces	CP001104.1; CP001105.1; CP001106.1	Complete Genome	-	3	-	2,144,190; 60,455; 62,6744	19321416	Holdeman and Moore, 1974
E. rectale	A1-86	Human faeces	FP929042.1	n/a	n/a	n/a	n/a	3,344,951	Unpublished	18599726
	M104/1	GI tract	FP929043.1	n/a	n/a	n/a	n/a	3,698,419	Unpublished	Unpublished
E. siraeum	V10Sc8a	Human faeces	FP929059.1	n/a	n/a	n/a	n/a	2,836,123	Unpublished	Unpublished
	DSM_15702	Human faeces	ABCA00000000.3	55	104,065	44	112,958	2,697,034	Unpublished	Moore *et al.*, 1976
	70/3	Human faeces	FP929044.1	n/a	n/a	n/a	n/a	2,943,413	Unpublished	18537837
R. hominis	A2-183	Human faeces	CP003040.1	Complete Genome	-	1	-	3,592,125	Unpublished	10712576
R. intestinalis	L1-82	Human faeces	ABYJ00000000.2	409	29,464	102	123,125	4,411,375	Unpublished	12361264
R. inulinivorans	A2-194	Human faeces	ACFY01000000	179	57,343	n/a	n/a	4,048,462	Unpublished	10712576

Chapter 4

General Discussion

4.1 General Discussion.

This thesis aimed to characterise the motility genotypes and phenotypes of flagellate bacteria that are commensal to the intestines of mammals, including humans. By focusing on *Lactobacillus ruminis* and closely related motile species of the *L. salivarius* clade, the research described in Chapter 2 provided the first in-depth characterisation of the *Lactobacillus* genes and proteins involved in flagellum-biogenesis. In Chapter 3, the motility genotypes of six *Eubacterium* and *Roseburia* species were examined by comparative genomics from which inferences regarding the regulation and expression of the motility phenotypes of these species could be drawn. Given their potential as immunomodulatory molecules, the flagellin proteins from these *Lactobacillus* and selected *Eubacterium* and *Roseburia* species were investigated, and in each case, the flagellin proteins were found to elicit a pro-inflammatory response characterised by IL-8 secretion from cultured human IECs. In Chapter 3, the flagellum-biogenesis genes of *Eubacterium* and *Roseburia* species were used to investigate if a relationship existed between the relative abundance of a target species in the intestinal microbiota and the ability to detect its genes for a specific pathway in the intestinal microbiome. The elucidation of this relationship was probably hindered by incomplete coverage of the target genomes in the metagenome database, leading to the conclusion that a greater depth of sequence data would be required to capture the full functional capacity of the intestinal microbiota at the level of species or strain. The motility phenotype of these commensal species is nevertheless a physiological feature of note, with potential as a host-interaction and niche-colonisation trait.

The genus *Lactobacillus* is a recognised source of bacteria for use in the food-industry and as probiotic supplements. Accordingly, the *Lactobacillus* species used for these purposes have been rigorously characterised. Several appendages and molecules may adorn the *Lactobacillus* cell surface, including pili (Kankainen *et al.*, 2009), exopolysaccharides (Sims *et al.*, 2011) and sortase-dependent surface proteins (van Pijkeren *et al.*, 2006). These appendages and extracellular molecules could facilitate interactions between the bacterium and its host, and may underlie the probiotic attributes of a particular strain. Therefore, as prominent cell-surface organelles, the flagella of *L. ruminis* and the other motile lactobacilli merit attention.

As discussed in Chapter 2, although more than one-hundred species are formally recognised in the genus *Lactobacillus,* only 12-14 are reported to be motile by means of flagella, most of which were first isolated and described in recent years (Table S2.1). Accordingly, the motility genotypes and phenotypes of these *Lactobacillus* species were not well characterised until now. As the majority of the known motile species belong to the *L. salivarius* clade as determined by 16S rRNA gene phylogeny (Felis & Dellaglio, 2007), flagellum-mediated motility is a trait of limited phylogenetic distribution within the genus *Lactobacillus*. A separate study which followed-on from the initial phylogenetic investigation on the origins of the *L. ruminis* motility genes that was presented in Chapter 2, confirmed that motility genes in the order *Lactobacillales* were inherited by horizontal gene transfer from the genus *Clostridium* (Forde, 2013). Thus, it seems likely that the last common ancestor of the lactobacilli inherited its motility locus vertically, and that a number of gene-loss events ensued as species diverged.

