# Characterization of commensal *Escherichia coli* through genetic content and Flagellin structure

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Thomas Hagemann Gutter

aus Santiago de Chile / Chile

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 05.07.2023 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter/-in: | Prof. Dr. med. Julia-Stefanie Frick |
| 2. Berichterstatter/-in: | Prof. Dr. Karl Forchhammer |

*For my parents (Para mis Padres)*

# Contents

# Abbreviations

| | |
|---|---|
| **5-ASA** | 5-aminosalicylic acid |
| **AIEC** | Adherent-invasive E. coli |
| **AP-1** | activator protein 1 |
| **BLAST** | Basic local alignment search tool |
| **Bp** | Base pair |
| **Breg** | Regulatory B cell |
| **CD** | Crohn’s disease |
| **CED** | Chronisch-entzündliche Darmerkrankungen |
| **CPI** | Cocktail of protease inhibitor |
| **CRP** | C-reactive protein |
| **CsA** | Cyclosporin A |
| **CV** | Column volumn |
| **DAEC** | Diffusely adherent E. coli |
| **DAP** | 2,6-Diaminopimelic acid |
| **DC** | Dendritic cell |
| **DM** | Darmmikrobiota |
| **DMEM** | Dulbecco's Modified Eagle Medium |
| **DMSO** | Dimethyl sulfoxide |
| **DNA** | Deoxyribonucleic acid |
| **DOC** | 2,5-Dimethoxy-4-chloroamphetamine |
| **DSS** | Dextran sodium sulfate |
| **EC** | Epithelial cells |
| **EcN** | Escherichia coli strain Nissle 1917 |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **EHEC** | Enterohemorrhagic E. coli |
| **ELISA** | Enzyme linked immunosorbent assay |
| **EM** | Electron microspcopy |
| **EPEC** | Enteropathogenic E. coli |
| **EU** | Endotoxin units |
| **FC** | Faecal calprotectin |
| **FCS** | Fecal calf serum |
| **FEP** | Flagella enriched preparations |

| | |
|---|---|
| **FK-AT** | Nuclear factor activating T cell |
| **FIiC** | Flagellin |
| **FPLC** | Fast protein liquid chromatography |
| **GI** | Gastrointestinal |
| **GM** | Gut microbiota |
| **HCS** | Histological colitis score |
| **HEK** | Human embrionik kidney |
| **HVR** | Hypervariable region |
| **IBD** | Inflammatory bowel disease |
| **IFIB** | Interfakultäre Institut für Biochemie |
| **IL** | Interleukin |
| **ILC** | Innate lymphoid cells |
| **IMAC** | immobilized metal affinity chromatography |
| **IMIT** | Interfakultäres Institut für Mikrobiologie und Infektionsmedizin Tübingen |
| **IPTG** | Isopropyl-β-D-thiogalactopyranoside |
| **IS** | Insertion sequences |
| **LAL** | Limulus amebocyte lysate |
| **LB** | Luria-Bertani |
| **LG** | logistic regression |
| **LPS** | lipopolysaccharide |
| **M P/A** | Presence/absence matrix |
| **MALDI-TOF** | Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. |
| **MAMP** | Microbe-associated molecular pattern |
| **MAPK** | Mitogen-activated protein kinases |
| **MS** | Mass spectrometry |
| **MSA** | Multi-sequence alignment |
| **MUC2** | Glycosylated protein mucin 2 |
| **MUSCLE** | MUltiple Sequence Comparison by Log-Expectation |
| **NF-ƙB** | Nuclear factor kappa B |
| **NGS** | Next generation sequencing |
| **NK** | Natural killer cell |
| **NLR** | Nod like receptor |
| **PAGE** | Polyacrylamide gel electrophoresis |
| **PBS** | Phosphate buffered saline |

| | |
|---|---|
| **PCR** | Polymerase chain reaction |
| **PNPP** | P-nitrophenyl phosphate |
| **PPAR-y** | Peroxisome proliferator activated receptor gamma |
| **PRR** | Pattern recognition receptors |
| **PSA** | Polysaccharide A |
| **RNA** | Ribonucleic acid |
| **RPM** | Revolutions per minute |
| **SCFA** | Short chain fatty acids |
| **SCID** | Severe combined immunodeficiency disease |
| **SDI** | Shannon diversity index |
| **SDS** | Sodium dodecyl sulfate |
| **SEC** | Size exclusion chromatography |
| **SFB** | Segmented filamentous bacteria |
| **SM** | Salmonella Minessota |
| **SPF** | Specific pathogen free |
| **ST** | Salmonella Typhimurium |
| **STEC** | Shiga toxin-producing E.coli |
| **TEV** | Tobacco Etch Virus |
| **Tfh** | Follicular helper T cell |
| **TIR** | Toll/interleukin-1 receptor |
| **TJ** | Tight junction |
| **TLR** | Toll-like receptor |
| **TNF** | Tumor necrosis factor |
| **Treg** | Regulatory T cell |
| **UC** | Ulcerative colitis |
| **UPEC** | Uropathogenic E. coli |
| **UV** | Ultra violet |
| **WGS** | Whole-genome sequencing |
| **WT** | Wild type |

# Summary

Inflammatory bowel disease (IBD) is a steady health concern in the developed world and an increasing problem in emergent economies. IBD manifests as inflammation in the intestinal tract, impacting significantly the quality of life of the patients coping with the disease and in turn representing a major burden to the healthcare system.

The precise IBD etiology is unknown. Nevertheless, factors such as genetic conditions of the individual, environmental elements and the gut microbiota composition (GM) are among the most relevant elements associated with IBD. The GM comprises the whole collection of microbes inhabiting the gastrointestinal tract of an individual. Functional alteration of the GM composition is termed dysbiosis. In this state, an increase in proteobacteria and a general reduction in diversity are strongly associated with IBD.

Diagnosis of IBD is mainly accomplished through endoscopic exploration, and treatment consists of immunosuppressive drugs aiming to reduce inflammation. The drawbacks are that endoscopy lacks predictive value and the therapeutic alternatives bring along considerable side effects for the patient.

In recent years the detection of biomarkers in patients´ specimens has been used as complementary diagnostic approach. Nevertheless, these biomarkers still lack in specificity, accuracy and sensitivity. Hence, there is an ongoing interest to identify additional biomarkers to predict inflammation during IBD.

For this, one of the research fronts aims at the genetic characterization of bacteria associated with a dysbiotic state or pathobionts. Up to date, characterization of pathobionts such as *Escherichia coli* has not produced positive results.

The advances in high throughput sequencing have produced a gigantic amount of data about the GM composition, but at the same time they have brought the challenge of analysis of big data sets. In this work we show that it is possible to tackle this problem with affordable computational resources and familiar statistical tools. We found four genetic markers (*IS3 family transposase, trbC, eheA* and *licT*) that correctly classified 83% of the *E. coli* isolates as IBD-specific.

In a second project we analysed the probiotic effects of the surface protein flagellin (FliC) from the symbiotic bacterium *E. coli* Nissle 1917 (EcN). We showed that FliC from EcN (EcN FliC) is able to confer protection to mice against dextrane sodium sulfate (DSS)-induced acute colitis and that this protection relies on the structure of EcN FliC.

# Zusammenfassung

Chronisch-entzündliche Darmerkrankungen (CED) sind in den Industrieländern ein ständiges Gesundheitsproblem und in Schwellenländern ein zunehmendes. CED manifestiert sich als Entzündung im Darmtrakt, die die Lebensqualität der Patienten, die mit der Krankheit leben müssen, erheblich beeinträchtigt und wiederum eine große Belastung für das Gesundheitssystem darstellt.

Die genaue CED-Ätiologie ist unbekannt. Dennoch gehören Faktoren wie genetische Veranlagung der Person, Umweltelemente und die Zusammensetzung der Darmmikrobiota (DM) zu den relevantesten Elementen im Zusammenhang mit CED. Die DM umfasst die gesamte Sammlung von Mikroben, die den Gastrointestinaltrakt eines Individuums bewohnen. Funktionelle Veränderungen der DM-Zusammensetzung werden als Dysbiose bezeichnet. In diesem Zustand ist die Zunahme von Proteobakterien und die Verringerung der Diversität stark mit CED assoziiert.

Die aktuelle CED-Diagnostik wird hauptsächlich durch Endoskopie durchgeführt. Die Behandlung von CED-Patienten besteht aus immunsuppressiven Medikamenten, die darauf abzielen Entzündungen zu reduzieren. Die Nachteile sind, dass die Endoskopie keine Aussagekraft für zukünftige Entzündungen hat und die bisherigen Therapien erhebliche Nebenwirkungen für den Patienten mit sich bringen.

Als komplementäre diagnostische Option entwickelte sich in den letzten Jahren die Einbeziehung von Biomarkern, denen jedoch immer noch an Spezifität, Genauigkeit und Sensitivität mangelte. Daher besteht ein anhaltendes Interesse an der Einbeziehung von neuen Biomarkern zur Vorhersage von entzündlichen Schüben bei CED.

Eine der Ansätze zur Lösung dieses Problems befasst sich mit der genetischen Charakterisierung von Bakterien, die mit einem dysbiotischen Zustand oder Pathobionten assoziiert sind. Bis heute hat die Charakterisierung von Pathobionten wie *Escherichia coli* keine förderlichen Ergebnisse erbracht.

Die Einbeziehung der Hochdurchsatz-Sequenzierung hat eine gigantische Menge an Daten über die DM-Zusammensetzung produziert, aber gleichzeitig die Herausforderung der Analyse dieser großen Datensätze mit sich gebracht. In dieser Arbeit haben wir gezeigt, dass es möglich ist, dieses Problem mit erschwinglichen Rechenressourcen und vertrauten statistischen Werkzeugen anzugehen. Wir fanden

vier genetische Marker (Transposase der *IS3-Familie, trbC, eheA und licT*), die 83 % der von CED-Patienten isolierten *E. coli* korrekt als solche und nicht als Isolate aus gesunden Spendern klassifizierten.

In einem zweiten Projekt analysierten wir das Oberflächenprotein Flagellin (FliC) aus dem symbiotischen Bakterium *E. coli* Nissle 1917 (EcN). Wir haben gezeigt, dass FliC von EcN (EcN FliC) Mäuse vor einer Dextrannatriumsulfat (DSS)- induzierten akuten Colitis schützen kann und dass dieser Schutz auf der Struktur von EcN FliC beruht.

# 1 Introduction

## 1.1 Inflammatory bowel diseases

Inflammatory bowel diseases (IBD) like Crohn's disease (CD) or ulcerative colitis (UC) are chronic inflammatory disorders of the gastrointestinal (GI) tract.

IBD presents both cycles of inflammation with strong clinical symptoms and phases of remission. The diseases differ in terms of location of inflammation and mucosal damage. CD is characterized by chronic inflammation of any part of the GI tract and has a progressive and destructive course [2]. Clinical presentation includes abdominal pain, diarrhoea, fatigue, weight loss, fever, growth failure, anaemia, recurrent fistulas as well as extraintestinal manifestations (EIMs). UC presents with diverse phenotypes, but it is observed predominantly as a superficial inflammation of the rectum with increased risk of rectal bleeding compared to CD. Both diseases can be associated with EIMs, like pyoderma gangrenosum, uveitis, spondyloarthropathy - axial arthropathy - and primary sclerosing cholangitis [3]. EIMs are more frequent in patients with CD than in patients with UC [4]. In addition, UC and CD are associated with an enhanced risk to develop colorectal cancer [5, 6].

Since the 1950s, an increase of autoimmune diseases including IBD has mainly been observed in developed countries as well as emerging economies. Today, the global distribution of IBD shows a high prevalence in Europe, with e.g., 505/100 000 ulcerative colitis cases in Norway and 322/100 000 Crohn's disease cases in Germany, as well as in North America, with 286 /100 000 cases of ulcerative colitis in the USA and 319/100 000 Crohn's disease cases in Canada [7].

The precise incidence in developing economies remains unclear but it is suggested that the acquisition of the so-called western diet, which is associated to the increase of wealth, is linked to an increase in IBD [8].

This idea is also reflected in the so called "hygiene hypothesis" which proposes an association between higher hygiene standards and a higher occurrence of immune-mediated diseases and atopic diseases. This concept was further extended into the microflora hypothesis, which postulates that disturbances in the microbiota composition in the gut early in childhood may cause immune imbalance and lead to IBD later on [9].

It is known that a genetic predisposition of the individual is crucial for induction of disease. Genome wide association studies revealed 163 gene loci associated with

IBD. 30 gene loci were specifically associated with CD, and 23 with UC [10]. An overview of such loci is given by Mohan *et al.* and provided here in **Table 1**.

Table 1. Genes and pathways associated with CD and UC (Adapted from Mohan and colleagues [11])

| Biological Pathway | Genes |
|---|---|
| inflammatory response | *IL21, IL1RN, CXCL8, IL25, ITGB2, NLRC4, NOD2, TNF, ICAM1, CXCR1, CXCR3, ALOX5, SPP1, CCL2, NLRP3, CCR6, CCR5, ELANE, IL10, SELE, F3, TNFRSF1A, CXCL10, IL2RA, LTA, TLR4, IL17A* |
| response to molecule of bacterial origin | *IL10, CXCL8, FASLG, PTPN22, NOD2, MPO, SELE, TNF, ICAM1, TNFRSF1A, CXCL10, IFNG, LTA, CCL2, NLRP3, CCR5, TLR4, ELANE* |
| response to lipopolysaccharide | *IL10, CXCL8, FASLG, PTPN22, MPO, SELE, TNF, ICAM1, TNFRSF1A, CXCL10, IFNG, LTA, CCL2, NLRP3, CCR5, TLR4, ELANE* |
| response to cytokine | *IL1RN, CXCL8, ITGA4, IL1R1, FASLG, SELE, F3, TNF, ICAM1, TNFRSF1A, CXCL10, IFNG, CXCR1, IL2RA, CXCR3, FASN, IL2RB, LTA, CCL2, CCR6, CCR5, IL17A* |
| response to other organism | *IL10, CXCL8, IL25, FASLG, PTPN22, NLRC4, NOD2, MPO, SELE, TNF, ICAM1, TNFRSF1A, CXCL10, IFNG, IL2RA, LTA, CCL2, NLRP3, CCR5, TLR4, ELANE, IL17A* |
| leukocyte migration | *IL10, CXCL8, ITGA4, IL1R1, ITGB2, SELE, TNF, MMP9, ICAM1, CXCL10, IFNG, CXCR1, CXCR3, CCL2, CCR6, CCR5, ELANE* |
| regulation of defense response | *IL10, IL21, IL20, IL1R1, ITGB2, PTPN22, NLRC4, NOD2, SELE, TNF, TNFRSF1A, IFNG, IL2RA, LTA, CCL2, NLRP3, TLR4, ELANE, IL17A* |
| regulation of inflammatory response | *IL10, IL21, IL20, IL1R1, NOD2, SELE, TNF, TNFRSF1A, IL2RA, LTA, CCL2, NLRP3, TLR4, ELANE, IL17A* |
| response to lipid | *IL10, IL1RN, CXCL8, FASLG, PTPN22, MPO, SELE, TNF, AREG, ICAM1, TNFRSF1A, CXCL10, IFNG, SPP1, LTA, CCL2, NLRP3, CCR5, TLR4, ELANE* |
| cell surface receptor signalling pathway | *IL1RN, CXCL8, IL25, ITGB2, PTPN22, FASLG, NOD2, TNF, AREG, ICAM1, CXCR1, CXCR3, CCL2, CTLA4, CCR6, CCR5, ITGA4, IL1R1, GPBAR1, MMP9, F3, TNFRSF1A, CXCL10, IFNG, IL2RA, IL2RB, LTA, SAG, TLR4, IL17A* |
| leukocyte cell-cell adhesion | *IL10, IL21, ITGA4, ITGB2, PTPN22, NOD2, SELE, TNF, ICAM1, IFNG, IL2RA, CCL2, CTLA4, NLRP3, CCR6, ELANE* |

Main players in the induction of autoimmunity are polymorphisms in genes encoding for regulation of adaptive immunity (IL10, IL23R, IL21, IL17A). In addition, polymorphisms in genes encoding for microbial recognition through pattern recognition receptors (PRRs) like, e.g., NOD1 and NOD2, TLR4 and TLR5, are highly associated

with IBD as well as polymorphisms in genes associated with the inflammasome activity (NLRP3, NLRC4)[12].

### 1.1.1 The intestinal microbiota

The prominent role of PRR polymorphisms in IBD pinpoints the third crucial player in the induction of these diseases: the gastrointestinal microbiota (GM), which is defined as "*the collection of bacteria, archaea and eukarya colonizing the GI tract*" [13]. The mucosal surface of the GI tract is colonized by at least 160 bacterial species [14] for a total of $10^{12}$ single bacteria actively interacting with each other and contributing to the homeostasis of the human host.

The microbiota mediates colonization resistance, hence protecting the host from infection and overgrowth of pathogenic bacteria, e.g., *Clostridiodes difficile,* or fungi like *Candida spp*. In case of *C. difficile,* it was shown that a reduced diversity in the GI microbiota enhances the risk of infection, and the restoration of diversity by faecal microbiota transplantation efficiently outcompeted *C. difficile* and led to clearance of infection [15]. Faecal microbiota transfer turned out to be more efficient in the treatment of recurring *C. difficile* infections as compared to antibiotic therapy [16].

The microbiota supports the host also by production of compounds that are essentially needed by the host. Members of the human microbiota are known to synthesize vitamin K as well as most of the water-soluble B vitamins, such as biotin, cobalamin, folates, nicotinic acid, pantothenic acid, pyridoxine, riboflavin and thiamine [17]. Other important compounds synthesized by the intestinal microbiota are short chain fatty acids (SCFA). Besides exerting various metabolic effects, SCFA contribute to epithelial homeostasis and to regulation of the immune system. The immunological effects of SCFA are, at least in part, mediated by the downregulation of the transcription of proinflammatory cytokines and the differentiation of lamina propria regulatory T (Treg) cells [18].

Hence, the intestinal microbiota strongly contributes to the intestinal as well as systemic immune homeostasis of the host. Consequently, different autoimmune diseases are associated with alterations in the composition of the intestinal microbiota that generally lead to a reduction in diversity [19].

The interaction between the host intestinal immune system and the intestinal microbiota is complex, dynamic and context dependent [20].

A mucus layer and a single epithelial layer separate the deeper mucosal tissue and the associated immune cells from the luminal content like the microbiota or food antigens [21]. The main component of the mucus layer is the glycosylated protein MUC2 (mucin 2), which was shown to additionally dampen the immunogenicity of pathobionts and pathogens by imprinting dendritic cells (DC) towards an increased secretion of the anti-inflammatory cytokine IL-10 [22, 23]. In addition, the intestinal microbiota is known to contribute to the integrity of the epithelial barrier by regulation of the expression of tight junction (TJ) and associated cytoskeletal proteins [24].

With respect to immune cells, the intestinal microbiota influences various cell types of the innate and adaptive immune system. One of the most important mechanisms is the interaction with DCs which contribute to the host immunity by sampling and presenting antigens and subsequently promoting different T cell responses, i.e., pro- or anti-inflammatory [25]. Another cell type of interest is the innate lymphoid cells (ILC). These cells represent a heterogeneous group of immune cells characterized by rapid cytokine and chemokine production. They promote anti-infectious responses and contribute to wound healing [26].

In adaptive immunity, the microbiota interacts with B cells, inducing their maturation into regulatory B (Breg) cells, which are essential players in the maintenance of intestinal homoeostasis [27] (**Figure 1**).

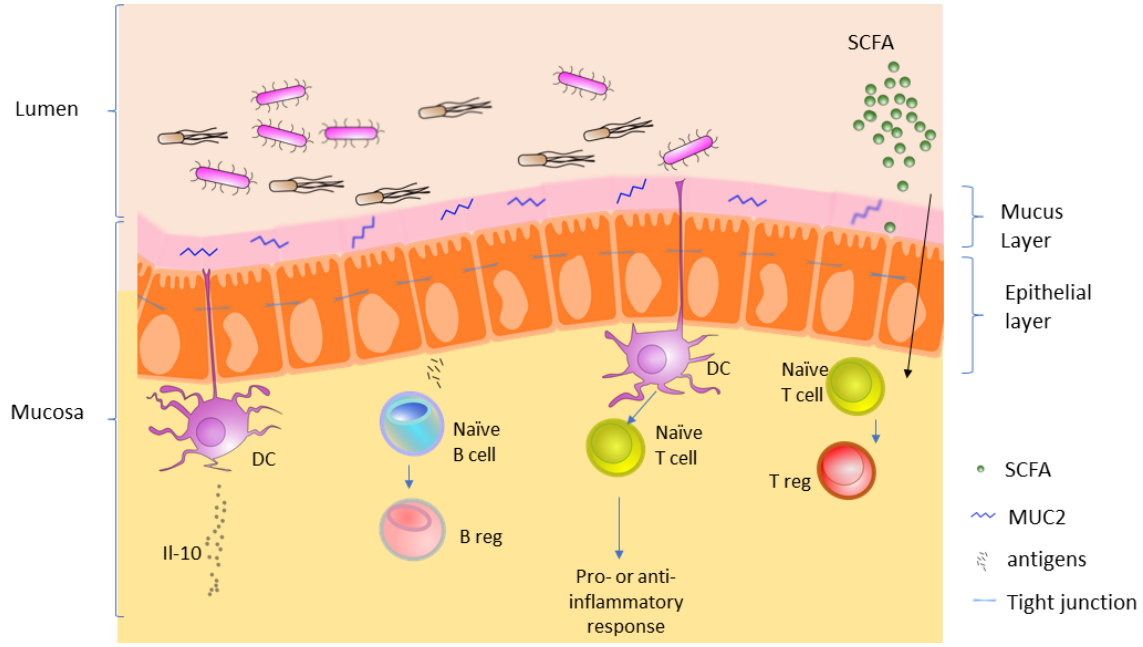



Figure 1. Scheme of the interaction of the gut microbiota and gut immune system. SCFA = short chain fatty acids, DC = dendritic cell, B reg = regulatory B cell, T reg = regulatory T cell, MUC2 = glycosylated protein mucin 2.

Concerning $CD4^+$ T cells, the intestinal microbiota influences both the proliferation of $CD4^+$ T cells as well as the polarization into the different T helper cell subsets, e.g., Th1, Th17 and into Tregs. The induction and activity of cytotoxic $CD8^+$ T cells is also influenced by the microbiota [28]. Recently, it was described that the microbiota interacts with follicular helper T cells (Tfh) [29]. Tfh are crucial for the formation of germinal centers of lymph nodes, they promote the generation of high affinity antibody responses and differentiation of memory B cells.

### 1.1.2 Pathobionts, Symbionts and Probiotics

The degree of complexity existing in the host-microbiota interaction demands new definitions for the participants in such interaction or at least a re-examination of those classic definitions. For the purpose of this manuscript we propose that commensal bacteria can be broadly classified as pathobionts, symbionts and probiotics depending on the outcome of the host-bacteria interaction (**Figure 2**)

Pathobionts are commensals that are harmless in a healthy host but can promote disease in a host with a specific genetic predisposition or under particular environmental conditions [30]. Symbionts are bacteria that establish a mutualistic relationship with the host, hence benefiting themselves and the host from this interaction [31]. Again, the symbiotic features of commensals may depend on host genetic and environmental factors [1].

Commercially available probiotics are defined as bacteria whose association in adequate doses with humans or animals can have positive health effects on the host [32]. As an example, the restoration of microbe diversity when the composition of the gut microbiota has been altered can be achieved through the intake of probiotics [33].