However, at least one *L. curvatus* strain, (a kind gift from Dr. A. Endo) is demonstrably motile under standard laboratory conditions. On the basis of 16S rRNA gene phylogeny, *L. curvatus* is more closely related to *L. sakei* than to *L. salivarius* (Felis & Dellaglio, 2007). Interestingly, Torriani and colleagues (Torriani *et al.*, 1996) describe *L. curvatus* subsp. *curvatus* and *L. curvatus* subsp. *melibosus* and mention that they are motile, but lose their motility upon repeated subculture. *L. curvatus* subsp. *melibiosus* was subsequently recognised as a later synonym of *L. sakei* subsp. *carnosus* (Koort *et al.*, 2004). Thus, *L. curvatus* and *L. sakei* subsp. *carnosus* are potentially motile. However, the strains described by Torriani and colleagues were not deposited in an international culture collection and were not available for motility analysis.

The accumulation of publically-available draft genome sequences for *Lactobacillus* species means that a genus-wide evaluation of the distribution of motility genes is possible. Of the 49 *Lactobacillus* species for which genome information is currently available, the genomes of only four species, specifically *L. ruminis* (x 4 genomes), *L. mali* (x 2), *L. acidipiscis* (x 1) and *L. vini* (x 1) harbour flagellum-biogenesis genes. However, the annotation of motility genes in the genome of a particular strain is insufficient evidence to infer a motile phenotype. Each of the nine *L. ruminis* strains studied in Chapter 2 harboured genes for flagellum synthesis, but six of these strains were non-motile and aflagellate under standard culture conditions. Flagellum-biogenesis genes are not present in the genomes of *L. curvatus* CRL705 and *L. sakei* 23K and these strains are not known to be motile. These observations nevertheless highlight the issue of intra-species variation in genotype and phenotype, and emphasise the

importance of strain-specific characterisation over generalised inferences of genotypes and phenotypes which could be misleading.

Genome sequences have a significant predictive-value which can be exploited to inform and direct research hypotheses. In Chapter 3, aspects of the regulation of motility gene expression in several presumed and verifiably motile *Eubacterium* and *Roseburia* species were inferred from an investigation of the motility gene content and arrangement in the draft genome sequences of these species. A similar approach was used in Chapter 2 to propose a model for the regulation of flagellum biogenesis in *L. ruminis.*

By comparing the motility gene arrangement of these species with that described for *L. ruminis* in Chapter 2, differences in the regulation and phenotypic expression of flagellum-biogenesis in these systems could be discerned. For example, unlike *L. ruminis* which is peritrichously flagellate, the *Eubacterium* and *Roseburia* species bear sub-terminal polar flagella. This is reflected in the annotation of genes controlling the polar localisation of flagella, (*flhF, flhG*) in the *Eubacterium-Roseburia* motility loci.

Furthermore, while the arrangement of genes at each locus in the *Lactobacillus, Eubacterium* and *Roseburia* genomes examined tended to reflect the predicted hierarchical order of their expression, the overall regulation of the *Lactobacillus* and *Eubacterium-Roseburia* motility systems are expected to differ. In particular, CodY and CsrA, global regulators of gene expression, were associated with the motility loci of the *Eubacterium* and *Roseburia* genomes examined. It was inferred that these regulators may function to couple motility-gene expression with nutrient availability in the environment of the bacterium. This may be of particular relevance to the study of motility in *R. inulinivorans*, a

species in which flagellin expression appears to be carbon-source dependent (Scott *et al.*, 2011). This finding suggests a strong role for carbon sources as chemo-attractants or chemo-repellants in the gut.

While a canonical global regulator of motility-gene expression was not readily identifiable from the *L. ruminis* genome sequences, a model for the regulation of flagellum-biogenesis in this species was proposed in Chapter 2 by merging genomic information with transcriptomic data. This led to the identification of a hypothetical protein, ANHS_518/LRC_15730, as a potential global regulator for the *L. ruminis* flagellum-biogenesis pathway. Furthermore, ANHS_518/LRC_15730 was predicted to be under the transcriptional control of an alternative, sigma factor, $\sigma^{70'}$ which could be an extracytoplasmic function sigma factor. Preliminary, unpublished, follow-up data from our lab suggest that expression of $\sigma^{70'}$ and ANHS_518/LRC_15730 are related to the non-motile phenotype. Thus, ANHS_518/LRC_15730 could serve as a molecular target for any future studies that seek to elucidate the regulatory network controlling flagellum biogenesis in *L. ruminis*.

The reasons for the different motility phenotypes of *L. ruminis* strains of human and animal origin remain undetermined. The environmental triggers of flagellum biogenesis and motility for *L. ruminis* strains isolated from humans are unknown, but could rely on *in vivo* factors that are not well translated to the *in vitro* culture environment. Results presented in Chapter 2 would support this hypothesis because a partial motility phenotype was restored to *L. ruminis* ATCC25644 following passage through the murine GI tract, and motility was not restored to this strain when cultured in media supplemented with alternative carbon and protein sources. The scientific literature suggests that the

Lactobacillus motility phenotype can be "sluggish" and difficult to demonstrate, and that motility can be "lost" by repeated subculture (Kandler & Weiss, 1986; Torriani *et al.*, 1996). This also implies that the motility phenotype can be difficult to observe in a laboratory environment. However, the presence of an intact motility locus in *L. ruminis* ATCC25644 is evidence that the motility of this strain is not lost, but is instead repressed under certain environmental conditions. Relief from this state of repression is likely to involve ANHS_518/LRC_15730 and $\sigma^{70'}$.

As proteins with significant immunomodulatory potential, the innate-immune response to flagellins of commensal bacteria merits investigation, particularly if these bacteria are intended for use in the food or health industries. Although the flagellin proteins of enteric pathogens such as *S. typhimurium* and *Vibrio cholerae* can trigger pro-inflammatory responses in epithelial cells (Gewirtz *et al.*, 2001; Harrison *et al.*, 2008; Tallant *et al.*, 2004), some pathogenic proteobacteria, including *Helicobacter pylori*, *Campylobacter jejuni* and *Bartonella bacilliformis*, have evolved flagellin molecules that evade recognition by TLR5 (Andersen-Nissen *et al.*, 2005). Thus, the possibility exists that commensal species, particularly those species that are autochthonous to the mammalian intestine and which have a long-term relationship with their host, might also produce flagellin proteins that fail to stimulate pro-inflammatory responses. However, as demonstrated in Chapters 2 and 3, the native flagellin proteins from several *Lactobacillus* species and from *E. rectale* and *R. inulinivorans*, stimulated secretion of the pro-inflammatory cytokine IL-8, from human IECs. Chapter 2 also shows that the pro-inflammatory response to flagellate *L. ruminis* ATCC27782 was achieved through stimulation of TLR5.

The flagellin-TLR5 interaction may result in either protective or harmful consequences for the host. Therefore, further research is required to determine the biological outcome and *in vivo* significance of the pro-inflammatory flagellin-induced signalling outlined in Chapters 2 and 3. TLR5 is basolaterally expressed on epithelial cells in the human colon (Gewirtz *et al.*, 2001; Rhee *et al.*, 2005), while the flagellin proteins of commensal bacteria, including those of the *Lactobacillus*, *Eubacterium* and *Roseburia* species studied as part of this thesis are thought to be confined to the intestinal lumen. This physical separation of receptor and ligand in the colon is expected to prevent the flagellins of commensal bacteria from signalling through TLR5 except in cases where the integrity of the epithelial barrier is breached and an inflammatory response would be required to limit systemic infection.

However, recent *in vitro* research suggests that the flagellin-exposed apical surfaces of IECs might internalise flagellin proteins in a TLR5-dependent manner resulting in IL-8 secretion from both the apical and basolateral surfaces (Eaves-Pyles *et al.*, 2011). Moreover, prolonged exposure of IECs to flagellin *in vitro* has been reported to induce a reversible tolerance to the stimulating protein via TLR5 signalling, without compromising immune responsiveness to other stimulants (Sun *et al.*, 2007a). Thus, if TLR5 is expressed on the apical surfaces of IECs in the human small intestine, as it is in mice (Bambou *et al.*, 2004), pro-inflammatory flagellin-TLR5 signalling, such as that described in Chapters 2 and 3 could contribute to homeostatic and protective signalling by commensal bacteria in the gut (Rakoff-Nahoum *et al.*, 2004) and could also establish a state of immune hypo-responsiveness to these flagellate commensals. Therefore, the pro-inflammatory signalling initiated by the commensal flagellins and flagellate

bacteria studied in Chapters 2 and 3 might not be wholly undesirable and might have a protective biological effect on the host *in vivo*. However, reports suggest that titres of antibodies to flagellins from some of the *Eubacterium* and *Roseburia* species studied in Chapter 3 and to other closely related species are higher in individuals with Crohn's disease than in healthy controls or in individuals with ulcerative colitis (Duck *et al.*, 2007; Lodes *et al.*, 2004). This suggests that these molecules contribute, perhaps in a generalised way (Sitaraman *et al.*, 2005), to the elevated immune responses that are associated with Crohn's disease.