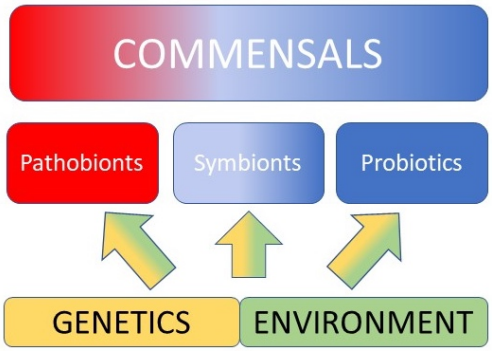



Figure 2. Schematic view of the proposed concept of pathobiont symbiotic and probiotic commensal bacteria.

### 1.1.3 Effects of pathobionts and symbionts on host physiology

A reduced diversity of the intestinal microbiota has been associated with an inflammatory state of the host GI tract [34]. Such a reduction can be caused by overgrowth of one species, e.g., *Escherichia coli.* Furthermore, the presence of specific pathobionts such as diffusely adherent *E. coli* (DAEC), the adherent-invasive *E. coli* (AIEC) and in mice *E. coli* mpk are also associated with IBD [35, 36].

The pathobiont *E. coli* mpk was shown to promote IBD in both *IL-2* deficient mice as well as *Rag1*$^{-/-}$ mice, and to induce a strong Th1/17 response [36, 37]. Neither in *Il2*$^{-/-}$ mice nor in *Rag1*$^{-/-}$ mice, the host can counteract this response by Tregs, resulting in overwhelming intestinal inflammation [36-38].

In the murine gut microbiota, *Helicobacter hepaticus* was described as the first pathobiont bacterium. *H. hepaticus* is causally linked to inflammation of the colon which ensues in SCID mice upon adoptive transfer of naïve CD4$^{+}$ T cells but not in wild type mice [30]. A similar effect was observed with segmented filamentous bacteria (SFB), other well studied pathobionts. SFB have been found to induce specific Th17 cells in the small intestinal lamina propria [39], which are associated with a variety of autoimmune inflammatory diseases.

In the last years, *Bacteroides vulgatus* mpk was shown to mediate anti-inflammatory effects on the host [36] . Hence, we consider this strain as a symbiont.

We were able to show that intestinal association with *B. vulgatus* mpk prevented the induction of IBD in *Il-2*$^{-/-}$ mice as well as *Rag1*$^{-/-}$ mice [36]. In T cell-transplanted *Rag1*$^{-/-}$ mice, administration of *B. vulgatus* mpk even abrogated the existing intestinal inflammation [40]. The protection against inflammation or respectively the abrogation of inflammation is associated with a decrease in the Th1/Th17 response of the host as well as with the induction of Treg cells [36, 38].

Another example of symbiotic bacterium is *Bacteroides fragilis*. It was shown that the expression of polysaccharide A (PSA) by *B. fragilis* prevented induction of colitis *in Rag1*$^{-/-}$ mice upon transfer of naïve CD4$^{+}$ cells [41]. It also prevented sterile and viral central nervous system inflammatory disorders [42]. In addition, it was shown that only a low percentage of *B. fragilis* strains harboured in IBD patients expresses PSA compared to *B. fragilis* strains in a healthy population [43].

### 1.1.4 Role of microbe-associated molecular patterns in symbiotic host-microbe interactions

Microbial recognition is mediated by PRRs like Toll-like receptors (TLRs) or Nod like receptors (NLRs) on epithelial cells, phagocytes and macrophages among others [20, 44]. The activation of these receptors is triggered by conserved, microbe-specific molecules, known as microbe-associated molecular patterns (MAMPs), such as e.g., lipopolysaccharide (LPS) or flagellin (FliC), which play an important role in inducing host intestinal immune reactions.

For example, *E. coli* mpk or its LPS, the ligand of the toll-like receptor 4 (TLR4), induces strong intestinal inflammation when introduced into a host with a T cell dysfunction (*Il-2*$^{-/-}$, *Rag1*$^{-/-}$) but it exerts a protective effect against dextran sodium sulfate (DSS)-induced intestinal inflammation in an immunocompetent host by induction of B reg cells and T reg cells [36, 37].

In contrast, *B. vulgatus* mpk or its LPS effectively suppresses inflammation in hosts with a dysfunctional T and B cell response but is not able to protect the immunocompetent C57BL/6 mouse against DSS induced colitis [45].

However, in the individual host, the structure of the MAMP appears to dictate either symbiotic or pathobiont features as shown for flagellin, the ligand of TLR5. The FliC of the *Escherichia coli* strain Nissle 1917 (EcN) protected mice against DSS-induced colitis, but not the FliC of *E. coli* MG1655. In addition, deletions inside *fliC*, the gene that encodes FliC in EcN resulted in loss of anti-inflammatory properties [1].

### 1.1.5 Current IBD diagnostics and therapeutics

IBD diagnosis is based on the detection of intestinal mucosal damage and inflammation. Even though there is no universal diagnostic standard, the most recent European guidelines include clinical evaluation, endoscopy, cross-sectional imaging, histological observation, stool analysis and biochemical tools [46]. Clinical evaluation, endoscopic and radiological findings are widely accepted tools to determine the severity of the disease. As objective measure of the observed mucosal damage, several different disease activity indexes have been created, each one based on different algorithms [47]. Histological observation has increased the sensibility of mucosal damage detection, nevertheless it is not included in disease activity index determination [48]. Stool analysis includes microbiological exploration and several

markers for mucosal damage [49, 50]. Beside faecal markers, also numerous serum markers are taken into account [51].

IBD therapy is based on immunosuppression by classical immunosuppressive substances like corticosteroids. Another leading substance in maintenance of remission is mesalazine (5-aminosalicylic acid or 5-ASA), an anti-inflammatory compound that acts in the colon. Its mechanism of action is not finally described yet, but one plausible target of 5-ASA is the peroxisome proliferator activated receptor gamma (PPAR-gamma). PPAR-y trans-represses the nuclear factor kappa B (NF-ƙB), a central signal transducer and transcriptional activator. So-called biologicals are helpful in more severe courses of disease. Currently, anti TNF-α antibodies (e.g., Infliximab) or antibodies blocking α(4)β(7) integrins on activated T cells (Vedolizumab) are used. Thiopurines (Azathioprin, 6-Mercaptopurin und 6-Thioguanin) are used as cytostatic and immunosuppressants in CD.

In a very severe disease, calcineurin inhibitors drugs such as cyclosporin A (CsA) and tacrolimus (FK-506) are used. CsA and FK-506 differ in their molecular structure and dosage, but they exert their immunosuppressive activities by inhibiting the activation of T cells through similar pathways [52]. CsA binds to the immunophilin cyclophilin A and this complex binds to and inhibits calcineurin. As a result, the nuclear factor activating T cell (NF-AT) cannot be dephosphorylated (activated) by calcineurin and translocate into the nucleus, leading to reduced T cell effector function.

Despite the advances in diagnostic and treatment, there are major challenges still to be addressed. Endoscopy and radiological approaches are invasive, costly and cannot predict an upcoming inflammation. Serum markers inform of systemic inflammation and the current gold standards therapies act systemically as well. This implies that the diagnosis is not specific and the treatment generates extensive side effects such as systemic immunosuppression, carrying with it susceptibility to infection, tumor growth, etc. as possible undesirable side effects.

## 1.2 Analysis of the *E. coli* pangenome

It has been shown that an early IBD diagnosis and intervention would substantially improve the patient's outcome [53]. For this reason, there is an ongoing interest in incorporating novel markers in IBD diagnostics, such as microbiological genetic traits that early fingerprint the host inflammatory condition.

Due to the fact that one of the hallmark changes in GM composition associated with IBD is the overgrowth of *E. coli* [54, 55], there has been extensive analysis of the whole genome content of *E. coli* strains inhabiting an inflamed niche, nevertheless no single genetic element has been found that links *E. coli* pathobionts to the onset of IBD [55]. One example is the pathobiont AIEC, which was suspected to promote induction of inflammation in a genetically predisposed host and found to be associated specifically with the ileal mucosa in patients suffering from CD [55]. AIEC strains present different features: they adhere to intestinal epithelial cells, can invade the epithelium, survive and replicate in macrophages but no unique genetic markers have been associated to this strain [56].

Next generation sequencing together with staggering advances in bioinformatics has been key in describing the *E. coli* genomic landscape. In such a way it has been possible to analyse 10 000 genomes from *E. coli* coming from human host, describing a total pangenome of 55 000 genes [57]. It has been estimated that the core genome of the whole species comprises 2 200 genes [58], which highlights the genetic richness of the whole species.

The GM composition varies through an individual life and between individuals, in good and poor health, making it extremely difficult to accurately correlate either condition [59]. The complex nature of the GM together with an abundant *E. coli* accessory genome makes the search of pathobiont markers for IBD highly difficult.

## 1.3 The MAMP flagellin.

FliC, the ligand of TLR5, is the main component of the bacterial flagellum which is a key microbial organelle. The function of flagella goes beyond propelling the bacterial cell and participating in adherence; the flagella and especially FliC contribute to immune modulation and can stimulate the host immune system in a pro or anti-inflammatory manner [60, 61]. As a whole, the flagellum can be considered as a multi protein complex whose assembly relies on more than 40 genes [62]. The flagellum is composed of 20 000 to 30 000 helically arranged FliC units [63] forming a rigid flagellar filament and the hook. This hook is a short flexible segment that connects the flagellar filament to the rotary motor [64] hence enabling the rotary motor to transmit turning moments to the filament and thereby allowing locomotion of the cell.

Conventionally, FliC is organized in the four morphological domains D0, D1, D2 and D3 (N'terminus-D0-D1-D2-D3-D2-D1-D0-C'terminus) [65], which are arranged from

the inside to the outside of the filament. D0 and D1 are part of the so-called conserved region (**Figure 3**) of FliC and are essential for the interaction with TLR5. The D0 domain contributes to TLR5 activation but has no or little effect on binding to TLR5, whereas the D1 domain makes substantial contributions to both high affinity binding and TLR5 signalling [66]. The D2 and D3 domains are part of the hypervariable region of the FliC. D2 and D3 are thought to stabilize each other, forming the most stable part of FliC [67]. In addition, the D3 domain, which is exposed on the surface of the filament part, contains the major antigenic determinants [68] and is highly resistant against proteolytic degradation [67, 69].

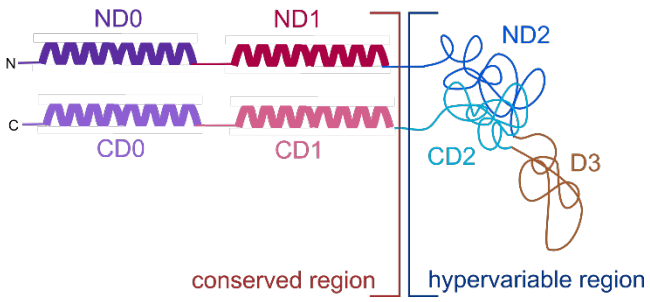



Figure 3. Schematic representation of the folded FliC. The conserved region is formed by the domains D0 and D1 on the N terminus (ND0, ND1) and C terminus (CD0, CD1). The hypervariable region comprises ND2 and CD2 domains and a central D3 domain (brown) (**Figure 3 a from [1]**)

Two different pathways for recognition of FliC have been described: extracellular FliC is recognized by TLR5; intracellular FliC is recognized by interleukin-converting enzyme protease-activating factor (Ipaf) and Nod-like receptor apoptosis-inhibitory protein-5 (Naip5) [70]. Here, I will focus on TLR5 activation by FliC. Among others, TLR5 is expressed by dendritic cells (DCs), natural killer (NK) cells, basolateral on epithelial cells (EC) and lymph node stromal cells [71]. Functional studies suggest distinct roles of TLR5 activation in EC *vs.* immune cells [72]. Upon contact, two FliC monomers bind to two TLR5 monomers generating a heterotetrametric unit [73]. Subsequently, the intracellular segment of the TLR5, the Toll/interleukin-1 receptor (TIR) domain, triggers a MyD88-dependent downstream activation cascade resulting in activation of mitogen-activated protein kinases (MAPK), like p38 and IκB, leading to the induction of transcriptions factors, e.g., the activator protein 1 (AP-1) and NF-κB, which results in transcription of a panel of at least 500 genes [74]. This panel includes

genes encoding for chemo attractants, stress induced genes like heat shock proteins, genes mediating anti-apoptotic effects as well as genes encoding for AMPs [75]. It was shown that activation of TLR5 on epithelial cells favours the expression of chemokines, whereas the activation on hematopoietic cells preferentially results in the production of pro-inflammatory cytokines [76].

In IBD, opposing effects have been described for FliC. On the one hand, *Tlr5*$^{-/-}$ mice develop spontaneous colitis [77], which indicates an anti-inflammatory role of FliC. On the other hand, it has been shown that in the mouse model of acute DSS-induced colitis, the activation of TLR5 leads to an exacerbation of severity of colitis [78, 79]. In humans, a common stop codon polymorphism in the ligand-binding domain of TLR5 (*TLR5*$^{392STOP}$), which is unable to mediate FliC signalling and acts in a dominant fashion [80], was negatively associated with CD in individuals of Ashkenazi Jewish ethnicity. This finding suggests that a reduced TLR5 function may protect from CD [81].

In summary, it can be said that so far neither the genetic content of *E. coli* nor the structure of FliC can distinguish symbiotic from pathobiont strains, and it is not clear on the basis of which mechanism microbial components induce different immune reactions in the host.

# 2 Aim of the work.

The aim of this work was to help clarify the distinctions existing between symbiotic and pathobiontic *E. coli*. Two projects were carried on for this purpose.

In the first project, a machine learning-based tool was developed that can differentiate between symbiotic and pathobiotic *E. coli* strains.

The second project aimed to investigate which structural components of FliC trigger an anti-inflammatory or inflammatory immune response in the host.

# 3 Results

## 3.1 Manuscript : “Characterization of *E. coli* carried by IBD patients through logistic regression”

Friday, 17. February 2022

Declaration on contributions

This is to certify the contribution of each author to the manuscript included in the thesis by Thomas Hagemann Gutter titled :

“Characterization of commensal *E. coli* through genetic content and Flagellin structure”

| Author | Author position | Scientific ideas % | Data generation % | Analysis & interpretation % | Paper writing % |
|---|---|---|---|---|---|
| Thomas Hagemann Gutter | 1 | 80 | 90 | 90 | 90 |
| Ulrich Schoppmeier | 2 | 10 | - | 10 | 10 |
| Walburga Dietrich | 3 | | 5 | | - |
| Yurgadül Zopf | 4 | | 5 | | |
| Julia-Stefanie Frick | 5 | 10 | - | | |
| Title of paper | | Characterization of *E. coli* carried by IBD patients through logistic regression. | | | |
| Status in publication progress | | In preparation for submission | | | |

Thesis Supervisor

Prof. Dr. med Julia Stefanie-Frick

### 3.1.1 Introduction

Inflammatory bowel disease (IBD) is a chronic inflammatory disorder of the gastrointestinal tract comprising Crohn's disease (CD) and ulcerative colitis (UC). Both CD and UC present cycles of inflammation with strong clinical symptoms (e.g., abdominal pain, diarrhoea and fatigue) and phases of remission [1]. Currently, IBD has a global distribution and a high prevalence in industrialized countries, such as the USA and European countries like Norway, with 319 CD and 505 CD cases per 100 000 inhabitants, respectively [2].

Up to now, the IBD etiology has not been fully explained. Nevertheless, the key influence of the individual's genetics, environmental factors and the composition of the gut microbiota (GM) has been widely reported. GM is defined as "the collection of bacteria, archaea and eukarya colonizing the GI tract"[3]. It fulfils essential functions for the host, such as protection against colonization or infection with pathogenic bacteria and production of compounds that are essentially needed by the host or participate in epithelial homeostasis, such as vitamins [4] and short chain fatty acids SCFA [5], respectively. One of the most important features of the gut microbiota is its contribution to the host immune homeostasis, both local and systemic. It is known that different autoimmune diseases are associated with alterations in the composition of the intestinal microbiota, in particular with a reduced GM diversity [6].

Currently, diagnosis of IBD relies on symptoms´ evaluation and direct observation of the gut mucosa through endoscopic examination [7]. Given its invasiveness, endoscopy does not allow a continuous monitoring of the mucosa and besides, it does not provide elements to predict mucosal inflammation. In order to overcome these drawbacks, there have been substantial efforts towards the identification and use of non- or mildly invasive biomarkers, which are at the same time also accurate and specific. In this regard, the C-reactive protein (CRP) and faecal calprotectin (FC) are currently the most accepted diagnostic markers of IBD, albeit only complementary due to their low specificity and sensitivity [8]. Hence, numerous potential markers have been and continue to be investigated that can be used simultaneously to strengthen the diagnostic confidence level [8].

To this end, high throughput sequencing analyses have been instrumental in unveiling changes in GM composition that specifically occur during inflammation of the intestinal mucosa. In this way, we have learnt that in IBD there is an overrepresentation of the phyla Bacteroidetes and Proteobacteria, the latter including the Enterobacteriacea

*Escherichia coli* [9]. Nevertheless, the highly dynamic nature of the GM between individuals with or without IBD makes it difficult to confidently and causally link the individual health status to a specific GM composition [10]. Given its association with IBD, further efforts have been directed towards exploring the pangenome composition of *E. coli* in search for genetic elements that profile *E. coli* under this condition [11]. Nonetheless, no unique marker has been found so far.

The GM contains at least 160 bacterial species [12] for a total of $10^{12}$ single bacteria actively interacting with each other. Therefore, the identification of molecular microbial markers for IBD is extremely complex. Despite the ongoing increase in genome sequences and computational tools, we believe that it is necessary at first to approach such a complicated problem with a simple methodology before scaling up to highly complex statistical methods requiring large informatics resources.

The aim of this report is to show that it is possible to classify *E. coli* isolated from non-IBD volunteers and IBD patients based on their genetic content using a simple statistical approach through logistic regression. Our findings support the suitability of an easy and approachable method for bacterial genetic profiling, such as the case of *E. coli* from IBD patients.

### 3.1.2 Results

#### 3.1.2.1 The *E. coli* pangenome consisting of IBD and non-IBD isolates has a low clonality.

Twenty-seven stool samples were collected from non-IBD volunteers and 23 stools samples from IBD patients were kindly provided by Prof. Dr. med. Yurdagül Zopf (University Clinic, Erlangen). Each stool sample was serially plated on sheep-blood agar and on endo-agar to specifically isolate single colonies of Gram-negative bacteria. Among those, *E. coli* was first identified morphologically and then by 16S sequencing. Parallel identification was carried out through MALDI-TOF (Diagnostic department, Medical Microbiology, University Hospital Tübingen). For each of the total 50 *E. coli* isolates, the chromosomal DNA was extracted using a commercial kit and subject to whole genome sequencing. After annotation, the pangenome of the whole cohort was obtained, revealing a total of 17687 genes. This information allowed us to generate a so-called presence/absence matrix (m P/A) displaying which genes in the pangenome were present or absent in each strain (see methods). From the matrix we were able to

correlate the presence of each single gene in the non-IBD group with its presence in the IBD group (**Figure 1**).

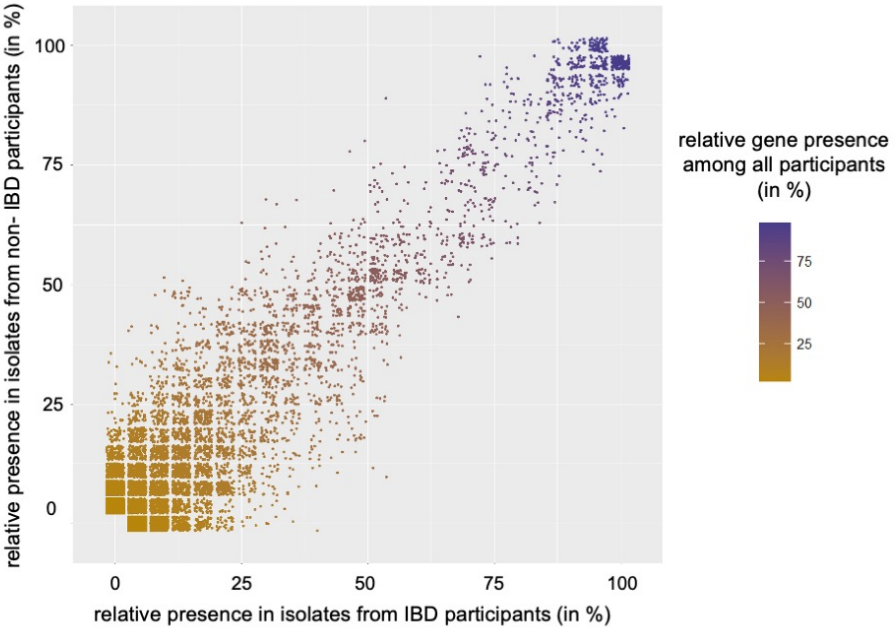



**Figure 1. Relative abundance of genes is similar in IBD and non-IBD and most of the genes have a low relative abundance**. The scatter plot was jittered to avoid overplotting and each dot represents a single gene. The color code depicts the relative presence of each gene across all the 50 *E. coli* isolates. If a gene is shared by few isolates is colored gold, while a gene shared by all the isolates is purple colored. The position of a gene in the plot shows its presence relative to the number of isolates on each group (non-IBD or IBD). As an example, if the "gene A" is found in 10 of the total 50 isolates, then its relative presence is 20% (hence, colored into the gold color code). If the same gene is found in 7 out of 23 IBD isolates, then its relative presence in the IBD group is 30,4 %. If the gene "A" is carried by 3 out of the 27 non-IBD isolates, then its presence is 11,1 % in that group. Quite similar to a coordinate system, "gene A" will be plotted on the "x axis" (IBD) on 30,4 and in the "y axis" (non-IBD) as 11,1.

Overall, this correlation analysis revealed an even distribution of genes, meaning that one group or the other was not clearly overrepresented in terms of gene presence. This is evidenced in the scatter plot in **Figure 1** where the distribution of most genetic elements describes a diagonal region in the plot, starting in the lower-left corner to the upper-right. This means that most genes are similarly represented within their respective groups (IBD and non-IBD). The periphery of the latter region is

characterized by genes that are unevenly represented in both groups and may provide higher discriminatory information with respect to a distinct gene content in the non-IBD and IBD group.

Most of the genes in the pangenome have a relative presence in both groups lower than 25 % (**Figure 1**, lower-left) while fewer higher than 90 % (**Figure 1**, upper-right), hinting to a cohort with low clonality.