Most of the genomes studied as part of this thesis included genes for more than one flagellin protein and these were differentially expressed. In Chapter 2, *L. ruminis* ATCC27782 was found to harbour two almost identical *fliC* genes at its motility locus, but only one of these genes was transcribed at high levels under standard laboratory conditions. The large protein band that was evident on SDS-PAGE gels loaded with *E. rectale* A1-86 and M104/1 flagellin protein preparations could correspond to two or more protein products of similar size and identical amino-terminal sequence. Protein products corresponding to at least three of the six genes encoding flagellins that were annotated in the *R. inulinivorans* A2-194 genome were recovered from cultures of this strain, though it was not always possible to determine which flagellins were represented in each protein band on the basis of SDS-PAGE resolution alone. The overlapping physical and chemical properties of these proteins hindered their separation and limited our ability to deduce which of the genes encoding flagellin were expressed on the basis of protein isolation alone. Gene-specific qRT-PCR could provide the necessary specificity to identify which of

the flagellin-encoding genes are expressed in these strains under a given set of culture conditions. This approach was used to distinguish expression from *L. ruminis* ATCC27782 *fliC1* and *fliC2* in Chapter 2.

The reasons why motile bacteria produce several flagellins are not always clear. Complex flagella could have structural properties that are distinct from those of simple flagella and which would be preferable for motility in a particular environment. For example, the complex flagella of *Agrobacterium* sp. H13-3, (formerly *Rhizobium lupini* H13-3) are more rigid and brittle than the flagella that are composed of a single type of flagellin (Trachtenberg & Hammel, 1992), but these complex flagella are apparently better suited to movement through viscous environments (Gotz *et al.*, 1982). *H. pylori* expresses two flagellin proteins in unequal stoichiometries (Kostrzynska *et al.*, 1991). Its minor flagellin, FlaB, is found at the hook-proximal region of the filament, while FlaA forms the remainder of the structure (Kostrzynska *et al.*, 1991). This spatial distribution of different types of flagellin could have structural or mechanical significance. In *Campylobacter coli*, wild-type bacteria with complex flagella are more motile than mutants bearing flagella that are constructed from one type of flagellin only (Guerry *et al.*, 1991). Thus, the improved functionality of complex flagella may support the expression and maintenance of multiple flagellins in these species. It could be possible that the *Eubacterium* and *Roseburia* species produce complex flagella with particular structural attributes that are more suitable for motility in an *in vivo* environment, such as through mucus. Nevertheless, as described in Chapter 2, expression of a single flagellin protein is sufficient to support flagellum formation and motility in *L. ruminis* and *L. mali*.

Post-translational glycosylation of flagellin proteins is often necessary for the assembly of complex polar, but not for the construction of simple polar or simple peritrichous flagella in bacteria (Logan, 2009). This and other post-translational modifications could contribute to stabilisation of the complex flagellar filament (Taguchi *et al.*, 2008). The occurrence of glycosyltransferase enzymes and rhamnose utilisation and carbohydrate modification enzymes at the motility loci of *L. ruminis* and *L. mali* respectively implies that post-translational modification of the flagellin proteins of these species could be possible. However, the glycosyltransferase gene immediately downstream of *fliC*1 in *L. ruminis* ATCC27782 includes a frameshift, and is presumed non-functional. This gene is absent from the *L. ruminis* ATC25644 motility locus. Furthermore, as described in Chapter 2, the estimated molecular weights of the *L. ruminis* and *L. mali* proteins after electrophoretic separation were consistent with those predicted from their respective genomes, suggesting that these proteins are not heavily modified by post-translational modifications. Similarly, the molecular weights predicted *in silico* for the *E. rectale* and *R. inulinivorans* flagellins examined in Chapter 3 were compatible with the migration of flagellin proteins from these species following electrophoresis, also suggesting that these proteins were not heavily modified.

As was mentioned in Chapter 3, the recent combination of genomics and metagenomics techniques with high-throughput sequencing technologies has enabled us to appreciate the compositional diversity of the GI microbiota of humans and the biological functions encoded therein without recourse to culture-based methods (Claesson *et al.*, 2011; Qin *et al.*, 2010). This is of relevance because the intestinal microbiota is now recognised as a tractable target for

prophylactic and curative medical treatments (Brandt *et al.*, 2012; Hunter *et al.*, 2012; Lawley *et al.*, 2012; Rojas *et al.*, 2012). Knowing the composition of an individual's microbiota and the functions it encodes would better inform interventions that seek to manipulate or modify the composition of the intestinal microbiota, such as dietary supplementation with prebiotics, probiotics or synbiotics (Furrie *et al.*, 2005; Tannock *et al.*, 2000), and faecal microbiota transplantation (Brandt *et al.*, 2012). Interventions such as these typically target genera or species that have a health-promoting phenotype that is presumed beneficial to the host, or which seek to either counterbalance or replace a microbiota of skewed composition (Lawley *et al.*, 2012; Sokol *et al.*, 2008). Moreover, knowing which biological functions are encoded by the gut microbiome could inform both drug-development and prescription of drug regimen. Given that intra-species genotypic and phenotypic diversity has been reported for many intestinal bacteria (Collins *et al.*, 2012; O'Donnell *et al.*, 2011), there is a need to resolve compositional and functional analyses of the microbiota to the level of strain. There is also a need to ensure that the functional capacity of low abundance, subdominant species in the gut is captured so that potentially desirable or undesirable traits are represented.

The relationship between the relative abundance of species and the representation of genes for the flagellum biogenesis pathway of *Eubacterium* and *Roseburia* species in 27 metagenomes generated by the ELDERMET project were investigated in Chapter 3. These species were chosen for investigation, because they had previously been identified as some of the most abundant species in the human gut microbiota (Aminov *et al.*, 2006; Qin *et al.*, 2010), and were expected to be represented in the ELDERMET databases. Flagellum-

biogenesis was selected as the function of interest because the genes responsible for this microbe-host interaction trait were well-defined. Furthermore, previous studies had reported that cell-motility functions were poorly or variably abundant in the gut microbiome (Kurokawa *et al.*, 2007; Turnbaugh *et al.*, 2009). This suggested that the representation of flagellum-biogenesis genes in the databases might also be sufficiently variable to allow a relationship between species abundance and species-specific gene detection to be discerned. This analysis was also expected to highlight if the current sequencing depth and assembly methods for metagenomes yield a resource that adequately represents functions at the level of species or strain. The outcome of the analysis described in Chapter 3 suggested that if microbiota or microbiome composition is intended for research purposes, a greater depth of sequencing is required to capture the biological functions contributed by low abundance or subdominant strains.

In the future, motile commensal bacteria could be developed by the biotechnology industry as vaccine delivery agents with enhanced adjuvancy. The development of several aflagellate commensal lactobacilli for vaccine delivery is already underway (Fredriksen *et al.*, 2012; Grangette *et al.*, 2001; Kajikawa *et al.*, 2012; Mohamadzadeh *et al.*, 2010). The adjuvancy of such vaccines could be improved by engineering the vehicle bacteria to express heterologous proteins from human pathogens on their cell surfaces. For example, *L. casei* and *L. acidophilus* have been engineered to display *Salmonella* flagellin on their cell surface (Kajikawa & Igimi, 2010; Kajikawa *et al.*, 2011; Kajikawa *et al.*, 2012), while *L. plantarum* has been modified to express invasin from *Yersinia pseudotuberculosis* (Fredriksen *et al.*, 2012). If flagellate lactobacilli suitable for use as vaccine delivery agents could be found, specific

genetic manipulations such as these to improve vaccine adjuvancy might be avoided.

As an efficient secretion apparatus, the flagellar T3SS of commensal bacteria could also be exploited for production and secretion of bioactive peptides. The *S. typhimurium* flagellar T3SS has already been used for this purpose (Singer *et al.*, 2012). Furthermore, flagellum display, a technique for the expression of foreign polypeptides or proteins on the surface of the bacterial flagellum, has been used for the development of recombinant vaccines and for protein expression purposes (Majander *et al.*, 2009). Flagellum surface display has several advantages over other surface display methods, including multivalent protein display and the simultaneous display of two different peptides in a single flagellum via fusions to either *fliC* or *fliD* (Majander *et al.*, 2005; Majander *et al.*, 2009). Commensal bacteria could be used as non-pathogenic alternatives for recombinant vaccine production. Furthermore, the machinery for post-translational modification of commensal flagellins could possibly be exploited to produce glycopeptides fused to flagellin for vaccination purposes, a concept that was proposed recently (Logan, 2009).

The flagellate, motile phenotype of commensal bacteria may be predicted from genome sequences but a laboratory-based confirmation of the phenotype is also required, because motility may be repressed in a strain-specific manner under certain environmental conditions. However, the expression or repression of the motility phenotype *in vitro* may not truly represent the phenotype expressed by the bacterium in the host. Although the overall flagellation pattern, type of flagellum and the regulation of its biosynthesis were predicted to vary among the motile commensals studied as part of this thesis, their flagellin

proteins elicited the same pro-inflammatory response, (secretion of IL-8), from IECs *in vitro*. Indeed, it remains to be elucidated if the commensal flagellin proteins elicit pro-inflammatory responses *in vivo*. Genomics facilitates genus-wide assessments of motility genotypes, while metagenomics provides a useful means to evaluate the overall compositional and functional diversity of the intestinal microbiota, but might require a greater depth of sequencing than that currently used by the scientific community in the preparation of metagenome databases if the functional contribution made by a particular species or strain to an ecosystem is to be determined. Finally, flagellate commensal bacteria could prove valuable to the biotechnology industry as vaccine delivery agents with significant natural adjuvancy, and as efficient tools for recombinant protein expression and surface display.

4.2 References

Aminov, R. I., Walker, A. W., Duncan, S. H., Harmsen, H. J., Welling, G. W. & Flint, H. J. (2006). Molecular diversity, cultivation, and improved detection by fluorescent *in situ* hybridization of a dominant group of human gut bacteria related to *Roseburia* spp. or *Eubacterium rectale*. *Appl Environ Microbiol* **72**, 6371-6376.