#### 3.1.2.2 Measuring relatedness using the whole genome does not allow discriminating between IBD and non-IBD *E. coli* isolates

Next, we wanted to find out how far or close the strains relate to each other in terms of whole genome content similarity and whether this distance may be related to the source of each *E. coli* strain, i.e., IBD patient's vs non-IBD volunteers. For this, we employed a clustering algorithm that enabled the calculation of the Euclidian distance between the strains (see methods). Strains that have a similar gene content will group together, as illustrated in the heatmap for the whole pangenome (**Figure 2**)

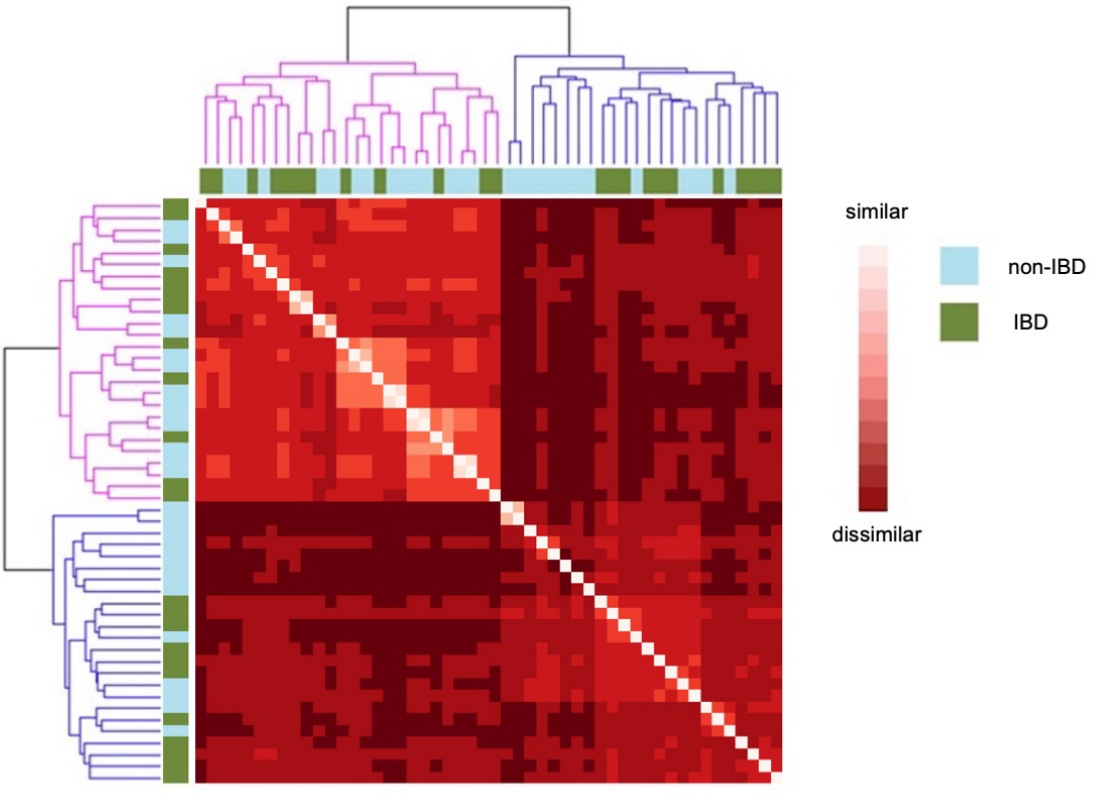



**Figure 2. The analysis of the *E. coli* pangenome does not show a link between gene content and the host source**. Closeness in terms of gene content between strains is represented by light-red, conversely distant strains are depicted in dark red. The strip on the left side of the heatmap shows the source of the *E. coli* isolates, with light blue

corresponding to non-IBD and olive to IBD. The heatmap was built based on the depicted dendrograms, which represent how two strains are related. Colors pink and blue in the dendrogram show the major clusters found.

This whole genome distance analysis revealed that all the isolates from the very beginning split in two major groups (pink and blue dendrograms in **Figure 2**), with little correspondence to the source where the bacteria were isolated from (**Figure 2**, left strip). Hence, we found no evidence that strains closely related in terms of gene content belonged to one specific group or the other. This observation together with the fact that those genes that have the same or highly similar gene presence in both groups do not provide discriminatory information about the IBD or non-IBD group prompted us to focus instead on those that are unevenly abundant in the non-IBD and IBD group (**Figure 2**). For this, the genes whose relative presence was more similar in both groups were filtered out, a process that was performed through the use of different thresholds, or “cutoffs” (see methods). As an example, a certain common gene that is represented by a 75 % in one group and 25 % on the other has a relative presence-difference of 50 % and will be included or not in the scatter plot depending on the chosen cutoff. By inspecting different cutoffs (supplementary Information, **S1 Figure**), we noticed a clearer discrimination between the two groups when we used a cutoff of 35 % (**Figure 3 and 4**)

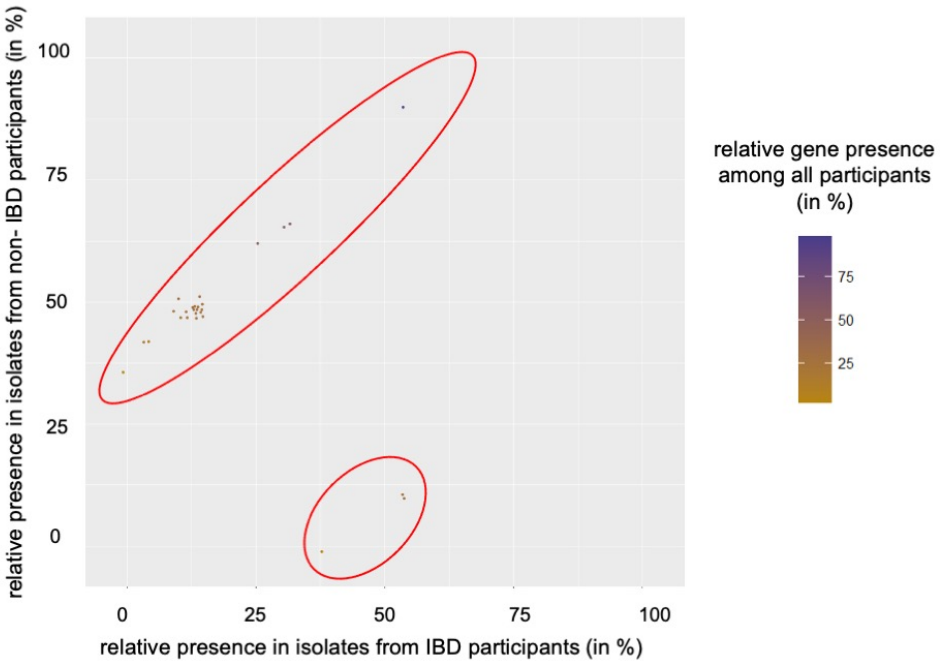


**Figure 3. The non-IBD and IBD group differ in terms of abundance of uneven shared genetic elements**. After removing those genes that were more similar based on a 35% cut-off, two clouds of genes (red ovoid) can be distinguished. The difference in number of uneven shared genes is evident when both groups are compared for a relative abundance of 50%.

The removal from the pangenome of those genes presenting similar abundance in both groups (within 35% cut-off threshold), uncovered differences in term of gene abundance. This can be appreciated better when comparing both groups at a 50% relative gene presence. At this threshold the non-IBD isolates present approximately 7-fold more genes than the IBD group. Besides, the non-IBD cloud is more dispersed compared to IBD. Both findings led us to investigate whether the relatedness of the strains changed as well.

#### 3.1.2.3 A clustering analysis solely based on uneven abundant genes hints to a different genetic profile between IBD-derived isolates and non-IBD ones.

Aiming to establish whether the dissimilarity in genetic content between non-IBD and IBD groups was reflected on the strain level, clustering was performed as described before (**Figure 4**).

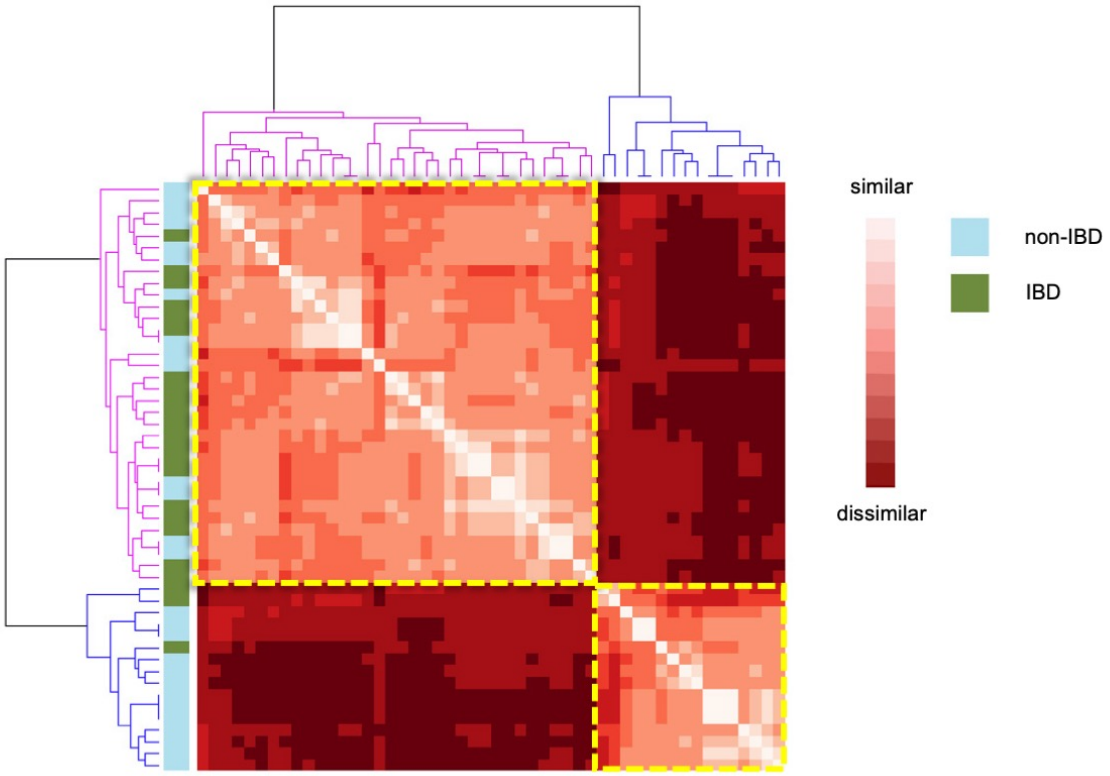



**Figure 4. *E. coli* isolates are found to form two clusters when genetic elements common to IBD and non-IBD strains are removed**. When depicting strain distance considering only those genes outside the difference of 35% cutoff, two major clusters emerge (highlighted with segmented yellow squares). Each cluster contains strains that are closer related in terms of gene content. The lower-right cluster contains mainly *E. coli* genomes from non-IBD cases but underrepresented compared to the total non-IBD cases. The upper-left cluster contains a higher number of strains from both groups but no predominance of one or the other is observed.

Even though we observed a structure when clustering genes based on the 35 % cut-off (**Figure 4**), the scattering observed for both groups (non-IBD and IBD) did not lead to a conclusive pattern of gene content and the respective origin of the strains. Altogether, the attempt to establish any kind of predictive power in order to classify both non-IBD or IBD groups employing the genes that are above the 35 % threshold was not successful. Nevertheless, the presence of clear distinctive clusters

(particularly observed for the low-right cluster on **Figure 4**) hinted to the existence of some degree of relatedness among strains within the same group. This led us to suspect that the gene content of the strains was not totally independent of their source.

#### 3.1.2.4 Classification strategy through logistic regression.

The obtained results prompted us to perform further analyses with a different approach that took into account all the genes from the pangenome. The goal was to build a model allowing the accurate classification of any *E. coli* genome as belonging to either non-IBD or IBD group based on a discrete number of genetic features. Logistic regression was chosen as the statistical approach for building such a model (detailed in methods). Briefly, the cohort of 50 *E. coli* isolates was randomly divided into two groups: one group of 12 isolates, which constituted the “validation data”, and a second group of 38 isolates (21 non-IBD and 17 IBD) as “training data”. The validation data group was used only to determine the accuracy of the final model as this process requires data that were never used before. In contrast, the purpose of the training data was to internally train and validate several models. In this way, we generated a high number of models and the final most accurate model along with the validation data was the proof of concept of this statistical approach. A scheme showing an overview of the process of model building and genetic feature selection is depicted in **Figure 5**.

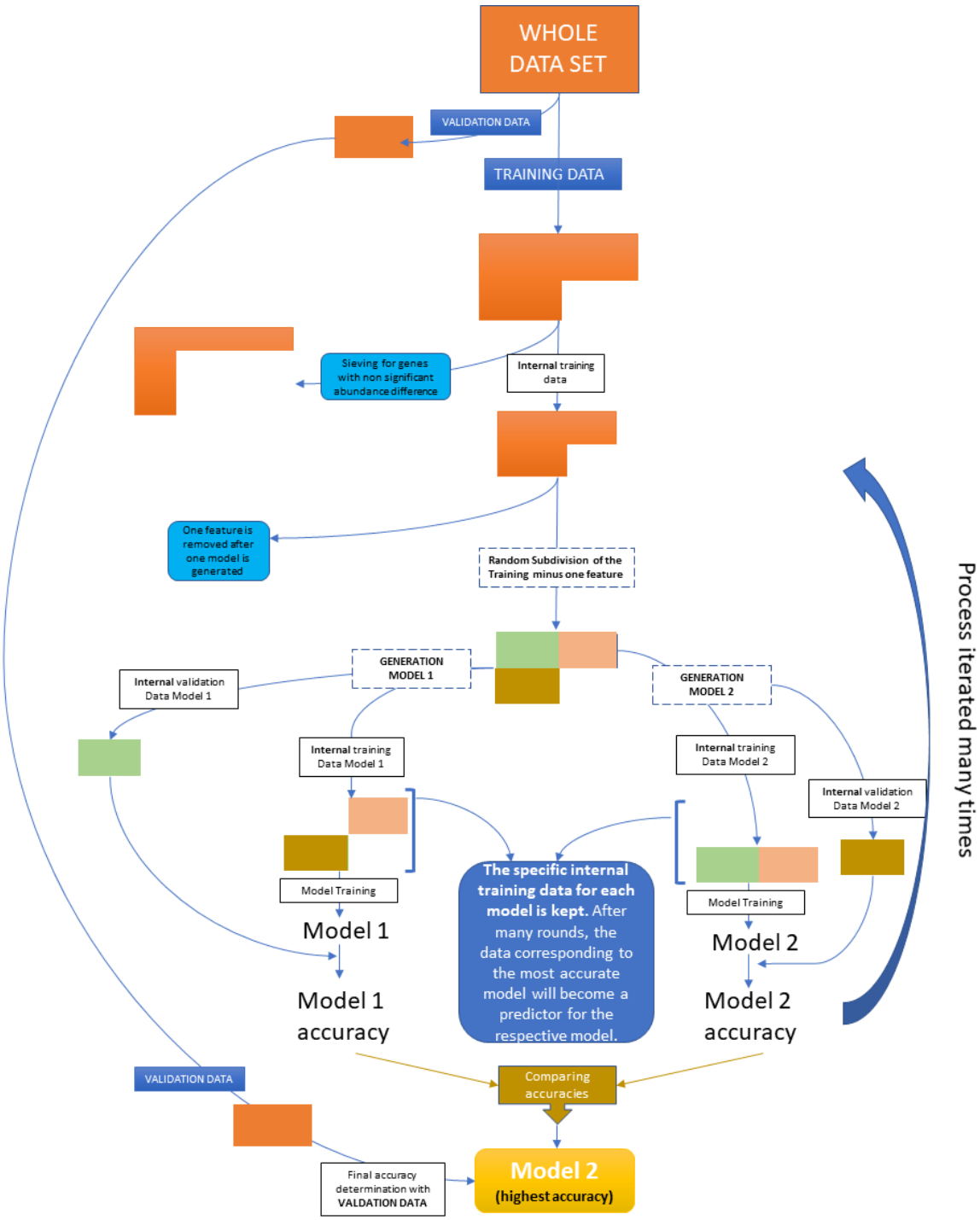



**Figure 5. Simplified example of the pipeline used to generate a final accurate model.** Before the modelling begins, the data is split. A portion of the data is kept outside the model generation process and used only at the very end to test the accuracy of the "best" model (or the most accurate classifier) among those generated. Therefore, these data are called validation data. The other portion of the data, or training data, is at first sieved, removing all those genes that have a similar relative abundance between both groups. The purpose of this is to remove the less informative genes in terms of finding differences between IBD and non-IBD isolates. The sieved training data are then used

to generate different models. In this example, the sieved training data are split into three parts (olive, peach and brown rectangles). Two folds of the sieved data (brown and peach) become what is called “internal training data” for model 1. The third part (olive rectangle) consists of the data used to evaluate the accuracy of model 1 and is called internal validation data. A similar pipeline is used for the generation of model 2, with the difference that the data set composition has now changed: the internal validation data are now represented by the brown ones while the internal training data are made of the olive and peach data. The process continues by reshuffling the sieved internal data and thereby, changing the composition of validation and training data set. This reiterative process allows generating as many models as desired. Note that each time a new model is created, one new feature is deleted from the data set, which allows measuring the impact of that feature as a predictor. In the end, the most accurate model and the remaining predictors are tested with the validation data set.

#### 3.1.2.5 Four genetic features correctly predict all non-IBD isolates and 83% of the IBD cases in the training data set.

The 38 isolates (or cases) constituting the training data for our actual analysis still contained a number of genes that were equally shared by the IBD and non-IBD group. For this reason, a proportional test was used to select only those genes that had a significantly different presence between both groups. This resulted in the selection of 38 genes, which were used for the modelling process. At the beginning, N=38 models were trained, each of which considered one different genetic feature less, hence N-1 =37 genetic features per model. The accuracy of each of the N models was tested with internal validation data. The model with highest accuracy in classifying the strains according to their origin was kept. The genetic feature that was removed for training to build up this model was discarded from our analysis. This allowed to reduce features and concentrate on those that directly impacted the accuracy of the current model. Next, in the same way 37 models with 36 features were built, the model with the highest accuracy was kept and the removed genetic feature was again no longer considered. This training process was reiterated until one final model remained with 3 features, that is, the minimum number of features possible with the training data remaining at this point. The process described above constitutes one single run and was repeated for a total of 83 runs.

Every single run gave as a result one best model and the features that served as predictors for that model. In order to select the most relevant features out of the 83 runs, the feature frequency was determined by counting in how many runs a defined

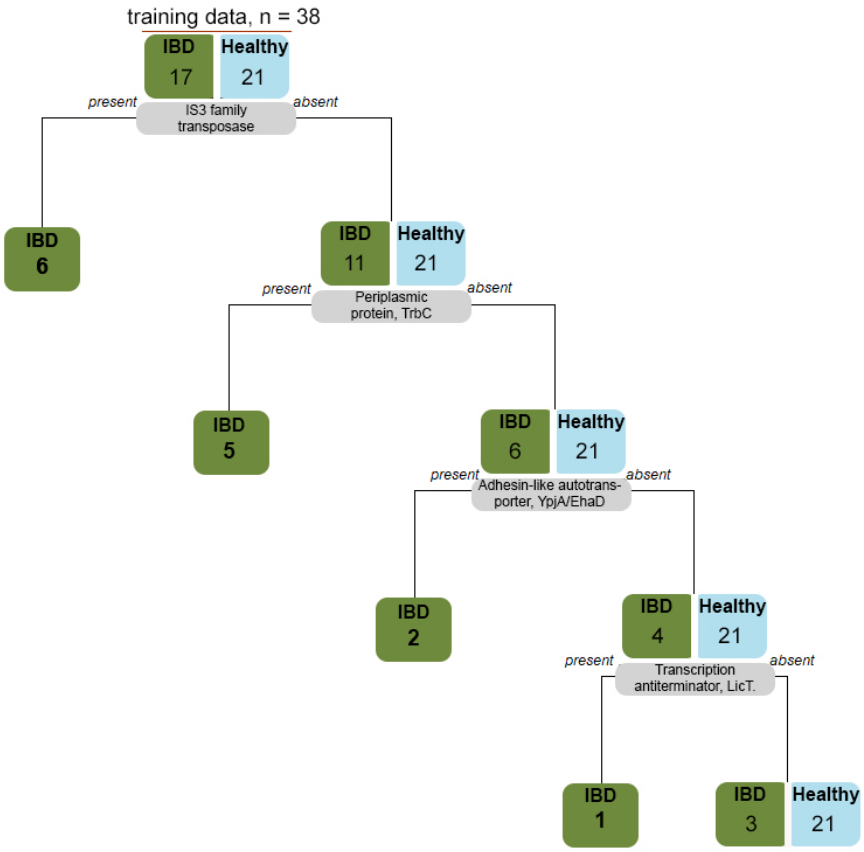
feature was found. This enabled us to list the features according to their frequency and inspect their presence in the training data according to the source of the isolate (supplementary information **S1 Table**) This led to the identification of four predictors that were almost only present in IBD cases (14 out of 17) and totally absent in the non-IBD training data (**Table 1**). This observation was confirmed by the training of a new model that considered these genetic features only.

**Table 1. Four genetic elements are always absent in non-IBD isolates**. By simple inspection of the most frequent genes out of 83 runs, we observed a group of 4 annotated genetic elements that were absent in the non-IBD isolates and distributed among 14 out of 17 IBD isolates. Note that these data correspond to our training data only.

| **Predictors corresponding to our classifier** | **IBD Isolates (n=17)** | **Non-IBD Isolates (n=21)** |
|---|---|---|
| Transcription antiterminator LicT | 5 | 0 |
| Putative autotransporter precursor | 6 | 0 |
| Type-F conjugative transfer system pilin assembly protein | 8 | 0 |
| Integrase core domain protein | 9 | 0 |

We noticed that the most frequent genes that were totally absent in the non-IBD training group were unevenly distributed within the IBD training group, with no IBD isolate containing all the genetic features at the same time. This can be better represented by a decision tree, which was constructed by simple inspection of the presence/absence of the mentioned genetic features among the isolates belonging to the training data set (a) and then for the isolates belonging to the validation data set (b). In the tree, each feature corresponds to a node where the classification is made based on the presence or absence of the feature itself (**Figure 6**).

a)


training data, n = 38
IBD 17
Healthy 21
present
IS3 family transposase
absent
IBD 6
IBD 11
Healthy 21
present
Periplasmic protein, TrbC
absent
IBD 5
IBD 6
Healthy 21
present
Adhesin-like autotrans-porter, YpjA/EhaD
absent
IBD 2
IBD 4
Healthy 21
present
Transcription antiterminator, LicT.
absent
IBD 1
IBD 3
Healthy 21

b)

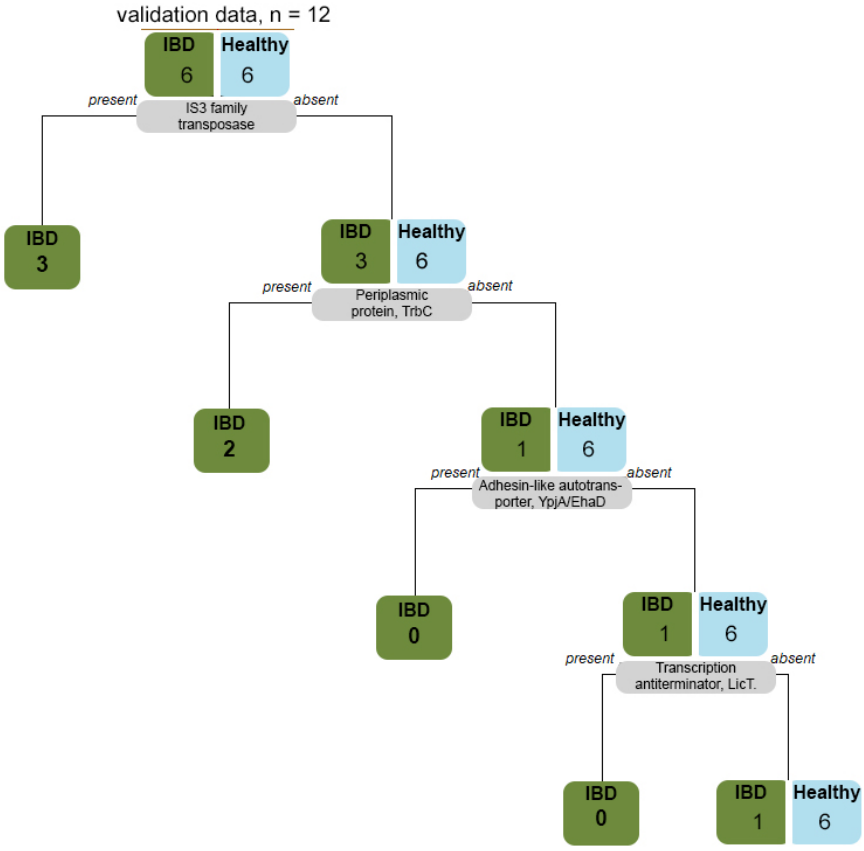

validation data, n = 12
IBD 6
Healthy 6
present
IS3 family transposase
absent
IBD 3
IBD 3
Healthy 6
present
Periplasmic protein, TrbC
absent
IBD 2
IBD 1
Healthy 6
present
Adhesin-like autotrans-porter, YpjA/EhaD
absent
IBD 0
IBD 1
Healthy 6
present
Transcription antiterminator, LicT.
absent
IBD 0
IBD 1
Healthy 6

**Figure 6. The identified predictors are dispersed in the IBD group. Both trees show on each node how the isolates are classified when a specific genetic feature is present/absent.** On each tree the color cyan represents non-IBD cases and olive IBD. **a)** Training data: The first decision classified 6 strains correctly as part of the IBD group, on the second decision 5 isolates more, next 2 isolates and by the end only three IBD strains were incorrectly classified, hence the prediction was accurate in 14 out of 17 IBD cases. **b)** Testing predictor accuracy with training data: When the same markers were used for the training data, all the non-IBD strains were again correctly classified and only one misclassification occurred assigning one IBD case to the non-IBD group, meaning 5 out of 6 correct classifications for IBD cases.