Andersen-Nissen, E., Smith, K. D., Strobe, K. L., Barrett, S. L., Cookson, B. T., Logan, S. M. & Aderem, A. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc Natl Acad Sci U S A* **102**, 9247-9252.

Bambou, J. C., Giraud, A., Menard, S., Begue, B., Rakotobe, S., Heyman, M., Taddei, F., Cerf-Bensussan, N. & Gaboriau-Routhiau, V. (2004). *In vitro* and *ex vivo* activation of the TLR5 signaling pathway in intestinal epithelial cells by a commensal *Escherichia coli* strain. *J Biol Chem* **279**, 42984 - 42992.

Brandt, L. J., Aroniadis, O. C., Mellow, M., Kanatzar, A., Kelly, C. P., Park, T., Stollman, N., Rohlke, F. & Surawicz, C. (2012). Long-term follow-up of colonoscopic fecal microbiota transplant for recurrent *Clostridium difficile* infection. *Am J Gastroenterol* **107**, 1079-1087.

Claesson, M. J., Cusack, S., O'Sullivan, O. & other authors (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4586-4591.

Collins, J., van Pijkeren, J. P., Svensson, L. & other authors (2012). Fibrinogen-binding and platelet-aggregation activities of a *Lactobacillus salivarius* septicaemia isolate are mediated by a novel fibrinogen-binding protein. *Mol Microbiol* **85**, 862-877.

Duck, L. W., Walter, M. R., Novak, J., Kelly, D., Tomasi, M., Cong, Y. & Elson, C. O. (2007). Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis* **13**, 1191-1201.

Eaves-Pyles, T., Bu, H. F., Tan, X. D., Cong, Y., Patel, J., Davey, R. A. & Strasser, J. E. (2011). Luminal-applied flagellin is internalized by polarized intestinal epithelial cells and elicits immune responses via the TLR5 dependent mechanism. *PLoS One* **6**, e24869.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of lactobacilli and bifidobacteria. *Curr Issues Intest Microbiol* **8**, 44-61.

Fredriksen, L., Kleiveland, C. R., Hult, L. T., Lea, T., Nygaard, C. S., Eijsink, V. G. & Mathiesen, G. (2012). Surface display of N-terminally anchored invasin by *Lactobacillus plantarum* activates NF-κB in monocytes. *Appl Environ Microbiol* **78**, 5864-5871.

Furrie, E., Macfarlane, S., Kennedy, A., Cummings, J. H., Walsh, S. V., O'Neil D, A. & Macfarlane, G. T. (2005). Synbiotic therapy (*Bifidobacterium*

longum/Synergy 1) initiates resolution of inflammation in patients with active ulcerative colitis: a randomised controlled pilot trial. *Gut* **54**, 242-249.

Gewirtz, A. T., Navas, T. A., Lyons, S., Godowski, P. J. & Madara, J. L. (2001). Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* **167**, 1882-1885.

Gotz, R., Limmer, N., Ober, K. & Schmitt, R. (1982). Motility and chemotaxis in two strains of *Rhizobium* with complex flagella. *Microbiology* **128**, 789-798.

Grangette, C., Muller-Alouf, H., Goudercourt, D., Geoffroy, M. C., Turneer, M. & Mercenier, A. (2001). Mucosal immune responses and protection against tetanus toxin after intranasal immunization with recombinant Lactobacillus plantarum. *Infect Immun* **69**, 1547-1553.

Guerry, P., Alm, R. A., Power, M. E., Logan, S. M. & Trust, T. J. (1991). Role of two flagellin genes in *Campylobacter* motility. *J Bacteriol* **173**, 4757-4764.

Harrison, L. M., Rallabhandi, P., Michalski, J., Zhou, X., Steyert, S. R., Vogel, S. N. & Kaper, J. B. (2008). *Vibrio cholerae* flagellins induce Toll-like receptor 5-mediated interleukin-8 production through mitogen-activated protein kinase and NF-kappaB activation. *Infect Immun* **76**, 5524-5534.

Hunter, C., Dimaguila, M. A., Gal, P., Wimmer, J. E., Jr., Ransom, J. L., Carlos, R. Q., Smith, M. & Davanzo, C. C. (2012). Effect of routine probiotic, *Lactobacillus reuteri* DSM 17938, use on rates of necrotizing enterocolitis in neonates with birthweight < 1000 grams: a sequential analysis. *BMC Pediatr* **12**, 142.

Kajikawa, A. & Igimi, S. (2010). Innate and acquired immune responses induced by recombinant *Lactobacillus casei* displaying flagellin-fusion antigen on the cell-surface. *Vaccine* **28**, 3409-3415.

Kajikawa, A., Nordone, S. K., Zhang, L., Stoeker, L. L., LaVoy, A. S., Klaenhammer, T. R. & Dean, G. A. (2011). Dissimilar properties of two recombinant *Lactobacillus acidophilus* strains displaying *Salmonella* FliC with different anchoring motifs. *Appl Environ Microbiol* **77**, 6587-6596.

Kajikawa, A., Zhang, L., Long, J., Nordone, S., Stoeker, L., LaVoy, A., Bumgardner, S., Klaenhammer, T. & Dean, G. (2012). Construction and immunological evaluation of dual cell surface display of HIV-1 gag and *Salmonella enterica* serovar Typhimurium FliC in *Lactobacillus acidophilus* for vaccine delivery. *Clin Vaccine Immunol* **19**, 1374-1381.

Kalmokoff, M. L., Allard, S., Austin, J. W., Whitford, M. F., Hefford, M. A. & Teather, R. M. (2000). Biochemical and genetic characterization of the flagellar filaments from the rumen anaerobe *Butyrivibrio fibrisolvens* OR77. *Anaerobe* **6**, 93-109.

Kandler, O. & Weiss, N. (1986). Regular, nonsporing Gram positive rods, Bergey's manual of sytematic bacteriology. In *Bergey's manual of sytematic bacteriology*, pp. 1208-1234. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore.

Kankainen, M., Paulin, L., Tynkkynen, S. & other authors (2009). Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proc Natl Acad Sci U S A* **106**, 17193-17198.

Koort, J., Vandamme, P., Schillinger, U., Holzapfel, W. & Bjorkroth, J. (2004). *Lactobacillus curvatus* subsp. *melibiosus* is a later synonym of *Lactobacillus sakei* subsp. *carnosus*. *Int J Syst Evol Microbiol* **54**, 1621-1626.

Kostrzynska, M., Betts, J. D., Austin, J. W. & Trust, T. J. (1991). Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *J Bacteriol* **173**, 937-946.

Kurokawa, K., Itoh, T., Kuwahara, T. & other authors (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181.

Lawley, T. D., Clare, S., Walker, A. W. & other authors (2012). Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. *PLoS Pathog* **8**, e1002995.

Lodes, M. J., Cong, Y., Elson, C. O., Mohamath, R., Landers, C. J., Targan, S. R., Fort, M. & Hershberg, R. M. (2004). Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest* **113**, 1296-1306.

Logan, S. M. (2009). Glycosylation of flagellins. In *Pili and flagella, current research and future trends*, pp. 155-172. Edited by K. F. Jarrell. Norfolk, UK: Caister Academic Press.

Majander, K., Korhonen, T. K. & Westerlund-Wikstrom, B. (2005). Simultaneous display of multiple foreign peptides in the FliD capping and FliC filament proteins of the Escherichia coli flagellum. *Appl Environ Microbiol* **71**, 4263-4268.

Majander, K., Anton, L., Kylvaja, R. & Westerlund-Wikstrom, B. (2009). The bacterial flagellum as a surface display and expression tool. In *Pili and flagella*, pp. 191-206. Edited by K. F. Jarrell. Norfolk, UK: Caister Academic Press.

Mohamadzadeh, M., Durmaz, E., Zadeh, M., Pakanati, K. C., Gramarossa, M., Cohran, V. & Klaenhammer, T. R. (2010). Targeted expression of anthrax protective antigen by *Lactobacillus gasseri* as an anthrax vaccine. *Future Microbiol* **5**, 1289-1296.

O'Donnell, M. M., Forde, B. M., Neville, B., Ross, P. R. & O'Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* **10 Suppl 1**, S12.

Qin, J., Li, R., Raes, J. & other authors (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229-241.

Rhee, S. H., Im, E., Riegler, M., Kokkotou, E., O'Brien, M. & Pothoulakis, C. (2005). Pathophysiological role of Toll-like receptor 5 engagement by bacterial flagellin in colonic inflammation. *Proc Natl Acad Sci U S A* **102**, 13610-13615.

Rojas, M. A., Lozano, J. M., Rojas, M. X. & other authors (2012). Prophylactic probiotics to prevent death and nosocomial infection in preterm infants. *Pediatrics* **130**, e1113-1120.

Scott, K. P., Martin, J. C., Chassard, C. & other authors (2011). Substrate-driven gene expression in *Roseburia inulinivorans*: importance of inducible enzymes in the utilization of inulin and starch. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4672-4679.

Sims, I. M., Frese, S. A., Walter, J. & other authors (2011). Structure and functions of exopolysaccharide produced by gut commensal *Lactobacillus reuteri* 100-23. *ISME J*.