The test of the predictors using the validation data confirmed what was observed for the training data. In particular, for the IBD cases, the predictive accuracy of training and validation data was 82,3% and 83,3%, respectively, highlighting the strength of the identified predictors.

### 3.1.3 Discussion

The currently available IBD biomarkers have been implemented as a valuable complement of the traditional diagnostic tests. However, while faecal calprotectin (FC) allows prediction of inflammation, its reliability is low and cannot replace the current endoscopic gold standards [7]. Hence, generating a panel of biomarkers would substantially increase the accuracy of IBD diagnosis.

This has led to the search for new biomarkers within the gut microbiota, one of the main elements influencing IBD [13]. The search for microbiological markers, such as unique changes in GM composition, has been heavily propelled in the recent years by the extensive use of next generation sequencing (NGS) [14]. NGS has allowed high throughput in-depth exploration of the genomic content of the bacterial populations inhabiting the gut. At the same time, the production of such a large amount of data presented a challenge from the point of view of data processing capabilities [15].

We believe that through simple statistical tools that neither require high expertise nor massive informatics resources it is possible to obtain meaningful and statistical sound results. With this goal, we chose to identify candidate markers as predictors that accurately classify a bacterium with an increased abundance in IBD patients such as *E. coli*.

Our approach consisted in exploring the relative abundance of each of the annotated genes represented in the pangenome of our cohort and investigating how they might

be related to the source the bacteria were isolated from, i.e., either IBD patients or non-IBD. By choosing the most unevenly distributed genes to produce and train different models and after many trainings runs, we could select the most accurate model together with 4 predictors. When these were tested on the validation data, we observed that 83.3% of the IBD strains could be accurately classified and no misclassification occurred for the non-IBD group.

According to their automatic annotation, these predictors encode a transposase of the IS3 family, a periplasmic protein (TrbC), an adhesin-like autotransporter (YpjA/EhaD) and a transcription antiterminator (LicT). The IS3 family transposase and the periplasmic protein TrbC are associated with recombination events and conjugation, respectively [16], suggesting the presence of mobile genetic elements in IBD-associated *E. coli* strains. This is consistent with the acquisition of virulence traits by human enteric opportunistic pathogens through intra- and interspecies transfer of genes that are located on mobile elements such as plasmids, conjugative transposons and integrons. EhaD or YpjA are putative autotransporters involved in adhesion, which is a known pathogenic feature of *E. coli* [17]. LicT regulates the ribosomal RNA gene *rrnB* from *E. coli* by transcriptional termination, and may be associated with metabolic adaptation [18]. Altogether, the encoded functions of the four predictor genes point to factors that favor pathogenic traits, which would be in line with the features of a bacterium associated with inflammation.

Nevertheless, it is far too early and out of the scope of this report to claim that the presence of these features in IBD strains fosters gut inflammation in IBD patients or increases the fitness of those strains under inflammatory conditions. Further investigation is required to address those issues.

Even though our results are statistically sound, we are aware of the limitations of the study, which made use of a numerically and geographical restricted cohort. Before claiming an immediate clinical use of our model and predictors as a diagnostic tool, our approach needs to be reproduced with a higher number of *E. coli* genomes, ideally with a broader geographical distribution.

Moreover, gene deletion and phenotypic evaluation would be of much interest to evaluate the role of the 4 identified predictors in *E. coli* pathogenesis and IBD etiology.

Overall, our simple statistical modelling through logistic regression allowed us to identify 4 genetic elements out of more than 17000 that classified with high accuracy IBD and non-IBD *E. coli* genomes.

We showed with this work that it is possible to tackle a complex problem of genetic characterization of two bacterial populations without high complicated algorithms or extensive computer resources. Hence, contributing with a method reachable not only to expert bioinformaticians but to a wider community of researchers

### 3.1.4 Material and Methods

#### 3.1.4.1 Sample preparation.

Sample collection

Stool samples from non-IBD individuals were collected among co-workers and students from the Tübingen university hospital between August and October 2016. The samples were processed within 2 hours from the time of collection. IBD stool samples were kindly provided from Prof. Dr. med. Yurdagül Zop (Universitätsklinikum Erlangen). The samples were collected over a six-month period and then shock frozen at -80°C. Volunteers received informed consent prior to sample collection.

The experimental protocols were approved by the ethics committees of the Universitätsklinikum Tübingen (043 / 2023BO2) and Universitätsklinikum Erlangen (n 145_16 B)

Bacterial isolation and sequencing

Each stool sample was at the same time plated on Columbia Blood Agar (Thermo scientific TM) and Endo-Agar plates. Preliminary *E. coli* identification was made through visual inspection of colony morphology and further identification was performed by MALDI-TOF (Bruker Daltonik Maldi Biotyper) and sequencing of the gene codifying for the 16S RNA ribosomal unit, followed by Blast analysis [19]. Bacterial DNA was isolated using the QIAGEN Genomic-tip 100/G kit, according to the manufacturer protocol. The DNA concentration was determined by fluorimetry (Qubit 2.0) Sequencing was performed using Illumina MiSeq sequencer as a paired end sequencing and a read length of 300 bp. The targeted fragment size was 500 bp.

Sequencing and assembly.

High quality trimming of the reads was performed and A5 [20] assembler was chosen. The scaffolds generated were evaluated according to their quality cut-off and those with lower quality than expected were discarded and the reads reassembled this time with the SPAdes [21] assembler pipeline. Annotation for all the scaffolds was

performed using rapid prokaryotic genome annotation, Prokka [22]. The annotated scaffolds quality was analysed using the tool CheckM version 1.1.3 [23].

#### 3.1.4.2 Statistical Methods

Whole pangenome analysis

The pangenome of the whole cohort (n=50) was determined through the pangenome pipeline tool Roary [24] under a paralog splitting setting of 95%. In this way the gene content of each strain arranged in the so-called gene presence/absence (m P/A) was obtained.

The (m P/A) was reformatted in order to perform the following analysis in R [25]. The relative presence of each of the genes from the pangenome was determined by counting how many isolates of each group harbored a certain gene with respect to all isolates belonging to the respective group (IBD or non-IBD). The ratio between the number of isolates with a certain gene and the total amount of isolates was expressed as a percentage in order to compare both groups. The same calculation was repeated for the whole pangenome and represented as a scattered plot.

Filtered pangenome analysis

The examination of unevenly represented genes was accomplished by filtering genes that were shared by both groups. For this purpose, the first step was to filter out genes that were found to be present in a 100% of the cases on either group with the aim to reduce bias in the prediction process. Next the difference of relative abundance of a certain gene between the IBD and non-IBD group was calculated, and this calculation was repeated for the whole pangenome. The pangenome was stepwise filtered starting with common genes that had a relative abundance difference of 10%, and then 20, 30, 35 and 40 % (supplementary information **S1 Figure**).

Heatmaps

The relative gene presence data was used to explore relatedness between the strains. The heatmap tool from R [25] was used under default settings to calculate the Euclidian distance between the strains and clustering using the complete linkage method. For this, the input data were the whole unfiltered pangenome content and the five different cut-offs mentioned above. The dendrogram representing relatedness of each pair of strain was generated by default from the same tool.

Statistical Modelling

Statistical analysis was performed using R [25], using the following libraries*.

| | | | |
|---|---|---|---|
| dplyr_1.0.5 | gridExtra_2.3 | purrr_0.3.4 | stringr_1.4.0 |
| forcats_0.5.1 | gridGraphics_0.5-1 | RColorBrewer_1.1-2 | tibble_3.1.1 |
| ggplot2_3.3.3 | lattice_0.20-41 | readr_1.4.0 | tidyr_1.1.3 |
| gplots_3.1.1 | plyr_1.8.6 | Rmisc_1.5 | tidyverse_1.3.1 |

*The code build for this work will be available on the Github repository.

Best predictors determination

The process for determining the best model was repeated for a total of 83 times. Hence 83 sets of predictors were obtained. The predictors obtained on each run were listed and then counted to see which predictors were the most frequently found (supplementary information **S2 Table**). By simple inspection, it was observed that four predictors were absolutely absent in non-IBD cases, hence they were considered as the most interesting predictors for the classification of cases as IBD or non-IBD.

3.1.5 Acknowledgments.

We thank Dr. Lena Michaelis (Institute of Microbiology and Hygiene (IMIT), University Hospital from the Tübingen University) and Dr Libera Lo Presti (Cluster of Excellence “Controlling Microbes to Fight Infection”, University of Tübingen) for proofreading this manuscript. We also thank Prof Dr Matthias Willmann for his support for the genomes assembly process. We specially thank the co-workers and students from the IMIT for providing us with the stools samples for this study.

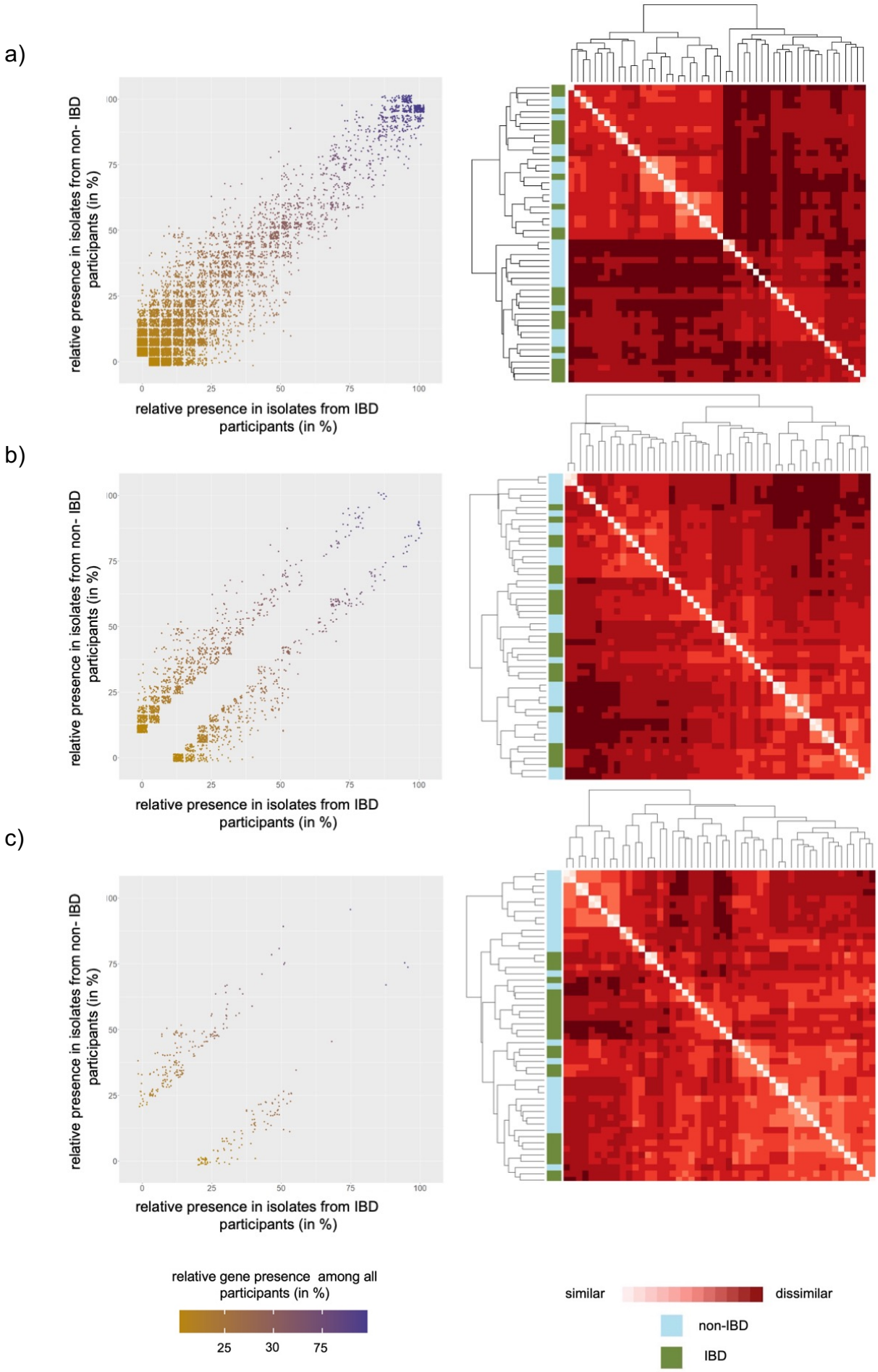
### 3.1.6 Bibliography.

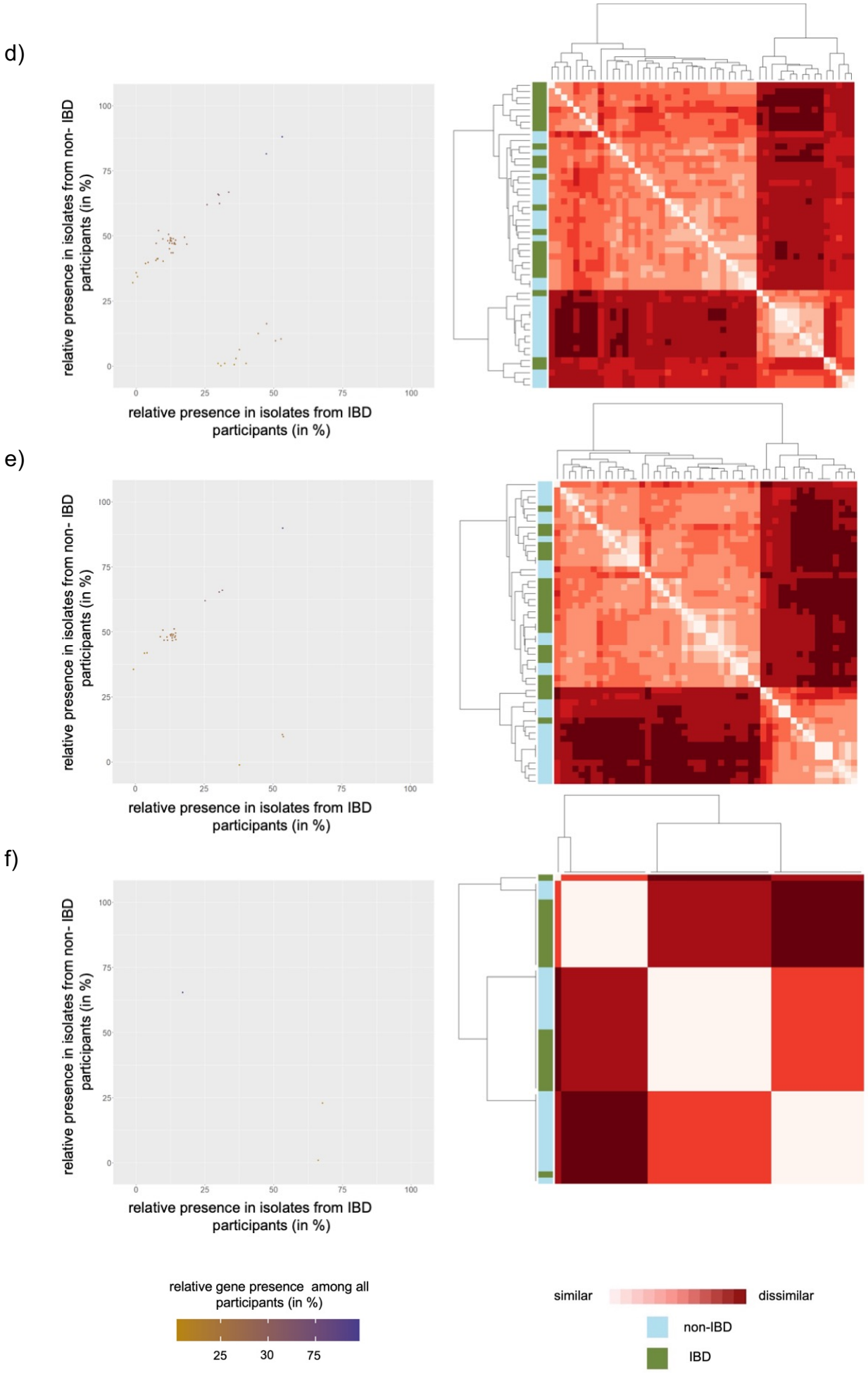

1. Villablanca, E.J., K. Selin, and C.R.H. Hedin, Mechanisms of mucosal healing: treating inflammatory bowel disease without immunosuppression? Nat Rev Gastroenterol Hepatol, 2022.
2. Ng, S.C., et al., Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet, 2017. 390(10114): p. 2769-2778.
3. Thursby, E. and N. Juge, Introduction to the human gut microbiota. Biochem J, 2017. 474(11): p. 1823-1836.
4. LeBlanc, J.G., et al., Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin Biotechnol, 2013. 24(2): p. 160-8.
5. Bachem, A., et al., Microbiota-Derived Short-Chain Fatty Acids Promote the Memory Potential of Antigen-Activated CD8(+) T Cells. Immunity, 2019. 51(2): p. 285-297 e5.
6. Matsuoka, K. and T. Kanai, The gut microbiota and inflammatory bowel disease. Semin Immunopathol, 2015. 37(1): p. 47-55.
7. Abreu, M.T. and N. Harpaz, Diagnosis of colitis: making the initial diagnosis. Clin Gastroenterol Hepatol, 2007. 5(3): p. 295-301.
8. Dragoni, G., T. Innocenti, and A. Galli, Biomarkers of Inflammation in Inflammatory Bowel Disease: How Long before Abandoning Single-Marker Approaches? Dig Dis, 2021. 39(3): p. 190-203.
9. Guo, S., et al., A Simple Fecal Bacterial Marker Panel for the Diagnosis of Crohn's Disease. Front Microbiol, 2019. 10: p. 1306.
10. Dubinsky, M. and J. Braun, Diagnostic and Prognostic Microbial Biomarkers in Inflammatory Bowel Diseases. Gastroenterology, 2015. 149(5): p. 1265-1274 e3.
11. Darfeuille-Michaud, A., et al., High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. Gastroenterology, 2004. 127(2): p. 412-21.
12. Yang, J., et al., Species-Level Analysis of Human Gut Microbiota With Metataxonomics. Front Microbiol, 2020. 11: p. 2029.
13. Qiu, P., et al., The Gut Microbiota in Inflammatory Bowel Disease. Front Cell Infect Microbiol, 2022. 12: p. 733992.

14. Caputo, A., P.E. Fournier, and D. Raoult, Genome and pan-genome analysis to classify emerging bacteria. Biol Direct, 2019. 14(1): p. 5.
15. Land, M., et al., Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics, 2015. 15(2): p. 141-61.
16. Shala-Lawrence, A., et al., The interaction of TraW and TrbC is required to facilitate conjugation in F-like plasmids. Biochem Biophys Res Commun, 2018. 503(4): p. 2386-2392.
17. Wells, T.J., et al., EhaA is a novel autotransporter protein of enterohemorrhagic Escherichia coli O157:H7 that contributes to adhesion and biofilm formation. Environ Microbiol, 2008. 10(3): p. 589-604.
18. Gorke, B. and J. Stulke, Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol, 2008. 6(8): p. 613-24.
19. Altschul, S.F., et al., Basic local alignment search tool. J Mol Biol, 1990. 215(3): p. 403-10.
20. Tritt, A., et al., An integrated pipeline for de novo assembly of microbial genomes. PLoS One, 2012. 7(9): p. e42304.
21. Bankevich, A., et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol, 2012. 19(5): p. 455-77.
22. Seemann, T., Prokka: rapid prokaryotic genome annotation. Bioinformatics, 2014. 30(14): p. 2068-9.
23. Parks, D.H., et al., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res, 2015. 25(7): p. 1043-55.
24. Page, A.J., et al., Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics, 2015. 31(22): p. 3691-3.
25. R Core Team, A language and environment for statistical computing. 2021, Foundation for Statistical Computing.

### 3.1.7 Supplementary information.

**S1 Figure. Stepwise filtered pangenome from 0 to 40 percent threshold.** For each threshold a scattered plot representing the pangenome was built (left panel) and the effect of the threshold in the relatedness between each strain was described thorough a heatmap (right panel) a) Corresponds to a 0% threshold, hence here the full pangenome was represented (left) and the relatedness between each strain is shown on the heatmap on the right side. b) Showing a 10% threshold. c) 20%. d) 35% e) 40% and f) 50%. The color code is the same as on the main text.


a)
b)
c)
relative presence in isolates from non- IBD participants (in %)
relative presence in isolates from IBD participants (in %)
0
25
50
75
100
relative gene presence among all participants (in %)
25
30
75
similar
dissimilar
non-IBD
IBD


d)
e)
f)
relative presence in isolates from non- IBD participants (in %)
relative presence in isolates from IBD participants (in %)
0
25
50
75
100
relative gene presence among all participants (in %)
25
30
75
similar
dissimilar
non-IBD
IBD

**S1 Table. Most frequent predictors listed according their presence in the non-IBD and IBD group**. "Count" means the frequency in which the predictor was found after 83 runs. "Total cases" stands for the total number of cases analysed in the training data that contained that particular predictor. "IBD" and "non-IBD" refers to the number of cases on each group that carried the predictor.

| Gene Function | Count | Total cases | IBD | non-IBD |
|---|---|---|---|---|
| Transcription antiterminator LicT | 25 | 5 | 5 | 0 |
| Putative autotransporter precursor | 68 | 6 | 6 | 0 |
| Type-F conjugative transfer system pilin assembly protein | 64 | 8 | 8 | 0 |
| Integrase core domain protei | 39 | 9 | 9 | 0 |
| Transposase IS66 family protein | 22 | 8 | 0 | 8 |
| Poly-beta-1,6-N-acetyl-D-glucosamine synthase | 31 | 9 | 0 | 9 |
| Hypothetical protein | 43 | 10 | 0 | 10 |
| Small toxic polypeptide | 17 | 8 | 1 | 7 |
| 2,6-dihydropseudooxynicotine hydrolase | 23 | 10 | 1 | 9 |
| Relaxosome protein TraM | 27 | 10 | 1 | 9 |
| Hypothetical protein | 21 | 15 | 2 | 13 |
| Hypothetical protein | 50 | 16 | 2 | 14 |
| Antitermination protein | 66 | 12 | 3 | 9 |
| Hypothetical protein | 38 | 14 | 3 | 11 |
| YHS domain protein | 27 | 17 | 3 | 14 |
| Hypothetical protein | 18 | 16 | 4 | 12 |
| Cytoskeleton bundling-enhancing protein CbeA | 20 | 8 | 6 | 2 |
| Reverse transcriptase | 24 | 8 | 6 | 2 |
| Bacterial Ig-like domain | 38 | 19 | 6 | 13 |
| Small toxic polypeptide | 21 | 16 | 9 | 7 |

**S2 Table. Most frequent predictors by the end of the modelling process**. The predictors were listed according to how frequent they were found by the end of the 83 runs.

| Count | Gene Function |
|---|---|
| 68 | putative autotransporter precursor |
| 66 | Antitermination protein |
| 64 | Type-F conjugative transfer system pilin assembly protein |
| 50 | hypothetical protein |
| 43 | hypothetical protein |
| 39 | Integrase core domain protein |
| 38 | Bacterial Ig-like domain (group 2) |
| 38 | hypothetical protein |
| 31 | Poly-beta-1,6-N-acetyl-D-glucosamine synthase |
| 27 | YHS domain protein |
| 27 | Relaxosome protein TraM |
| 25 | Transcription antiterminator LicT |
| 24 | Reverse transcriptase (RNA-dependent DNA polymerase) |
| 23 | 2,6-dihydropseudooxynicotine hydrolase |
| 22 | Transposase IS66 family protein |
| 21 | small toxic polypeptide |
| 21 | hypothetical protein |
| 20 | Cytoskeleton bundling-enhancing protein CbeA |
| 18 | hypothetical protein |
| 17 | Terminase-like family protein |
| 15 | Carbohydrate acetyl esterase/feruloyl esterase precursor |
| 15 | Bacteriophage lysis protein |
| 15 | hypothetical protein |
| 15 | hypothetical protein |
| 14 | Phage major capsid protein E |
| 14 | Integrase core domain protein |
| 14 | hypothetical protein |
| 12 | hypothetical protein |
| 10 | hypothetical protein |
| 10 | hypothetical protein |
| 9 | IS1 transposase |
| 9 | hypothetical protein |
| 9 | D-cysteine desulfhydrase |
| 7 | hypothetical protein |
| 6 | hypothetical protein |
| 5 | D-cysteine desulfhydrase |
| 5 | hypothetical protein |
| 3 | TraX protein |

## 3.2 Symbiotic properties of EcN FliC

One of the classical tool to classify *E. coli* is the detection of the highly diverse FliC (antigen H) [82]. The polymorphic nature of FliC resides mostly in its hypervariable region (HVR) [68], which, from a structural point of view, allows the terminal conserved regions to come close to each other when each FliC monomer is folded [82]. Of note, it is in its monomeric form that FliC activates TLR5 [83].

Since FliC exhibits a strong immunogenicity activity in general, and in particular it is a remarkable IBD antigen [84], we wondered whether and how the FliC structure would impact TLR5 activation.

To address this question, we first explored the structural features of FliCs from commensal *E. coli* strains isolated from healthy individuals or IBD patients and then dig into those of a known *E. coli* symbiotic commensal.

### 3.2.1 *E. coli* FliC structure is different between commensals inhabiting an inflamed and a non-inflamed environment

The first step was to determine whether *E. coli* inhabiting a niche under homeostatic conditions carried a structurally different FliC from *E. coli* inhabiting a higher stimulated niche. For this, the *fliC* nucleotide sequences belonging to *E. coli* isolated from stool specimens of non-IBD donors (n=27) and IBD patients (n=23) were extracted and translated in order to perform a multi-sequence alignment (MSA).

In order to gain a clear insight into how the structure of FliC differed between the non-IBD and IBD group, the Shannon diversity index (SDI) was calculated. Briefly, the MSA containing all the FliC sequences was generated (n=50). Then, in order to compare each group, the MSA was separated into groups based on the origin (non-IBD and IBD) of the bacterial isolates. The aim of this step was to obtain an MSA of the same length so that each amino acid had always a fixed position.

Then, the Shannon diversity index (SDI) [85] was calculated for each column of the MSA. The SDI allows to compare diversities [86] and has been widely used in ecology and population genetics. It considers “richness” (the total number of observations performed) and “relative abundance” (ratio between each different observation and the total of observation) as variables. For example, if 10 proteins are aligned, the first position is going to have a methionine on all of them. Hence, the variables in this case are “richness=10” (10 methionine in that position) and “relative abundance” = 1 (there are 10 methionine out of a total number of 10 amino acids). The SDI is the natural

logarithm of the relative abundance, which is zero (log (1)=0) in the example illustrated above (see methods, section 5.1.25 for a complete example). We decided to calculate the SDI as a way to compare the diversity between two group of proteins on each amino acid position. Therefore, we determined the ΔSDI, defined as $SDI_{group1}$ minus the $SDI_{group2}$

By the end, the variability difference was determined as a way to compare how conserved or variable were the FliC sequences belonging to each group (**Figure 4)**

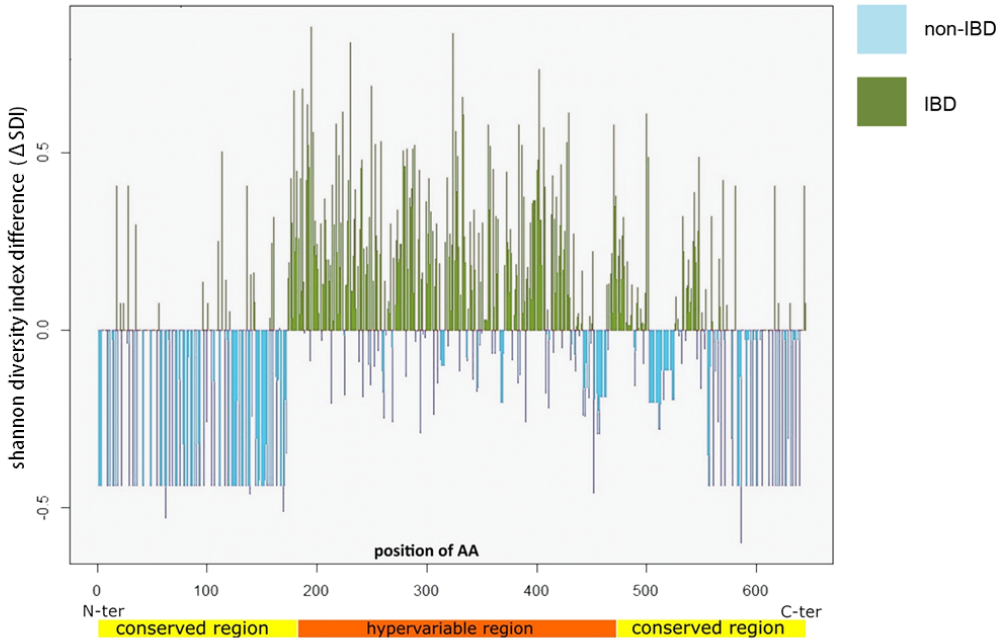



Figure 4. Flagellin structure is different between *E. coli* isolated from non-IBD and IBD individuals. Shannon Diversity Index difference (ΔSDI) was obtained by subtracting the non-IBD Shannon Diversity Index (SDI) from the IBD SDI on each position. ΔSDI >0 means that in that amino acid position, the SDI is higher for IBD (olive) than non-IBD individuals (cyan).

This analysis showed that both conserved regions in non-IBD isolates exhibited an overall higher variability than the ones of IBD isolates, as indicated by ΔSDI <0 (cyan). Interestingly, this tendency was found to be inverted when having a look at the hypervariable region, where the non-IBD isolates displayed a lower intragroup variation compared to IBD, as indicated by ΔSDI >0 (olive).

Taken together, *E. coli* isolates from stool specimen of IBD patients showed a much diverse FliC HVR amino acid sequence compared to non-IBD, whereas isolates from non-IBD donors showed a higher diversity in the conserved region.
The two groups can thus be distinguished according to their FliC diversity profiles. This suggests a correlation between the characteristics of the FliC amino acid sequence and the origin of the *E. coli* strain, i.e., a healthy vs an inflamed intestinal environment. This finding raised the question whether such correlation is causal, that is, whether there is an impact between the FliC primary structure and homeostasis.

### 3.2.2 FliC of *E. coli* strain Nissle 1917 protects against colitis

In order to study the impact of FliC on intestinal homeostasis, we decided to use as a model a known symbiotic bacterium such as EcN, which is a probiotic *E. coli* strain that can be used for maintenance of remission in patients suffering from ulcerative colitis. EcN has been proven as effective as the standard anti-inflammatory agent mesalazine [87]. The protective effect of EcN seems to rely on immune-modulation, antimicrobial activity and bacterium-epithelium crosstalk [88]. Nevertheless, the mechanism of action of EcN remains not completely elucidated.
In order to analyse the impact of the flagella of EcN (henceforth EcN FliC) on the immunoregulatory, anti-inflammatory effect of EcN, we induced acute colitis in specific pathogen free (SPF) wild type C57BL/6 mice by treatment with dextran sodium sulfate (DSS). DSS treatment produces in mice a disease pattern similar to IBD [89] also with respect to the response to treatment [89]. DSS treated mice were administered viable EcN or an isogenic mutant deficient in *fliC* (EcNΔ*fliC*). The inflammation progress was followed for 7 days through the measurement of body weight variation. At the end of the experiment, sections of the colon were stained and inspected for inflammation (**Figure 5**).

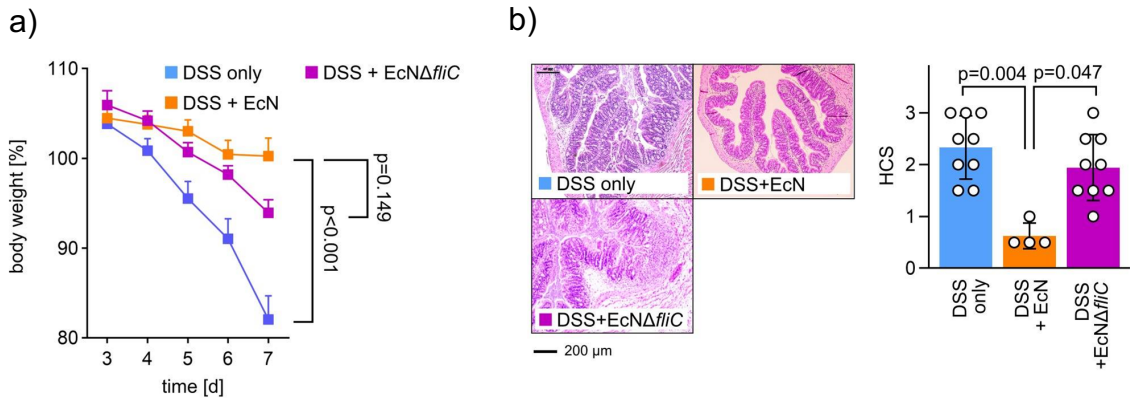



Figure 5. EcN FliC ameliorates DSS-induced colitis in mice. SPF C57BL/6 WT mice were given 3.5% DSS in drinking water at day 0. Two treatments were performed, one group of mice received EcN (DSS + EcN) and another the EcN mutant lacking *fliC* (DSS + EcNΔfliC). The control group received no bacteria (DSS only) **a)** Body weight variation **b)** Stained colon sections (left) and Histoscore (HCS, right), 7 days after start of DSS administration. (**Figures 1 c and 1 d from [1]**)

We observed that in the DSS+EcN treated mice the weight remained more stable than in mice treated with the DSS + EcNΔ*fliC* and the control group. Immunostaining of colonic sections of mice in the different groups and quantification of the histoscore as a proxy for disease severity confirmed that EcN but not EcNΔ*fliC* protected mice against DSS-induced intestinal inflammation. This result shows that FliC contributes to the homeostatic properties of EcN.

In order to test whether the observed protective effect is a general attribute of flagellated *E. coli* or a particularity of the EcN FliC, we administered flagella enriched preparations (FEP) from EcN and two non-symbiotic *E. coli* strains, MG1655 and mpk, to mice under DSS conditions (see **Figure 6**). Of note, FEPs are bacterial extracts devoid of living bacteria.

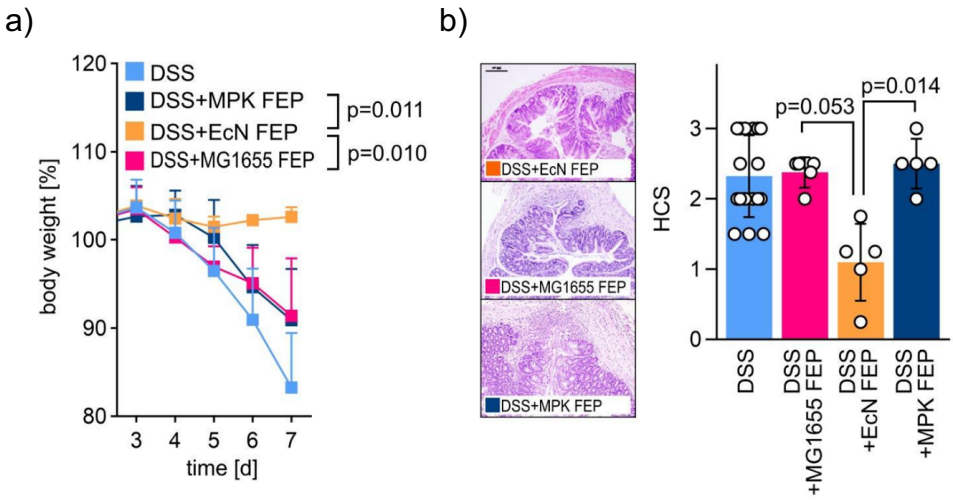



Figure 6. Protective properties of FliC are exclusive of EcN. SPF C57BL/6 WT mice were treated as described in **Figure 5**. Three treatments were performed, the first group of mice was given flagella enriched preparation from EcN (EcN FEP), the second group FEP from *E. coli* mpk (MPK FEP) and the third group FEP from *E. coli* MG1655 (MG1655 FEP). **a)** Body weight variation in the different mice groups. **b)** Stained colon sections (left) and Histoscore (right). (**Figures 2 c and 2 d from [1]**)

We observed that neither *E. coli* MG1655 nor *E. coli* mpk FEPs were able to significantly protect mice against intestinal inflammation. Hence, the protective effect turned out to be specific for EcN. This observation hints to the presence of particular characteristics of EcN FliC.

Since FliC triggers toll-like receptor 5 (TLR5) signalling, we asked whether the observed protective effects were mediated through TLR5. For this, human embryonic kidney cells overexpressing the murine TLR5 (mTLR5-HEK293) were used as a reporter system. When sensing FliC, TLR5 triggers the secretion of IL-8, which can be measured as readout through the enzyme linked immunosorbent assay (ELISA). For this, FEPs extracts from EcN, *E. coli* MG1655 and *E. coli* mpk were used as stimuli for mTLR5-HEK293 cells (**Figure 7 a**). Secondly, as a way to evaluate the impact of EcN FliC alone, TLR5 activation was tested with wild type EcN and EcN lacking *fliC* (**Figure 7 b**).

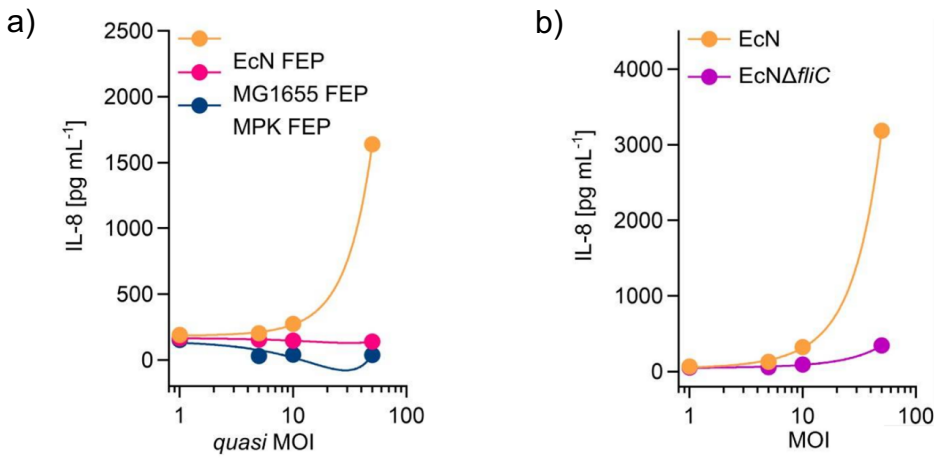



Figure 7. EcN FliC induces stronger TLR5 activation in comparison to non-symbiotic *E. coli*. mTLR5-HEK293 cells were stimulated during 24h using different moi, the IL-8 signal was measured through ELISA. **a)** FEPs produced from different moi (quasi moi) from EcN (EcN FEP) and the non-symbiotic *E. coli* strains MG1655 (MG1655 FEP) and mpk (MPK FEP) were used as stimuli**. b)** Viable EcN (EcN) and EcN lacking *fliC* (EcN Δ*fliC*) were used as stimuli. (**Figures 2 e and 2 f from [1]**)

We observed higher secretion of IL-8 in EcN FEP-stimulated mTLR5-HEK293 than in cells stimulated with non-symbiotic strains. Moreover, the secreted levels of IL-8 were dose dependent. When mTLR5-HEK293 were stimulated with live bacteria, we barely observed any secretion of IL-8 in response to EcN Δ*fliC* as compared to wild type EcN. Taken together, these results point out that non-symbiotic *E. coli* FEPs are unable to activate TLR5 and that the activation of TLR5 by living EcN is partly dependent on its own FliC.

In order to explore whether the distinctive effect of EcN FliC was influenced by unique features of EcN genetic background, two different recombinant bacteria were generated. The first was *E. coli* MG1655 lacking its own *fliC* and complemented with EcN *fliC* (MG1655Δ*fliC*::*fliC* EcN). The second was EcN lacking its own *fliC* and complemented with the *E. coli* mpk *fliC* (EcNΔ*fliC*::*fliC* MPK). The correct assembly of the flagellum in both mutants was checked through electron microscopy (see section 6.1.1, S1 Figure from [1]). The induction of colitis was performed as described before (see **Figure 5**) and mice were administered with the generated recombinant strains (**Figure 8**).

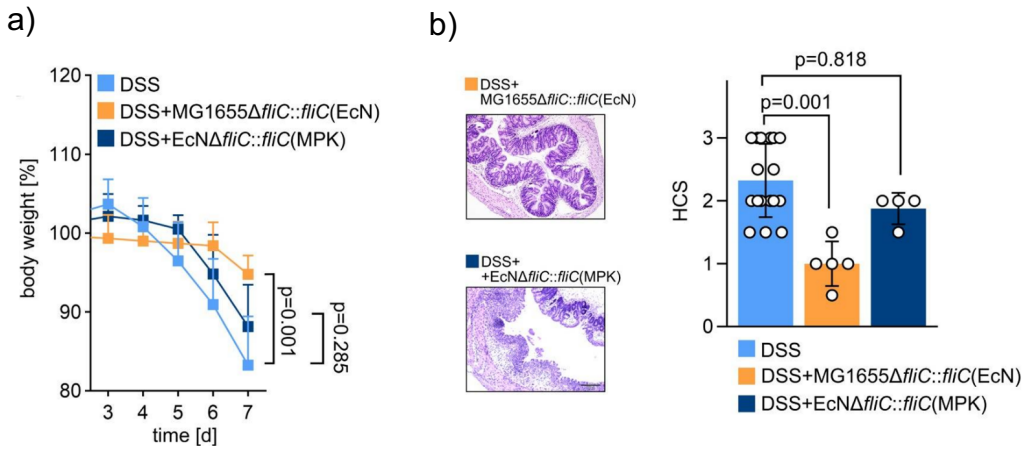



Figure 8. EcN FliC protective properties are independent of the bacterium genetic background. DSS treatment was performed as before. **a)** Body weight variation **b)** Stained colon sections (left) and Histoscore (right). (**Figures 2 g and 2 h from [1]**)

When administered with MG1655Δ*fliC*::*fliC* EcN, DSS treated mice showed a mild weight loss along a seemly undamaged colonic tissue compared to mice treated with EcN carrying the MPK FliC.

Overall, we observed that the heterologous expression of FliC from a symbiont into a non-symbiotic strain conferred protective properties to the latter. These results reaffirm the protective effect of EcN FliC and distinguish it from the contribution of the EcN genetic background.

In line with this, mice deficient in TLR5 developed severe colitis upon DSS treatment even during administration of EcN, indicating that the protective effect is TLR5-mediated and, thus, TLR5-ligand dependent (**Figure 5 c from [1]**).

### 3.2.3 Generation of high purity EcN FliC.

In order to further clarify the effect of the isolated protein alone, we decided to recombinantly express and purify ultra-pure EcN FliC, henceforth named rFliC (EcN). With this aim a collaboration with the group of Prof. Dr. Roman Jerala (National institute of Chemistry, Slovenia) was started, given their experience in purifying chimeric flagellin [90][73]. In Tübingen, we cloned EcN *fliC* via Gibson assembly into the IPTG inducible expression vector pET-19b, and transformed the newly generated vector into *E. coli* NiCo21(DE3) cells, which are engineered for protein expression. Altogether, we aimed to produce rFliC (EcN) using a polyhistidine-tag system [91] on its N-terminus in order to purify the protein via immobilized-metal affinity chromatography (IMAC). This technique exploits the ability of the poly-His tag to bind to nickel sepharose resins [91].

IMAC can be performed under different conditions in order to find the best compromise between purity, time and speed. We first used self-assembled gravity flow columns and compared them with a commercial prepacked column connected to a peristaltic pump (**Figure 9**).

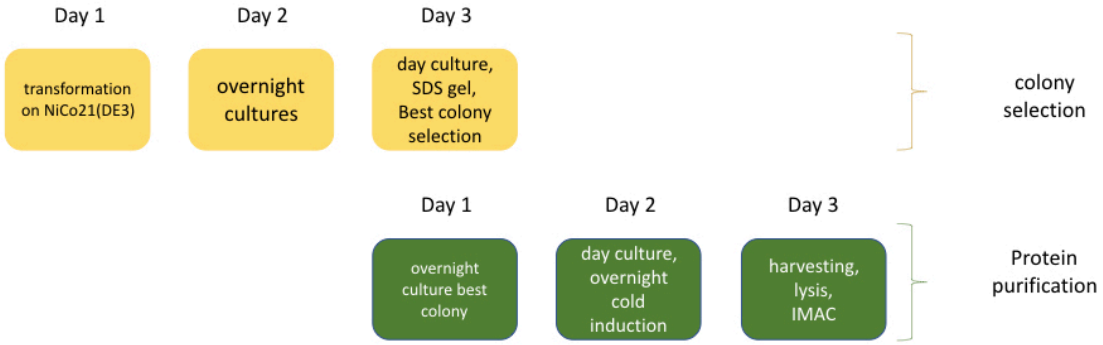



Figure 9. Chronology of the best colony selection and protein purification.

### 3.2.4 Purification standardization.

The first step (as performed by our collaborators in Slovenia) consisted in the isolation of an *E. coli* colony that produced the highest amount of protein, for this reason several single colonies were isolated from the same transformation mixture (**Figure 10 a**). The gravity flow IMAC is an inexpensive technique that allows to purify multiple samples at the same time, since its operation is not dependent from an external pump. The column was manually assembled, and filled with 1 ml of nickel resin or so-called slurry. The purification process is passive, as the pressure exerted by the weight of the sample allows the contact of the protein with the nickel-charged resin, and in turn capillarity pulls the entire sample through the column.

IMAC through prepacked column is more expensive and dependent on an external mechanical pump, yet at the same time advantageous due to the fact that the pressure applied from the pump into the column typically increases the speed of the process. An additional advantage of the prepacked column is that, in case of the need for scaling up the purification, it can be easily plugged to a fast protein liquid chromatography (FPLC) system, which makes the purification process fairly automatic once it is standardized.

The purity of the isolated protein was primarily evaluated using sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie blue. The SDS gel was loaded with different aliquots collected along purification process (**Figure 10 b and 10 c**).