Singer, H. M., Erhardt, M., Steiner, A. M., Zhang, M. M., Yoshikami, D., Bulaj, G., Olivera, B. M. & Hughes, K. T. (2012). Selective purification of recombinant neuroactive peptides using the flagellar type III secretion system. *MBio* **3**.

Sitaraman, S. V., Klapproth, J. M., Moore, D. A., 3rd, Landers, C., Targan, S., Williams, I. R. & Gewirtz, A. T. (2005). Elevated flagellin-specific immunoglobulins in Crohn's disease. *Am J Physiol Gastrointest Liver Physiol* **288**, G403-406.

Sokol, H., Pigneur, B., Watterlot, L. & other authors (2008). *Faecalibacterium parusnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731-16736.

Sun, J., Fegan, P. E., Desai, A. S., Madara, J. L. & Hobert, M. E. (2007). Flagellin-induced tolerance of the Toll-like receptor 5 signaling pathway in polarized intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* **292**, G767-778.

Taguchi, F., Shibata, S., Suzuki, T., Ogawa, Y., Aizawa, S., Takeuchi, K. & Ichinose, Y. (2008). Effects of glycosylation on swimming ability and flagellar polymorphic transformation in *Pseudomonas syringae* pv. tabaci 6605. *J Bacteriol* **190**, 764-768.

Tallant, T., Deb, A., Kar, N., Lupica, J., de Veer, M. & DiDonato, J. (2004). Flagellin acting via TLR5 is the major activator of key signaling pathways leading to NF-kappaB and proinflammatory gene program activation in intestinal epithelial cells. *BMC Microbiol* **4**, 33.

Tannock, G. W., Munro, K., Harmsen, H. J., Welling, G. W., Smart, J. & Gopal, P. K. (2000). Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20. *Appl Environ Microbiol* **66**, 2578-2588.

Torriani, S., Van Reenen, G. A., Klein, G., Reuter, G., Dellaglio, F. & Dicks, L. M. (1996). *Lactobacillus curvatus* subsp. *curvatus* subsp. nov. and *Lactobacillus curvatus* subsp. *melibiosus* subsp. nov. and *Lactobacillus sake* subsp. *sake* subsp. nov. and *Lactobacillus sake* subsp. *carnosus* subsp. nov., new subspecies of *Lactobacillus curvatus* Abo-Elnaga and Kandler 1965 and *Lactobacillus sake* Katagiri, Kitahara, and Fukami 1934 (Klein et al. 1996, emended descriptions), respectively. *Int J Syst Bacteriol* **46**, 1158-1163.

Trachtenberg, S. & Hammel, I. (1992). The rigidity of bacterial flagellar filaments and its relation to filament polymorphism. *J Struct Biol* **109**, 18-27.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T. & other authors (2009). A core gut microbiome in obese and lean twins. *Nature* **457**, 480-484.

van Pijkeren, J. P., Canchaya, C., Ryan, K. A. & other authors (2006). Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* **72**, 4143-4153.

Acknowledgements

I would like to thank the many people who directly and indirectly contributed to my PhD research experience and made this thesis possible.

Sincere and heartfelt thanks to Prof. Paul O'Toole. Thank you for your encouragement over the last few years, for listening to my ideas and for imparting the wisdom of your research experiences. It has been a pleasure to work in your lab.

To Prof. Paul Ross, thank you for your advice, suggestions and feedback on my work, and for always extending a warm welcome whenever I visited Moorepark.

To the local and international collaborators and the individuals who contributed to the research included in this thesis, thank you! The assistance and suggestions you provided greatly enhanced the quality of each study.

To all the scientists who joined FSB Lab 339 and BSI office 4.11 since October 2007 - you are too many to name individually – but it was wonderful to work in the dynamic environment that we all created, and I learned a lot from working with you! Thank you to the past and present members of FSB Labs 335, 337, 340 and BSI Lab 4.25 for advice, assistance and good company. I wish all of you continued success in your respective careers, wherever they may take you!

To the past and present members of the U. C. C. Microbiology department who maintained the communal equipment, who solved computer problems, who delegated the demonstrating, who handled all the orders and who kept the autoclave trolleys from becoming a health hazard – thank you for keeping everything running smoothly! We all benefited from the efficient environment that you helped to create.

To my friends, Sinéad, Aisling and Lucy. Thank you for taking time out to share research and non-research experiences with me.

Finally, a few words of thanks to my family... To my parents, Mary and Miah, thank you for your unfaltering belief in my ability to make it this far, and for your encouragement throughout my many student years. It would not have been possible without your support. To Fiona, John, my grandmother and my extended family - thank you for your interest in my work, and for reminding me that there is always life beyond the lab! To Gavin, thank you for your unlimited patience, encouragement and advice, and for helping me to keep everything in perspective when the going got tough. Without all of you, this would have been so much harder.