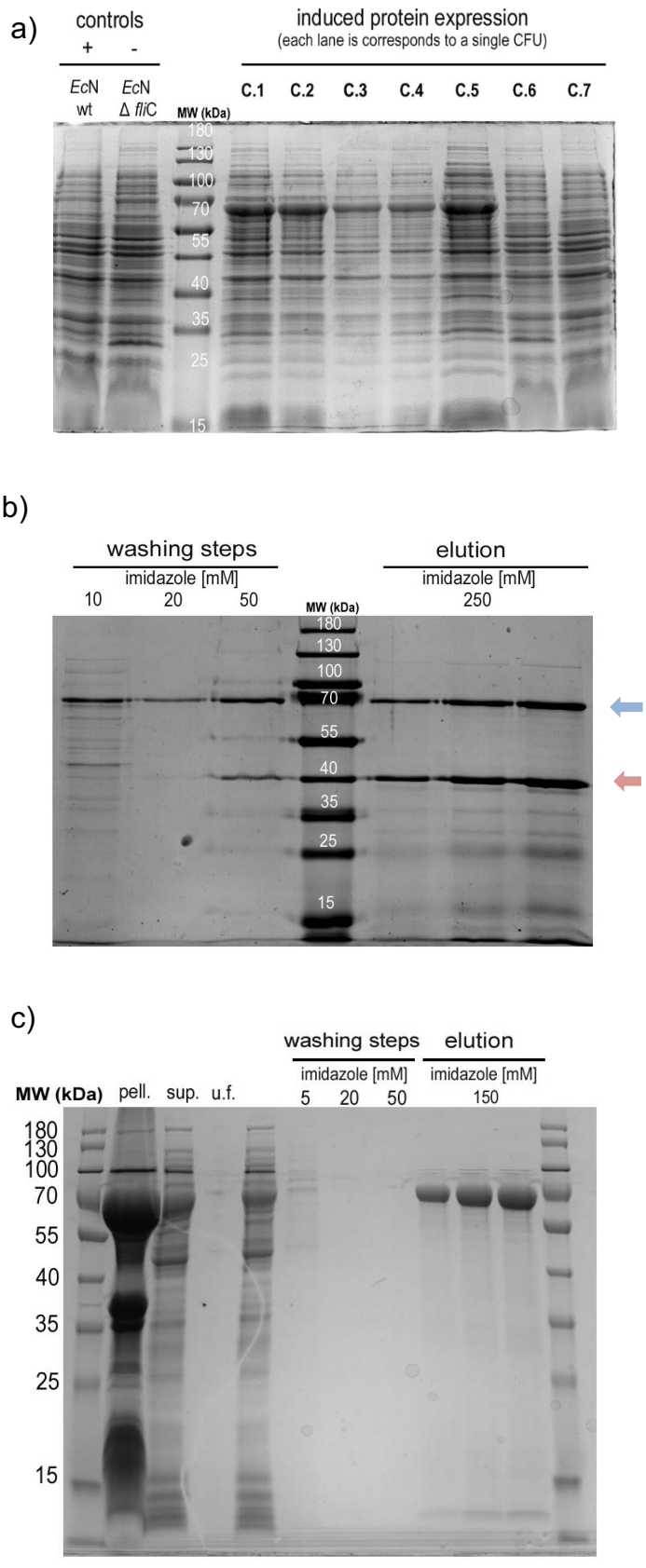



Figure 10. Screening of r*fliC* (EcN) expressing *E. coli* colonies and protein purification via IMAC performed with two different techniques:

**a)** Best producing colony selection. Each lane of an SDS gel (C.1 to C.7) was loaded with an aliquot of a single, IPTG-induced and lysed *E. coli* colony carrying the expression vector encoding for r*fliC* (EcN). As controls, wild type EcN and EcN lacking the flagellin gene (EcN Δ*fliC*) were used. The fifth colony (C.5) was selected for the following step.

**b**) IMAC gravity flow : from left to right, the bacterial lysate was washed in three steps (10, 20 and 50 mM Imidazole). Two proteins were eluted with 250 mM imidazole, one of the expected molecular size of 65 kDa (blue arrow) and another one of a size of 40 kDa (red arrow).

**c)** IMAC prepacked column : Collected fractions during IMAC are shown in the SDS gel left to right. Bacterial lysate as pellet (pell.) and supernatant (sup.) was loaded, next to it, lysate that did not bind to the column or unbound fraction (u.f). Loaded were also fractions collected from the column after washing with 5, 20 and 50 mM imidazole and finally three fractions collected during the elution step with 150 mM imidazole.

Prior to the gravity flow column purification, seven *E. coli* colonies carrying the rFliC encoding-pET-19b vector were isolated and screened for expression of the recombinant protein. The fifth colony was selected for rFliC purification as it expressed the highest amount of protein of the expected molecular size (65 kDa) as assessed by SDS gel analysis (**Figure 10 a**, C.5 lane). EcN wild type did not show a noticeable band compared with the negative control or the other induced colonies, most likely because the endogenous *fliC* is expressed at low levels, unlike the episomal *fliC* gene, which is expressed under the T7 bacterial promoter. Overall, it seemed that although all the cultures were originated from the same transformation mixture, the isolated colonies responded differently to the protein transcription induction. The reason for this might rely on differences in the recovery of *E. coli* from the freezing/thawing process [92].

During the IMAC itself, in both techniques (**Figure 10 b and c**) different fractions were collected to follow the course of the purification according the respective protocols. As expected, the protein was in both the sample for the gravity flow protocol and prepacked column (**Figure 10 b and c**). Analysis of the unbound fraction confirmed that rFliC (EcN) strongly adhered to the nickel resin.

For the gravity flow purification, some of rFliC (EcN) (65 kDa) was lost during the washing steps, mostly with 10 and 50mM imidazole (**Figure 10 b**). During the final elution with 250 mM imidazole, an additional compound of 40 kDa was coeluted along the protein of interest (65 kDa). The nature of this band was not examined but since the elution step was performed at room temperature, the protein might have been partially degraded during the purification process. Nevertheless, the protein concentration was calculated through Bradford assay to be 3,6 mg/ml, but this included also the coeluted compound. During the purification on prepacked column it was observed in the SDS gel (**Figure 10 c**) that washing with increasing concentrations of (5 mM, 20 mM and 50 mM) removed more unspecific elements without noticeable amount of protein in comparison with the elution fraction with 150 mM imidazole. No band at 40 kDa was observed in this case.

Overall, employing gravity flow as a purification technique was a process that presented two main disadvantages: the expected length of the purification and additionally the occurrence of clots of bacterial lysate in the column, which made the whole process even slower.

As observed by SDS gel analysis, barely any contaminant was observed during the purification on prepacked column and, by the end of the process, rFliC (EcN) at a final concentration of 1,9 mg/ml was isolated. For this reason, we went ahead employing the protein eluted from the prepacked column.

To validate the purity of the protein eluted from the prepacked column we analyzed whether the eluate contained lipopolysaccharide (LPS), which is a usual but undesired contaminant of protein preparations. With this aim, LPS concentration was determined through the Limulus amoebocyte lysate (LAL) assay. The assay is based on the sensibility of amoebocytes from the blood of marine arthropod horseshoe crab to LPS [93]. The activation of the amoebocyte pro-clotting enzymes in presence of LPS is measured through a chromogenic substrate (**Figure 11**).

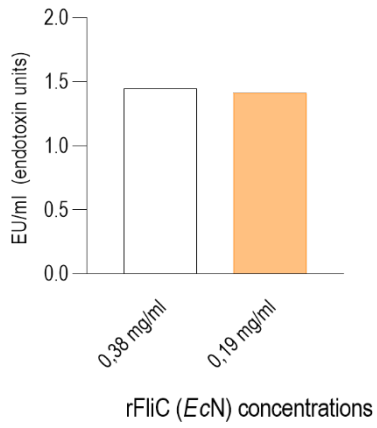



Figure 11 Endotoxin determination of rFliC (EcN) isolated using a prepacked column. The dialyzed and filtrated protein was diluted on a ratio 1 to 5 and 1 to 10 and endotoxin units per ml (EU/ml) were measured through the Limulus amoebocyte lysate test. The 1:10 dilution (orange) presented 2,2% less endotoxin that 1:5 (blank column).

Two different dilutions of the protein were assayed in order to rule out any artefact generated by an inhibitory LPS concentration. In this situation a typically amount of endotoxin measurement is lower than 0,5 EU/ml.

The test indicates presence of LPS below 1,5 EU/ml, which is approximately three times lower than the amount generally present in a crude preparation such as flagella enriched preparations. Although the measurements were performed using increasing dilutions of the protein, we measured a similar amount of LPS.

Finally, mass spectrometry (MS) analysis confirmed the predicted amino acid sequence of rFliC (EcN) (Performed by Dr Karsten Boldt. Eye clinic, Tübingen). This verification is of paramount importance given the fact that the recombinant protein will also be used for protein crystallization (**Figure 12**)

Figure 12 Mass spectrometry result from rFliC (EcN). The figure shows the amino acid sequence of rFliC (EcN). Yellow color means that the expected residue was identified. Green means that modification in the amino acid were detected. White signals the amino acid sequence of the histidine tag and the cleavage site (Dr Karsten Boldt, Eye Clinic, Tübingen)

The MS results presented a 96% coverage. The detected modifications corresponded to deamination for asparagine (N) and glutamine (Q) and to oxidation for methionine (M). All of these modifications are expected due to sample processing.

Overall, through Gibson assembly it was possible to quickly clone *fliC* into an expression system. The use of a prepacked column offered a clean and reproducible method for rFliC (EcN) purification.

### 3.2.5 rFliC (EcN) effectively protected mice against DSS induced colitis

Once obtained a standardized method for producing ultra-pure flagellin, the next step was to study its protective effect. With this aim, DSS-treated mice were administered either rFliC (EcN) or commercial flagellin from the non-symbiotic *E. coli* MG1655 as a control (**Figure 13**).

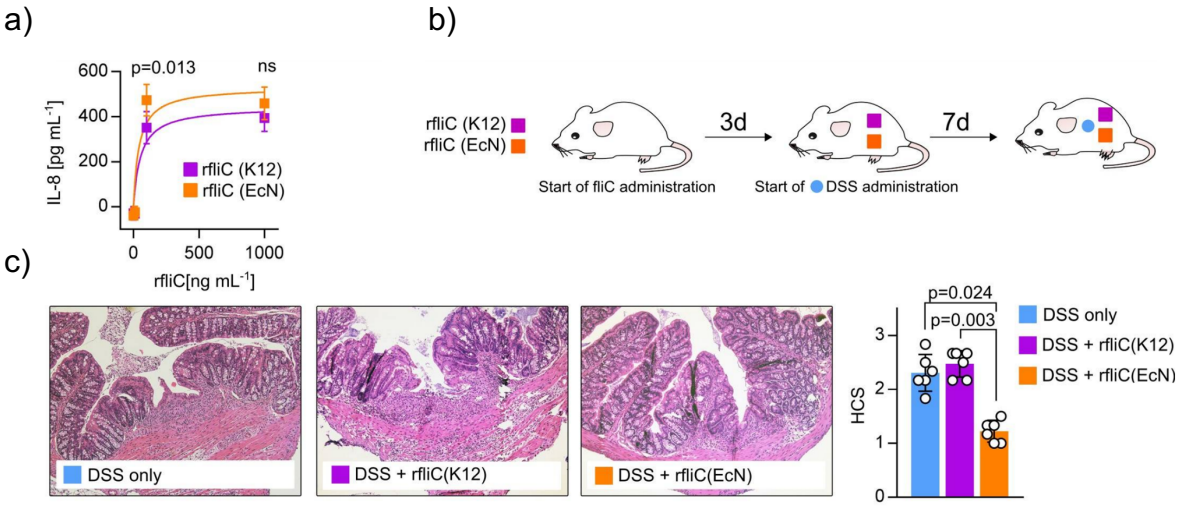



Figure 13. rfliC (EcN) confers protection to DSS-treated mice. **a)** mTLR5-HEK293 cells were stimulated during 24 with rfliC (EcN) or rfliC (K12**) b)** SPF C57BL/6 WT mice were given intragastral rfliC (EcN) or rfliC (K12), at day three the mice received 3.5% DSS in drinking water. Mice were monitored until day 7 **c)**. Stained colon sections (left) and Histoscore (right) 7 days after DSS administration. (**Figures 6 a, 6 b and 6 c from [1]**)

Both rFliC (EcN) and rFliC (K12) were able to induce IL-8 secretion when administered to mTLR5-HEK293 cells, which validated the immunogenic activity of rFliC (EcN).

The *in vivo* experiment showed that the colonic surface of rFliC (EcN)-administered mice presented significative less inflammation compared to the control condition and to the administration of the flagellin from MG1655.

This means that rFliC (EcN) protected mice against DSS induced colitis and that this effect was not exerted by the non-symbiotic MG1655.

Overall, this confirmed that EcN FliC mediates protection, hinting that EcN FliC may possess distinctive structural elements not present in non-symbiotic strains.

In order to test the occurrence of such structural differences, MSA was performed using the amino acid sequences of FliC EcN, FliC MG1655 and FliC mpk (**see Figure**

**3C from [1]).** A schematic representation was built to highlight the degree of similarity among the sequences (**Figure 14**).

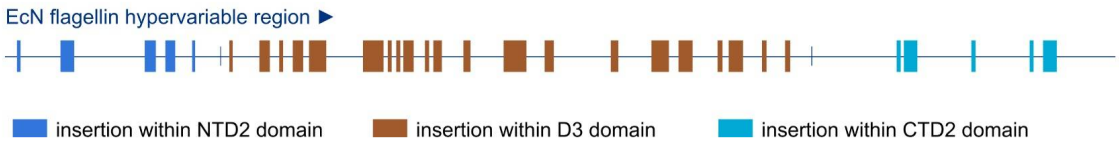



Figure 14. EcN FliC hypervariable region present numerous insertions. Schematic representation of insertions present on EcN FliC HVR and absent on MG1655 and mpk. (**Figure 3 d from [1]**)

Most of the dissimilarity was observed in the hypervariable region (HVR). Interestingly, the differences found mostly refer to the presence of short amino acid sequences in EcN FliC that are absent in the HVR of FliC from non-symbiotic strains.

These differences could be explained by the presence of numerous insertions of amino acids in the HVR of EcN, making the sequence substantially longer compared to the HVR amino acid sequences of FliC from other commensals.

This finding prompted us to determine whether the HVR significantly contributed to the observed protective effect. For this, we generated an EcN strain that expressed a mutant version of *fliC* lacking a vast part of the HVR (EcN *fliC*ΔHVR). In the DSS mouse model of gut inflammation EcN *fliC*ΔHVR did not confer any protection to mice, hence revealing that the HVR is indispensable for the protective effect of FliC (**Figure 4 c from** [1] ) Taken together, this analysis demonstrated that the most of the symbiotic properties of EcN reside in the HVR.

### 3.2.6 Purification and crystal structure of *E. coli* Nissle flagellin

In the next step, we addressed the question whether the detected difference in the amino acid sequence between EcN flagellin and the flagellin from *E. coli* K12 and *E. coli* mpk [1] had an impact on the structure of the molecule. Therefore, we started a cooperation with Prof. Thilo Stehle and Michael Braun from the interdisciplinary Institute for Biochemistry in Tübingen to generate EcN FliC crystals and in silico models of the protein structure.

Our collaborators analyzed the molecular mass of our rFliC (EcN) and its proper folding and set different crystal conditions. However, no crystals for rFliC were obtained. For

this reason, a new strategy was chosen: our collaborators scaled up the amount of protein produced and we generated different new constructs.

The generation of crystals is a stochastic process; hence, in order to maximize the chances of obtaining crystals, we generated new rFliC (EcN) variants containing deletions in different parts. Overall, the approach consisted in studying parts of the protein in order to build the best possible model of the wild type FliC EcN structure.

For this, I generated different mutants through enzymatic assembly of DNA (**Figure 15**).

In different organisms the conserved region has been described to present a structure of the type α-helix that is fairly conserved among all types of FliC. At the same time the presence of such α-helices, such as the D0 and D1 domain, tend to generate aggregates [94]. For this reason, we aimed to generate the construct FliC ΔD0D1, which lacks the complete conserved region of the protein, and the FliC ΔD0 construct which lacks the D0 domain only.

In order to study in-vitro the effect of the conserved regions alone, a variant lacking only the hypervariable region was generated (FliC ΔHVR).

Since the D1 domain contains a region that has been shown to be essential to TLR5 activation [95], we aimed to study the structure of this region generating a construct expressing only this domain (FliC ΔD0 ΔHVR). Finally, in order to analyse the contribution of a previously not described domain on EcN FliC, two more mutants lacking this domain were built, namely, FliC ΔD4 (EcN) and FliC ΔD4 long (EcN). Of note, a D4 domain was recently described also for enterohemorrhagic *E. coli* O157:H7 (EHEC) and enteropathogenic *E. coli* O127:H6 (EPEC) [96].

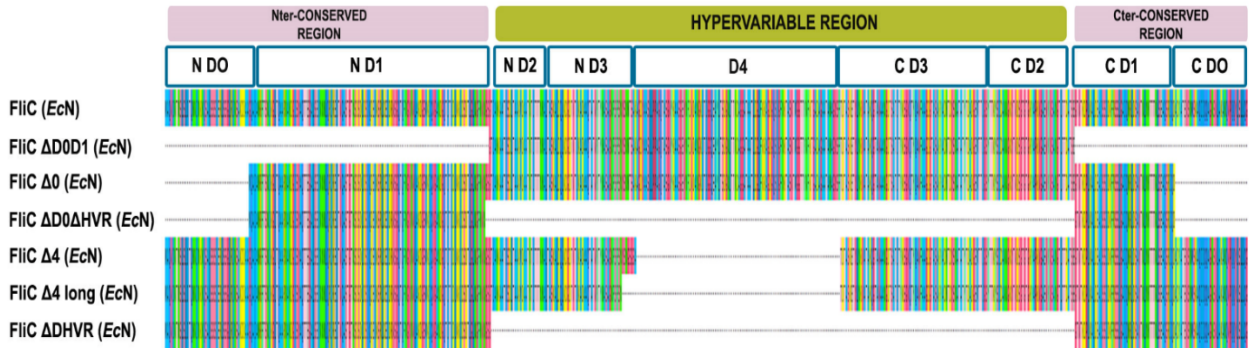


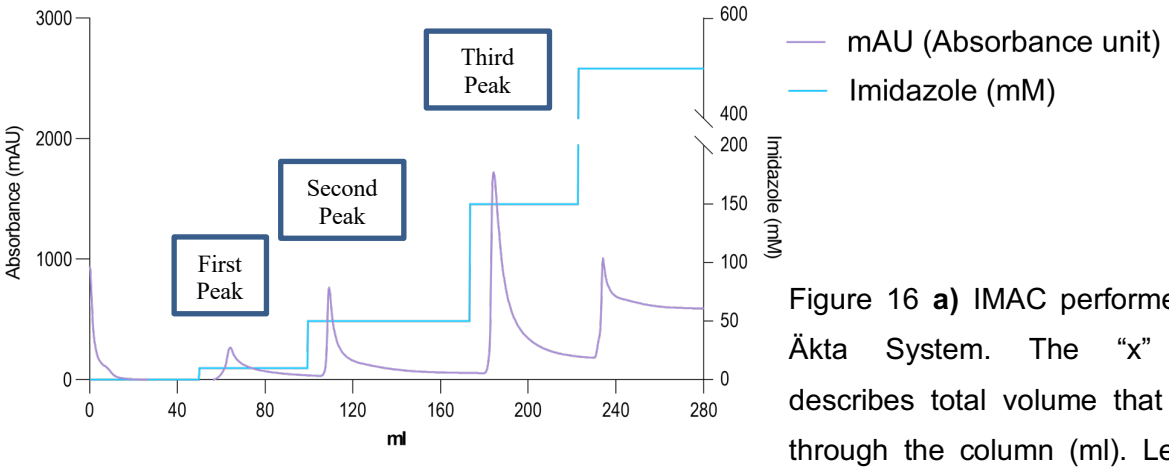

Figure 15. Schematic representation of rFliC (EcN) and its truncated versions. From top to bottom: full length **FliC (EcN)**, which has a length of 595 amino acids. **FliC ΔD0ΔD1 (EcN):** N-terminal deletion between M(1) and N(176) and C-terminal from G(501) to the end of the C-terminus. **FliC ΔD0 (EcN)**: N-terminal deletion between M(1) and Q(48) and C-terminal deletion from G(556) to the end of the C-terminus. **FliC ΔD0ΔHVR (EcN)** : N-terminal deletion between M(1) and Q(48) and C-terminal from G(556) to the end of the C-terminus. Besides a deleted region between G(179) and G(501). **FliC ΔD4 (EcN)**: deletion between N(254) and S(376**). FliC ΔD4 long (EcN)**: deletion between N(251) and T(372) and **FliC ΔHVR (EcN)** which contains a deletion between G(179) and A(501).

Our collaborators optimized the expression of the different FliC variants for crystallization and in parallel, we optimized the purification process to produce simultaneously the different FliC variants at the amount and purity required for in vitro experiments.

### 3.2.7 Purification of rFliC (EcN) performed with prepacked column on a FPLC system increases the speed and yield of the protein purification.

For the parallel purification of different FliC variants, we decided to use a semi-automated purification system in order to increase speed and reproducibility of the purification process. With this aim a fast protein liquid chromatography (FPLC) system was used. This FPLC system (Äkta system from Cytiva, before GE healthcare) allowed the use of the same prepacked columns used for affinity chromatography. The system uses a pump and a mixer that automatically set the desired imidazole concentration. It also contains an ultra violet (UV) light detection system of absorbance at 280 nm, this allows continuous monitoring of compounds carrying aromatic residues such as tyrosine and tryptophan. The first construct to be purified was rFliC (EcN) (**Figure 16**).

a)




b)

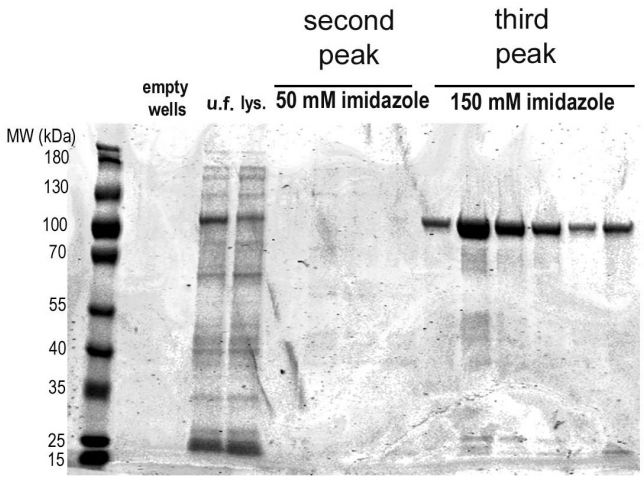



Figure 16 **a)** IMAC performed on Äkta System. The "x" axis describes total volume that went through the column (ml). Left "y" axis shows Absorbance units (mAu). Each peak reflects high amount of bacterial lysate eluted from the column. The right "y" axis refers to imidazole concertation (cyan), which removes impurities and ultimately elutes the His-tagged protein on the third peak**; b)** SDS gel after IMAC. Fractions collected from the second and third peak were collected and loaded on the gel, the first peak (washing step) was not collected. The third peak showed an elution of the expected molecular size (65 kDa).

The IMAC proceeded for a total of 280 ml. This means 280 times the volume of the prepacked column or “column volumes” (CV). Each washing and elution step was performed on an average of 80 CV, This is considerable higher than the previous IMAC approaches.

Each peak on the UV profile correlates with the respective imidazole concentration. When the peak is reached, the signal flattens, meaning that components susceptible to be removed with that imidazole concentration were successfully removed. After each imidazole wash and elution, the UV baseline detection increases. This is due to the fact that imidazole absorbs UV radiation as well.

SDS gel analysis of the IMAC eluate showed that the full-length protein was eluted with 150 mM imidazole. The elution fractions loaded on the gel showed a band thickness that correlated to the shape of the peak, hence validating that the peak apex contained the highest amount of protein.

The whole eluate was pooled together, dialyzed and concentrated. Finally, the protein concentration was determined and compared to the amount of protein eluted with the previous IMAC settings (**Figure 17**).

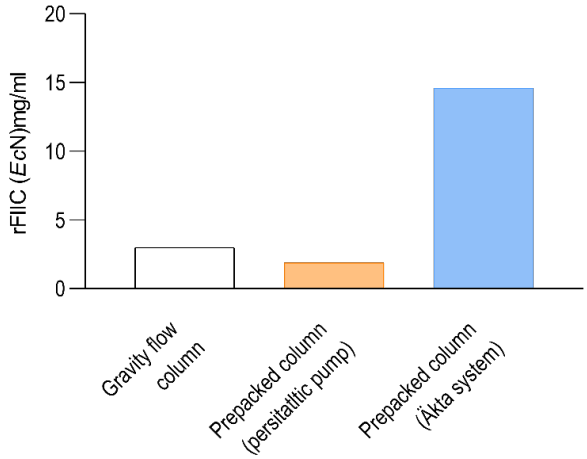



Figure 17. Protein yield. The FLPC system presented the highest protein yield after IMAC. Gravity flow presented a yield of 3 mg/ml compared with the 1,9 mg/ml of the prepacked column connected to the peristaltic pump. When the later column was plugged into the Äkta system the yield increased to 16 mg/ml.

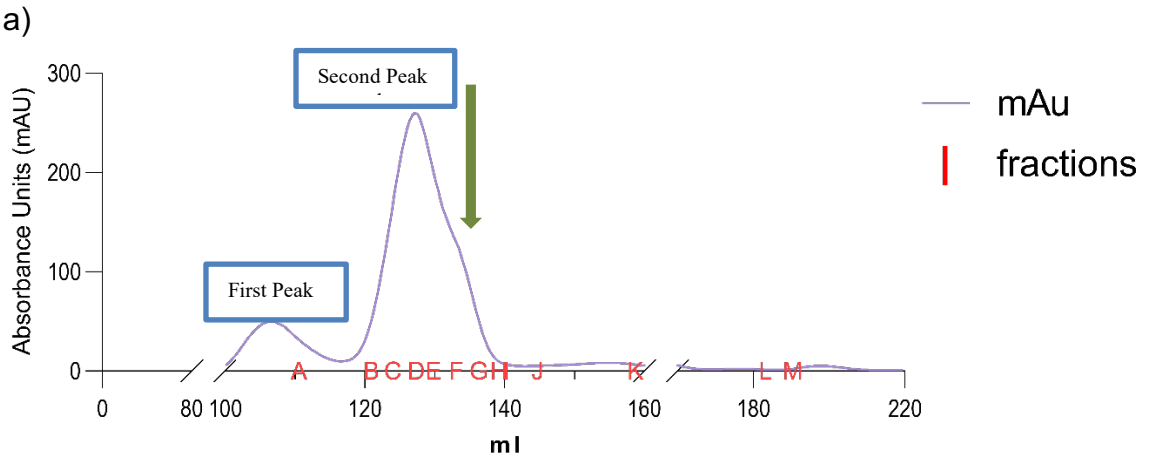
The total amount of protein recovered was 16 mg from a starting culture of 1 L. This is almost 8 times higher than what was recovered with the peristaltic pump. One of the main reasons for this difference is that during FPLC all fractions above the UV baseline

were pooled together, while for the eluate obtained with the peristaltic pump pooling was done only for those samples clearly visible on the gel.

Overall, FPLC allowed a thoroughly wash out of undesired compounds which is reflected on a clear band on the SDS gel. The UV peak apex during the elution step is a reliable sign of a high concentration of the desired protein. In vitro experiments require less amount of protein compared to crystallization, meaning that the selection of the fraction corresponding to the apex and the immediate surrounding is enough, hence saving time collecting all the fractions and avoiding the risk of protein precipitation during the concentration step.

Due to the fact that each FliC mutant variant has a distinctive molecular size, it was decided to include size exclusion chromatography (SEC) following the IMAC. SEC increases further the purity of the protein and importantly, as the separation is based on the molecular weight, it allows calculating the molarity of the isolated rFliC (EcN) mutant variants and species. In this context a species means that the isolated protein can come in the form of a monomer, dimer or higher oligomers.

Determination of the molarity of the isolated species enabled performing *in vitro* experiments with the same amount of protein for all the different variants taking into account their different mass. The first construct to be purified through SEC was rFliC (EcN) (**Figure 18**)



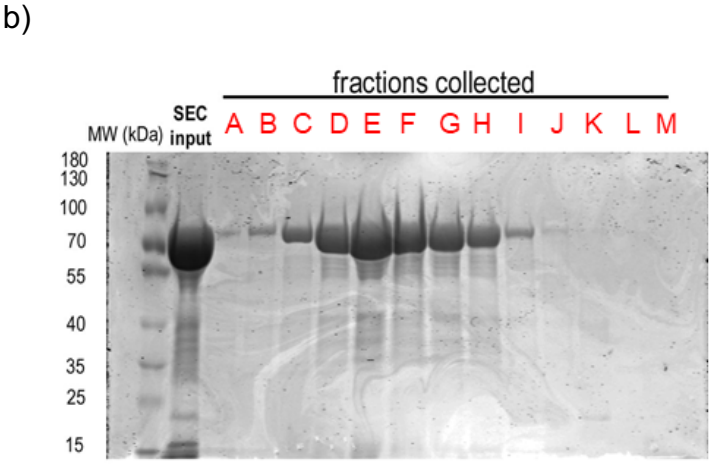




Figure 18. rFliC (EcN) SEC on Äkta System**.** Following IMAC; the protein was further purified trough separation by molecular size. **a)** UV profile: Each peak relates to UV absorbance (same as **Figure 16**), the first peak corresponds to protein aggregates, meanwhile the second refers to oligomers. **b)** SDS gel after SEC. The late fractions (lanes F and G) were collected.

The SEC run for rFliC (EcN) presented two most noticeable peaks, the first (**Figure 18**) corresponding to the species that, given their big size, were not retained by the column, and the second one that fitted with the separation range of the calibrated column (see methods) for a molecular size of rFliC (EcN) oligomers.

The green arrow in **Figure 18** points to a protrusion in the UV profile, which typically indicates that smaller species are being eluted together with oligomeric forms. Higher size oligomeric forms such as trimers or tetramers are less accurate to calculate due to the fact that they are distributed on the total area of the second peak. In contrast, the late elution pointed to in **Figure 18** has a narrower profile, hence providing a better accuracy, in this case for the monomeric form of rFliC (EcN)

Once obtained high purity rFliC (EcN), we studied the induction of innate immune response in vitro. For this, mTLR5-HEK293 cells were employed. TLR5 is the transmembrane receptor for flagellin, when intestinal epithelial cells are stimulated with

flagellin, they secrete in turn interleukin-8 [97]. On this in-vitro assay, the secretion of IL-8 was measured through ELISA. As a control for detection of LPS activity, HEK293 cells overexpressing TLR4 were used. The TLR4 detects LPS and activates a downstream cascade that ultimately turns on the expression of IL-8 (**Figure 19**).

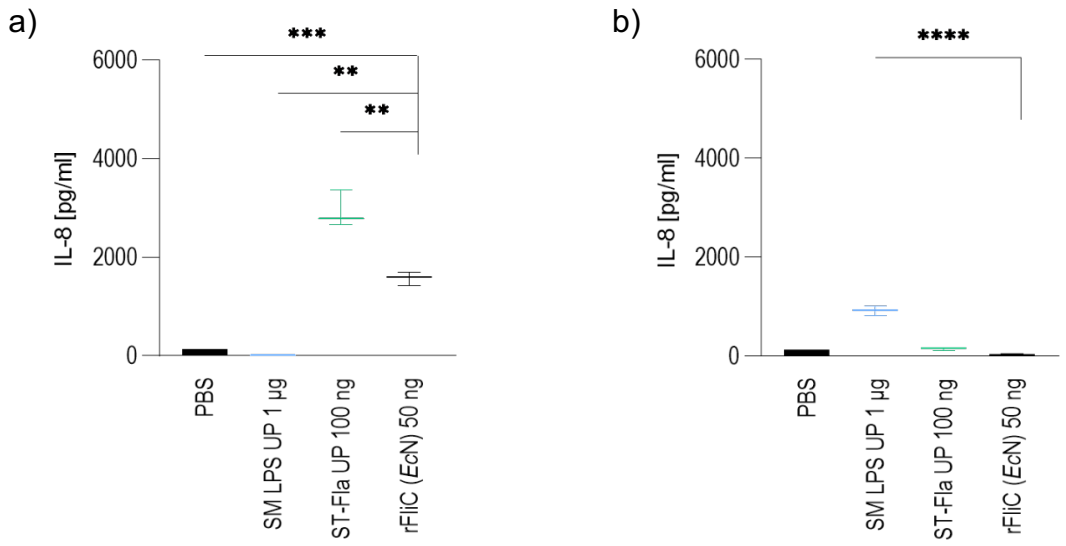



Figure 19 In vitro determination of immunogenic activity of rfliC (EcN). **a)** mTLR5-HEK293 and **b)** mTLR4-HEK293 cells were simulated with PBS (as assay control), *Salmonella* Minessota LPS (SM LPS 1), *Salmonella* Typhimurium Flagellin (Fla) and rFliC (EcN). Each condition was assayed in triplicates and the collection of supernatant and ELISA measurements were made after 24 hrs.

As shown in **Figure 19 a),** when the stimulus used was rFliC (EcN), the IL-8 signal was significantly higher than when PBS and the negative control (*Salmonella* Minessota LPS, SM LPS 1) were used as stimuli, and lower than when the positive control was used (*Salmonella* Typhimurium Flagellin, Fla). Instead, as shown in **Figure 19 b),** there was not statistical difference between rFliC (EcN) and PBS, but a significant difference with the positive control (ST-Fla UP)

rFliC (EcN) isolated from the prepacked column IMAC connected to the FPLC system was able to trigger a TLR5 response, while no signal was observed when the protein was used to stimulate mTLR4-HEK293, once again confirming the purity of rFliC (EcN). The same SEC conditions were used for the other rFliC (EcN) mutant variants (**Figure 15**). Purification of variants with deletions inside the hypervariable region of rFliC proved to be more challenging due to the generation of aggregates. This meant that

less protein was recovered and that each purification had to be followed by a thorough cleaning of the column using desaturating agents. One possible explanation for this problem is that without the HVR, the FliC monomers have a higher degree of freedom to interact which each other, making more likely a higher rate of polymerization and eventually aggregation.

Eventually, all the FliC variants mentioned above were produced through FPLC, and the molarity of the purified species was determined. The use of these variants enabled us to assess the immunogenic contribution from different domains of EcN FliC.

### 3.2.8 HVR-domains play a role in immunogenicity of FliC EcN

The immunogenic properties of the deleted domains generated were evaluated through stimulation of mTLR5-HEK293 cells with equimolar amounts of the purified EcN FliC variants (**Figure 20**).

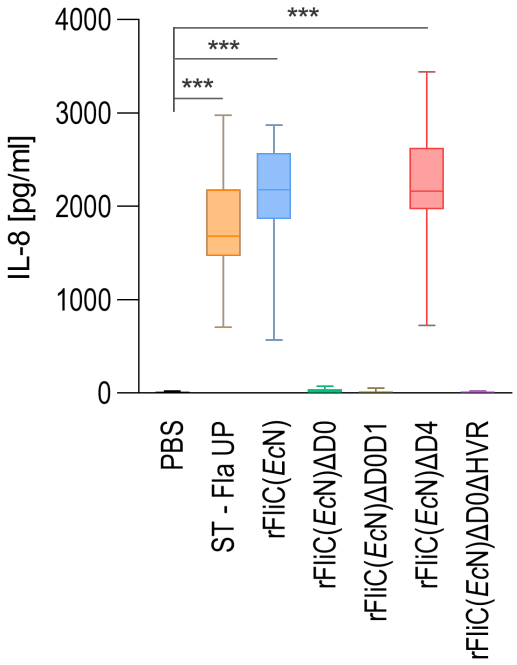



Figure 20. The central region of the HVR is not essential for TLR5 activation. mTLR5-HEK293 cells were stimulated during 24h using equimolar amounts of each of the EcN FliC mutants generated.

Mutants carrying deletions over the D0 domain failed to induce secretion of IL-8. The deletion of the D4 domain did not abrogate TLR5 activation.

### 3.2.9 FliC EcN harbors a D4 domain in its HVR

Meanwhile our collaborators, Prof. Thilo Stehle and Michael Braun (IFIB, Tübingen), were able to solve the structure of the FliC ΔD0D1 mutant protein as well as that of a FliC variant containing only the D1 and D2 domain. These two structures were superposed to model EcN FliC containing the D1 domain and the HVR (**Figure 21**).

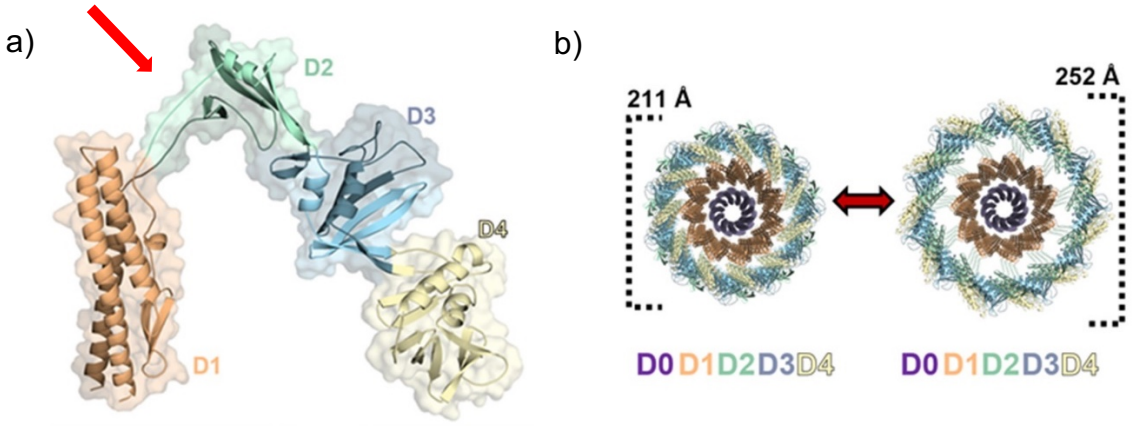



Figure 21. EcN FliC HVR harbours an additional domain. **a)** crystal structure of the D1 (orange), D2 (green), D3 (blue) and D4 (yellow) domains. The HVR is connected to the D1 domain of the conserved region through a linker region indicated with a red arrow. **b)** In silico models of EcN FliC flagella. Shown is the axial view of two possible conformations of the flagellar filament according to the angle between the conserved Domains D0/D1 and the hypervariable region. Compact conformation of the flagella (left panel) shows a diameter of 211 Ångström (Å) and the elongated conformation (right panel) with a diameter of 252 Å. (Dr Michael B. Braun and Prof Thilo Stehle, IFIB, Tübingen)

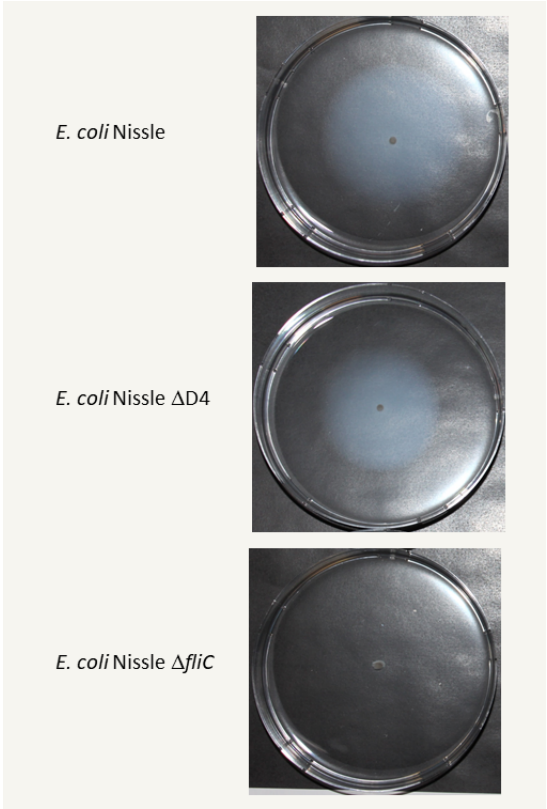
The solved crystal structure of the EcN FliC shows an HVR with three globular domains: D2, D3 and D4. The D1 domain is connected to the HVR through a linker from N176 to A182 and from A182 to T497. The structure of the D4 domain was recently solved via cryo-EM for the flagellar type H7 and H6 of pathogenic *E. coli* [96]. This is the first time that the FliC structure from a symbiotic *E. coli* is described. The predicted model of the flagellum was built using the published structures of the assembled conserved region of *Bacillus subtilis* [98]. According to this prediction, EcN flagellum may be thicker when the linker is elongated or thinner when the linker is compacted.

Since upon polymerization FliC monomers are stacked upon each other with the HVR domain facing outwards, we asked the question whether the HVR plays a role in stabilizing the whole flagellum and particularly what the contribution exerted by the D4 domain to such stability might be.

### 3.2.10 D4 domain does not influence the function of the EcN flagellum

In order to evaluate the contribution of the D4 domain to bacterial motility, an EcN strain expressing a FliC variant lacking the D4 domain was generated. As a control an EcN strain lacking *fliC* was generated and a swarming assay was performed (**Figure 22**).

a)




b)

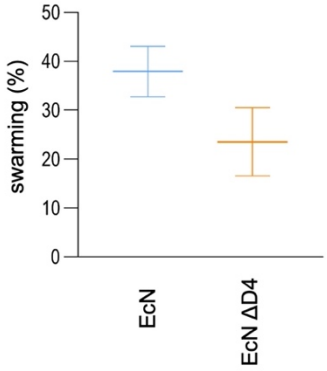



Figure 22. Role of the D4 domain in flagellum motility. **a)** To test for the ability to swarm, which reflects locomotion of the respective EcN wild type or isogenic mutant strain, 2 µL from a mid-log culture of EcN (top), EcN ΔD4 (middle) or EcN Δ*fliC* (bottom) were inoculated in the middle of a swarming **b)** Bacterial swarming was measured as ratio between the whole plate surface and the surface reached by the bacterium.

EcN wild type possesses a functional flagellum, for this reason, it was expected to swim (see **Figure 22 a**, top image), whereas the *fliC* deficient mutant strain EcNΔ*fliC* was not (see **Figure 22 a**, bottom image). The ability of the EcNΔD4 mutant strain to swarm was mildly impaired (see **Figure 22 a**, middle image). This result indicates that the ability of EcN to swarm partially depends on the presence of the D4 domain, suggesting that the D4 might alter the structure of the flagellum to the point that its function is altered.

# 4 Discussion

## 4.1 Prediction Model

The GM comprises an active and dynamic community fulfilling crucial functions such as production of essential metabolites and overall influencing the immunity of the host [99]. Typically, at least 1000 different species colonize the gut under homeostatic conditions [100], 90 % of which belong to the phyla Bacteroidetes and Firmicutes [101]. Besides the influence of age and genetics, the GM is altered by factors such as drugs and diet [102]. When the alteration turns out to be permanent in terms of composition and function, it is defined as dysbiosis [103].

It is postulated that under certain conditions harmless bacteria may display pathogenic qualities [104, 105], and in those cases they are termed pathobionts.

The dysbiotic state often associated with IBD is characterized by an increase in Proteobacteria [59] such as certain opportunistic *E. coli* like AIEC [102]. Conversely, other *E. coli* strains such as EcN hinder inflammation [87].

It is not clear if dysbiosis is the cause or consequence of IBD but characterization of *E. coli* associated to inflammation will aid IBD diagnostics through the use of such traits associated to inflammation as IBD microbiological biomarkers. Detection of the onset of the shift in microbiota composition and early IBD diagnosis would also make early treatment possible. For this reason, bacterial elements associated with pathogenicity have been thoroughly studied as a way to characterize *E. coli* pathobionts. For example, the *E. coli* fimbriae. This surface element is encoded by the operon *fim* which contains *fim A*, an abundant gene in *E. coli* from CD patients [106], and *fimH* which has been associated with AIEC pathogenicity [107]. Other genes associated with AIEC are *ompA*, encoding an outer membrane protein and *dsbA*, which encodes an oxidoreductase relevant to the assembly of pathogenicity factors [108][109]. Nevertheless, all these genes have also been found in non-pathogenic *E. coli* [107], making it extremely challenging to distinguish between symbionts and pathobionts

For this reason, we aimed at developing a simple method to identify possible genetic elements that serve to profile *E. coli* isolated from IBD considering all genes, irrespective of whether they are potentially involved in pathogenicity or not. Our approach was to build the *E. coli* pangenome from a cohort of IBD or non-IBD individuals and then using logistic regression (LG) as a machine learning tool in order

to produce a list of the genes that most accurately classified *E. coli* as coming from IBD or non-IBD individuals.

With this straight forward approach, starting from a pangenome of more than 17500 annotated genes, we found 4 genes (IS3 family transposase, *trbC*, *eheA* and *licT*) that serve as predictors discriminating IBD from non IBD *E. coli* strains with an accuracy of 83%.

At first glance the annotated predictors do not represent the canonical virulent elements attributed to pathogenic *E. coli* pathogenicity. Interestingly, however, a BLAST [110] search using the nucleotide sequences of the predictors “Periplasmic protein TrbC” and “IS3 family transposase” as a query showed both elements to be present in the whole sequenced genomes of seven uropathogenic *E. coli* (UPEC) strains [111]. This unexpected finding led us to suspect an association between these predictors and *E. coli* pathogenic traits, hence encouraging us to further investigate the putative role of these markers. The gene *trbC* encodes for a periplasmic protein belonging to the type 4 secretion system (T4SS), playing a fundamental role in pore formation and mating of bacteria during conjugation [112]. This is relevant because during inflammation the frequency of bacterial conjugation is increased as a way to quickly transfer genetic information in order to propagate fitness or virulence factors among bacteria [113]. Furthermore, a recent report from Elhenawy and colleagues pointed out that the T4SS plays as well a central role as AIEC colonization factor in CD, most likely through biofilm formation[114]. The capacity to form biofilm is a known feature of AIEC [115] and in IBD it is believed to confer the bacterial community resistance to clearance from the host immune system and antibiotics [116].

Insertion sequences (IS) are short DNA sequences moving within the same genome or between different genomes [117]. IS3 is the biggest IS family, and was reported to be relevant on a genomic island associated with virulence in a clinical *E. coli* strain related to UPEC (BH100) [118]. When analysing which genetic elements are favored to persist on pathogenic *E. coli* genomes and hence, under positive natural selection, Petersen and colleagues reported that transposable elements such as IS were positively selected in *E. coli.* [119]. Recently, Hawkey and colleagues showed that the accumulation of IS within Shigella species impacted the diversity of the group, optimizing their genome size compared to other *E. colis* [120] .Therefore, more than a pathogenic trait itself, the presence of IS3 might be the result of intense horizontal gene transfer within *E. coli* under inflammation. The next marker found was *ehaA*, which

encodes an autotransporter associated with biofilm formation in *E. coli* 0157:H7 and Shiga toxin-producing *E.coli* (STEC) [121]. LicT (on *E. coli* named BglG) is an antiterminator described to exert transcriptional control over the ribosomal RNA *rrnB*, allowing the transcription to proceed beyond this transcription terminator [122]. In *E. coli* it is believed to play a role in fitness, allowing the utilization of β-glucosides under metabolic adaptation [123].

Overall, three of the four candidate markers found are involved in horizontal gene transfer and biofilm formation. Even though our aim was only to show that through a simple statistical approach it was possible to classify *E. coli* isolated from IBD, our results align with the fact that adhesion is a an important trait of pathogenic *E. coli* [124].

To date, most of the studies aiming to characterize pathogenic traits in *E. coli* rely on some kind of comparative genomics approach that considers two steps: first, the identification of certain *E. coli* as a member of previously reported phylogroups associated to pathogenicity, and secondly, studying the abundance of genes previously reported to be associated with pathogenicity as well as the identification of single nucleotide polymorphisms SNPs, or transcriptomic analysis aiming to reveal metabolic changes relevant to *E. coli* under inflammation. The exception are evolutionary studies, such as positive selection analysis, meant to find which genes are being selected to remain over generations in the genome [125]. This is accomplished by calculating the ratio between non-synonymous and synonymous mutations. When a gene accumulates non-synonymous mutations is deemed as being under a positive selection [126]. This notion considers first all the genes involved under positive selection and later the nature of those genes, which represents an unbiased view of the genome content.

Machine learning has importantly contributed to analysis and interpretation of NGS data. Diverse statistical analyses with increasing level of complexity have been incorporated to cope with the equally increasing volume of sequencing data, seemingly making unthinkable to approach genome analysis without complicated algorithms.

Nevertheless, in comparative studies the so-called classical statistical tools such as LG has proved to be as effective as complex algorithms [127]. Furthermore, LG is actively used as a predictive tool of biomarkers associated with inflammation. For example, when deciding which cytokines were associated with the pathobiontic potential of *E. coli* isolates from mice with healthy microbiota, Kittana and colleagues

used LG as predictor tool. They found that IL-6 was the best predictor of inflammation, a result that was then corroborated by diminishing DSS induced inflammation by blocking IL-6 [128].

We are aware that higher representation is needed as well as phenotypical assays such as deletion mutagenesis and adhesion assays to test the biological relevance of the markers we found. However, it seems that the predictors hints to genetic traits related to colonization as relevant markers to classify *E. coli* in IBD and with this to predict the onset of microbiota shift composition.

## 4.2 Flagellin structure as homeostasis determinant.

Primarily recognized as a part of the locomotive organelle, i.e., the flagellum, flagellin has been acknowledged for its notable capability to modulate the immune system [97] [75, 129]. For example, its therapeutic potential as adjuvant for vaccination in cancer is actively investigated [130].

Most *E. coli* strains are flagellated and this motility facilitates the contact of bacterial FliC with its cognate PRR TLR5, which is expressed on the gut epithelium [131]. This interaction is mediated by the conserved region of FliC [83], comprised of the D0 and D1 domain, which is essential for TLR5 activation [95]. The hypervariable region acts as a pivot allowing the folding of the protein and consequentially on flagellum stability. Beyond this structural role, the FliC HVR plays a role in immunogenicity and antibody response [132, 133].

Our initial analysis of the *E. coli* FliC amino acid sequence showed that there is a profile distinguishing *E. coli* from IBD and non IBD individuals. Since EcN is a commensal that displays known probiotic properties [134], it was an ideal model to study whether its FliC contributes to its beneficial effects. In this study we could demonstrate that EcN is able to protect mice from DSS induced inflammation and that effect is in part dependent on EcN FliC and particularly on its HVR. Our results point out that the observed protection is mediated through TLR5 signalling, leading to secretion of IL-22 from intestinal dendritic cells and downregulation of inflammatory cytokines such as IL-6, IL-1β, IL-17 and TNFα.

The structural data we produced showed a strikingly similarity with the recently published FliC structure from EHEC and to a lesser extent to EPEC [96]. In the same work, Kreutzberger and colleagues observed that in EHEC and EPEC, the outer ring of the flagellum filament is made of HVR regions long enough to interact with each

other, hence stabilizing the filament. This aligns with our observations that living EcN lacking the D4 domain has a slight swarming defect.

Since to activate TLR5, two flagellin monomers are required [73], it is possible that the HVR of each EcN FliC monomer serves to generate a stable complex with TLR5 in a similar fashion that the HVRs interact with each other stabilizing the flagellum. Even though our results on TLR5 activation with EcN FliC lacking the D4 domain were inconclusive, the diminished protection observed when living EcN carried a shorter HVR might be the result of a weaker contact between each of the two-monomers necessary to activate TLR5, generating a less stable TRL5-FliC complex.

Interestingly, the EcN FliC amino acid sequence is identical to the one of another pathogenic strain, the uropathogenic *E. coli* CFTO73. Overall the genetic similarity between EcN and CFTO73 [135] has led to suggest that EcN might originate from CFTO3 [136].

In this regard, it is interesting to notice that CFTO3 is able to seemly modulate the immunity of the host as observed by inhibiting secretion of proinflammatory cytokines in mice testicular cells [137]. Furthermore, the early secretion of IL-10 at the onset of CFTO73 infection is mediated by FliC [138] and TLR5 polymorphisms have been associated to recurrent urinary infection [139].

This raises the question of how flagellin from a symbiont such as EcN might exert beneficial effects if it is so similar to the flagellin from a pathogen. In this regard, the context of the bacterium-host interaction matters here. For example, in an experimental design to challenge the epithelial protection by EcN against EHEC O157:H7, Pradhan and Weiss showed that EcN was able to protect the epithelia but unable to replicate, suggesting that the observed protective effect was the product of the stimulation that EcN exerted on the host immune system [140]. This illustrates the complexity and dynamics existent in the host-bacteria interaction.

In conclusion, we showed that LG as a simple yet powerful prediction tool pointed out the importance of non-canonical elements associated to bacteria-host interaction as potential IBD biomarkers. Overall, we are suggesting that these candidate genes must be further investigated, and in doing so enrich our understanding of the unknown dynamics underlying the interaction of *E. coli* with the human host.

Furthermore, we found that EcN FliC exhibits a structure resemblance to that of pathogenic *E. coli,* which stresses one more time that the genetic context of the bacteria and the host are highly dynamic. It is for this reason that we believe that in

order to increase the much-needed accuracy and sensitivity of IBD diagnostics, it is necessary to go even further, for example, taking the presence and structure of bacterial markers such as flagellin into account.

# 5 Material and Methods.

## 5.1 Methods

### 5.1.1 Preparation of chemically competent cells.

The bacterial culture was plated on LB-agar plate without antibiotic and incubated at 37 °C for 16 hours. Then, a colony from the plate was used to inoculate 4 ml of LB-medium and incubated at 37 °C for 16 hours and 160 rpm. Afterwards, the overnight culture was used to start a day-culture with a total volume of 400 ml LB-medium. Then, the culture was incubated at 37 °C and 3 000 g until $OD_{600}$ = 0,4. Then, the culture was incubated on ice for 20 minutes. The bacteria cells were then harvested at 4 °C for 10 min and 3000 g. The supernatant was discarded and the pellet resuspended in a volume of 24 ml of cold 0,1 M $CaCl_2$. The bacteria were then incubated on ice for 30 minutes. Afterwards, the culture was centrifuged as before, the supernatant was again discarded and resuspended in 6,4 ml of storage solution (cold 0,1 M $CaCl_2$, 15% glycerol). Finally, aliquots of 200 µl were generated in microcentrifuge tubes and shock-frozen with liquid nitrogen for -80 °C storage.

### 5.1.2 Preparation of electro competent cells.

Bacteria from LB-stock were inoculated on LB-agar plate without antibiotics and incubated at 37 °C for 16 hours. Then a colony from the plate was used to inoculate 5 ml of LB-medium and incubated at 37 °C for 16 hours and 160 rpm. Afterwards, 1 ml of the overnight culture was used to start a day-culture in a total volume of 100 ml LB-medium. Then, the culture was incubated at 37 °C and 160 rpm until $OD_{600}$ = 0,4. Then, the culture was incubated on ice for 10 minutes. The bacteria cells were then harvested at 4 °C for 10 minutes and 2 500 g. In order to minimize bacterial damage, acceleration and deceleration of the centrifuge was set to a minimum. The supernatant was then discarded and the pellet was resuspended with a volume of 50 ml cold sterile water and then centrifuged at 4 °C for 10 minutes and 2 500 g. This step was repeated one more time and then the bacterial pellet was resuspended in 500 µl of 15% glycerol. Finally, aliquots of 50 µl were generated in microcentrifuge tubes and shock-frozen in liquid nitrogen for -80 °C storage.

### 5.1.3 Transformation-heat shock.

Chemically competent cells stored at -80° C were thawed on ice. Then, either 30 ng of plasmid DNA or the total of Gibson-mix were slowly added and gently mixed. The mixture of plasmid and bacteria cells was let incubate for 1 hr and then incubated for 45 seconds into the thermoblock. Then, the mixture was put back on ice for 5 min. 800 µl of previously warmed LB-medium were gently added to the plasmid-bacteria mixture and then incubated into the thermoblock for 15 min at 37°C and 450 rpm agitation. After incubation the mixture was plated on LB-medium with antibiotics.

### 5.1.4 Transformation-electroporation.

Electrocompetent cells stored at -80° C were thawed on ice. At the same time, the electroporation cuvette was cooled down on ice. Afterwards, 20 ng of plasmid DNA were slowly added to electrocompetent cells and gently mixed. The mixture was then slowly added to the cuvette and electroporated with 2 400 volts during 5 milliseconds. Immediately after, the cuvette was removed from the electroporator and 800 µl of prewarmed LB-medium was added. The total volume was then transferred to a microcentrifuge tube and incubated for 1 hour at 37 °C and 450 rpm.

### 5.1.5 PCR.

The PCR protocols below was followed. The annealing temperature varied according the melting temperature of the primers. Concentrations of the reagents in the PCR reaction were followed accordingly the tag-polymerase (see materials) manufacturer recommendation.

| STEP | DURATION | TEMPERATURE |
|---|---|---|
| Initial denaturation | 3 minutes | 95 °C |
| 1) denaturation | 10 seconds | 95 °C |
| 2) annealing | 5 second | - |
| 3) extension | 30 seconds/ kilobase | 72° C |
| repeat 1) to 3) | 35 times | - |
| final extension | 5 minutes | 72 °C |

### 5.1.6 Cloning through enzymatic assembly (“Gibson assembly”)

The cloning strategy was based on the method published by Gibson and colleagues [141]. Briefly, we designed in total three primer pairs (here named as :GB_F/ R, OP_F/ R and SQ_F/ R). GB_F/ R had a maximal length of 60 base pair (bp). On the 5´extreme of the primer, the first 40 bp were designed to have perfect complementary homology

with the cloning vector (henceforth “vector”) and the next 20 bp were homologue to the designed DNA segment to be cloned (henceforth “insert”). OP_F/ R was designed to open (or linearize) the vector in a way that the 5´and 3´ ends of the vector have perfect complementary homology to GB_F/ R. SQ_F/ R was designed to verify by sequencing the success of the cloning.

The PCR product from GB_F/ R and OP_F/ R was purified and incubated with DpnI at 37 °C during 1 hour in order to digest the DNA template. Then 1,5 µl of the PCR product from GB_F/ R and OP_F/ R were gently added to the Gibson-mix (AG Wagner, IMIT) and then incubated at 50 °C during 1 hour. Finally, 3 µl of the Gibson reaction are used to transformation on chemocompetent cells.

### 5.1.7 Cloning of EcN FliC.

Generation of the construct pET-19b-FliC: genomic DNA was isolated from an overnight culture of 5ml LB-medium of *E. coli* Nissle 1917 using the kit DNeasy Blood & Tissue Kit (Qiagen) under manufacturer settings. The Gibson cloning was performed using the primers *Fl_pETb_GB_F* and *Fl_pETb_GB_R* (this serve as GB_F/ R as described before). The native expression vector used was pET-19b (Novagen) carrying a 10 histidine-tag in the N-terminus, followed by a cleavage enterokinase site. The transcription is under the control of the T7 promoter and the promoter region is inducible with IPTG. The pET-19b vector was opened using the primers *Fl_pETb_OP_F* and *Fl_pETb_OP_R* (they serve OP_F/ R). The EcN fliC sequenced was cloned between the sites XhoI and NdeI. As a note, the cloning was performed as mentioned before, the restriction sites serve here only as a reference.

### 5.1.8 Generation of pET-19b–TEV.

Our collaborators observed proteolytic degradation when they removed the histidine-tag with enterokinase. Aiming to increase specificity of the cleave site, we exchanged the enterokinase site for the highly specific Tobacco Etch Virus (TEV) cleavage site [142]. For this, using the native pET-19b as template and the primers pairs *pETb__to_TEV_mega_F* and *pETb__to_TEV_mega_F*, we generated an amplicon that carries the TEV nucleotide sequence but lacks the enterokinase site. This resulting amplicon is used next as the so called “megaprimer”. In the next step the megaprimer is used to amplify though PCR the whole pET-19b vector, where the enterokinase site was exchange for the TEV site.

The principle behind this method is to generate a long double strand DNA fragment (493 bp. in this case) that is almost a 90% complementary to the template. But carrying at the same time a relatively short not complementary region in the middle. (15 bp in this case). In this way, during the PCR reaction the long double stranded DNA fragment melts and due the long homology region it aligns to the template "forcing" the integration of the short region that is not complementary.

### 5.1.9 Cloning of EcN FliC and deletion mutants.

Cloning of EcN FliC on pET-19b–TEV was performed as described on section 5.1.7. Similar procedure was performed for the rest of the constructs, as shown in the scheme bellow.

| CONSTRUCT | VECTOR | TEMPLATE | INSERT | VECTOR |
|---|---|---|---|---|
| FliC (EcN) | pET-19b–TEV -FliC | pET-19b-FliC (insert ) pET-19b–TEV (vector) | *Fl_TEV_GB_F* *Fl_TEV_GB_R* | *Fl_TEV_OP_F* *Fl_TEV_OP_R* |
| FliC ΔD0D1 | pET-19b–TEV-FliC ΔD0D1 | pET-19b–TEV-FliC | *Fl_ΔD0_GB_F* *Fl_ΔD0_GB_R* | *Fl_ΔD0D1_OP_F* *Fl_ΔD0D1_OP_R* |
| FliC ΔD0ΔHVR | pET-19b–TEV-Δ0DΔHVR | pET-19b–TEV-FliCΔD0 | *Fl_ΔD0ΔHVR_GB_F* *Fl_ΔD0ΔHVR_GB_R* | *Fl_ΔD0ΔHVR_OP_F* *Fl_ΔD0ΔHVR_OP_R* |
| FliC ΔD4 | pET-19b–TEV-FliC ΔD4 | pET-19b–TEV-FliC | *Fl_ΔD4_flex_GB_F* *Fl_ΔD4_flex_GB_R* | *Fl_ΔD4_flex_OP_F* *Fl_ΔD4_flex_OP_R* |
| FliC ΔD4 long | pET-19b–TEV-FliC ΔD4 long | pET-19b–TEV-FliC | *Fl_ΔD4_long_GB_F* *Fl_ΔD4_long_GB_R* | *Fl_ΔD4_long_OP_F* *Fl_ΔD4_long_OP_R* |
| FliC ΔDHVR | pET-19b-tev-FliC ΔHVR | pET-19b–TEV-FliC | *Fl_ΔD0ΔHVR_GB_F* *Fl_ΔD0ΔHVR_GB_R* | *Fl_ΔD0ΔHVR_OP_F* *Fl_ΔD0ΔHVR_OP_R* |
| FliC ΔD4 | pSB890-FliC-Nissle ΔD4 | pSB890-FliC-Nissle +/- 1 kb | *Fl_ΔD4_flex_GB_F* *Fl_ΔD4_flex_GB_R* | *Fl_ΔD4_flex_OP_F* *Fl_ΔD4_flex_OP_R* |

### 5.1.10 Sequencing,

All the constructs were verified through sanger sequencing using the primers *Pet_Seq_F* and *Pet_Seq_R.*

### 5.1.11 Protein purification-Colony selection.

The transformed bacteria were plated on selective agar-LB and incubated at 37 °C for 16 hours. After incubation 5 colonies were randomly chosen to separately inoculate 5 culture tubes with 5 ml of LB medium with ampicillin (100 µg/ml) 37 °C for 16 hours and 160 rpm (an additional tube inoculated with sterile water is incubated as sterility control). At the same time each colony was propagated on a gridded selective agar-LB plate with the aim to identify it later.

Afterwards an aliquot of the overnight culture was diluted at a ratio 1:10 in order to determine $OD_{600}$. With this measurement it is possible to calculated the volume of the aliquot of the overnight culture needed to reach a final $OD_{600}$ = 0,1 in a total of 5 ml of LB-medium with antibiotics. This was done for all the overnight cultures, which were then incubated at 37 °C for 2 hours and 160 rpm. When each culture reached an $OD_{600}$ = 0,4 – 0,6, it was induced with 1mM IPTG. The incubation continued at 37 °C for 4 hours and 160 rpm. Next, the cultures were incubated on ice for 15 minutes. Afterwards, 1 ml from each culture was taken and centrifuged at 4 °C for 10 minutes and 15 000 rpm. Next, the supernatant was discarded and the pellet resuspended with 100 µl $H_20$. Then, 15 µl of resuspended culture was lysed and run on SDS gel. Finally, the gel was stained with Comassie blue and the colony that showed a thicker band at the expected molecular size was selected for IMAC.

### 5.1.12 Culture induction and harvesting.

The colony that showed the highest expression of the protein was used to start a pre-culture in 50 ml of LB medium with antibiotics and incubated at 37 °C for 16 hours and 160 rpm. This pre-culture was used to start the day-culture with an $OD_{600}$ = 0,1 in a total volume of 500 ml LB-medium with antibiotics at 37 °C for 2 hours and 160 rpm. When each culture reached an $OD_{600}$ = 0,4 – 0,6, it was induced with 200 µM IPTG and incubated at 18 °C for 16 hours and 90 rpm. After the incubation on ice, the whole culture was centrifuged at 4 °C for 45 minutes and 17 000 rpm and the pellet was discarded.

### 5.1.13 Protein purification-IMAC through gravity flow.

The pellet obtained after harvesting the bacteria was resuspended in lysis buffer (20 mM Tris-HCl [pH 8.0], 10 mM Imidazole, 1 mM $MgCl_2$, 0,1% DOC) to a ratio of 10 ml lysis buffer per gram of pellet. Immediately, CPI (cocktail of protease inhibitor) was added to a concentration of 1µl/ml and gently mixed. Then, 60 Units/ml of benzonase

were added and gently mixed, followed by a 15 -minute-long incubation on ice. Afterwards, sonication was performed for 10 minutes with an amplitude of 35%. Each pulse was 1 -second -long, followed by a pause of 2 seconds. The lysate was then centrifuged at 4 °C for 45 minutes and 35 000 rpm. In between the empty column was filled with a volume of 5 ml of nickel agarose resin solution, this represents an effective volume of 2 ml of nickel resin or column volume (CV). Since the resin is stored in ethanol, at first it was equilibrated with 10 CV of sterile water. Then, the resin was equilibrated with 5 CV of binding buffer (300 mM NaCl, 20 mM Tris-HCl [pH 8.0], 10 mM Imidazole). The lysate supernatant was directly added to the resin after centrifugation. Then, the mixture of resin and lysate was incubated at 4°C for 2 hours and 100 rpm. After the incubation, the mixture was let settle down inside the column and then washed with 1 CV of washing buffer I (300 mM NaCl, 20 mM Tris-HCl [pH 8.0], 10 mM Imidazole). Then, the column was washed with 5 CV of buffer II (300 mM NaCl, 20 mM Tris-HCl [pH 8.0], 20 mM Imidazole). A last washing step was performed with 5 CV buffer III (300 mM NaCl, 20 mM Tris-HCl [pH 8.0], 50 mM Imidazole). Finally, the elution buffer (300 mM NaCl, 20 mM Tris-HCl [pH 8.0], 250 mM Imidazole) was added and all the fractions coming out of the column were collected. The elution continued until $OD_{280}$ < 0,02. All the fractions collected from the column were pooled together and dialyzed at 4 °C for 16 hours with PBS in order to remove the elution buffer.

5.1.14 IMAC through peristaltic pump.

The pellet obtained after harvesting the bacteria was resuspended in lysis buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 5 mM Imidazole) to a ratio of 10 ml of lysis buffer per gram of pellet. Immediately, CPI was added to a concentration of 1µ/ml and gently mixed. Then, 60 Units/ml of benzonase were added and gently mixed, followed by a 15-minute-long incubation on ice. Afterwards, sonication was performed for 10 minutes with an amplitude of 35%. Each pulse was 1-second-long, followed by a pause of 2 seconds. The lysate was then centrifuged at 4 °C for 45 minutes and 35 000 rpm.

A prepacked column of 1 ml (HisTrap, Cytiva) containing the nickel sepharose resin was plugged to a peristaltic pump (Pharmacia-P1). The column was equilibrated with 10 column volumes (CV) of sterile water with a speed of 0,1 ml/min. Then, the column was equilibrated with binding buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 5 mM Imidazole) with a speed of 0,5 ml/min. After centrifugation, the supernatant was loaded

into the column overnight in cold room. The next day the column was washed with 10 CV of buffer I (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 5 mM Imidazole) with a speed of 1 ml/min. Then, the column was washed with 5 CV of washing buffer II (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 20 mM Imidazole). Next, the column was washed with 5 CV of washing buffer III (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 50 mM Imidazole). Finally, the elution buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 150 mM Imidazole) was added with a speed of 0,5 ml/min and all the fractions coming out of the column were collected. The elution continued until the $OD_{280} < 0,02$. All the fractions collected from the column were pooled together and dialyzed at 4 °C for 16 hours with PBS in order to remove the elution buffer.

### 5.1.15 IMAC through FPLC (Äkta sytem).

The FPLC system (henceforth Äkta) possesses different inlets from where the stock solution was pumped into the Äkta, which through a mixer can produce the desired buffer concentration to be used during the purification process. In order to reduce the likelihood of contaminations, all the buffers were previously sterilised (depending on the solution it was used and autoclave or a 0,22 µm filter). All the solutions used were thoroughly degassed to avoid accumulation of gases inside the Äkta and damage to the machine and the sample. Before each purification, the pumps of the Äkta were run with water at a speed of 10 ml/min for 4 minutes.

Lysis was carried out as described before for the peristaltic pump IMAC, but now the sample was additionally filtered with a 0,22 µm filter.

The prepacked column was plugged to the Äkta and equilibrated with 10 CV of water. The filtered sample was loaded into the column with a speed of 0,5 ml/min. Then, the washing and elution steps were performed as described before. Since the system possesses its own UV detection system, we did not perform 280 nm measurements.

### 5.1.16 Size exclusion chromatography (SEC)

The mixture of agarose and dextrose together with the length of the column (600 mm) allows the separation of compounds according to their size. A commercial size exclusion column was used (HiLoad 16/600 Superdex 200 pg) plugged to the ÄKTA system. Before the SEC run, the column was equilibrated with 2 CV of water and then 2 CV PBS buffer (both filtered and degassed). The column was calibrated using a calibration kit for proteins having a known molecular weight between 30 and 200 kDa (Sigma-Aldrich MWGF200). The sample injection proceeded slowly in order to avoid

the entry of air into the column. The run proceeded overnight with a speed of 1 ml/min. Fractions of 1 ml were collected. And aliquots collected in the region of the expected elution of the protein were used for SDS-PAGE.

### 5.1.17 SDS-Page.

From each sample, an aliquot of 20 µl was diluted with water and sample buffer (Bio-Rad Sample buffer 4X Laemmli and 2-mercaptoethanol) to a final volume of 60 µl inside a microcentrifuge tube. The mixture was heated up to 95 °C during 10 minutes and then centrifuged 10 min at 15 000 rpm. From each sample, 5 µl were used for SDS-PAGE.

Precast polyacrylamide gels with a 4 to 15% gradient (Bio-Rad) were used. The polyacrylamide gel was run using a vertical electrophoresis chamber from the system Mini-PROTEAN Tetra Cell (Bio-Rad). Once the board system was assembled, it was filled with SDS running buffer (50 mM Tris, 384 mM Glycin, 0,1% SDS). Before loading the sample, each well from the gel was washed to remove polyacrylamide residues.

The gel was run at 140 volts for 45 minutes. Afterwards, the gels were stained with Coomassie blue and destained with water (Abcam Instant Blue).

5.1.18 Agarose gel.

Agarose gels were prepared with concentrations ranging between 0,8 and 1,2 % depending on the expected amplicon size. The solvent used for preparing the gel was TAE 0,5% (Tris-acetate-EDTA).

The staining reagent used was a commercial alternative to ethidium bromide (SYBR Safe- Thermofisher). The stain solution was used to a proportion of 1 to 1000.

The samples were diluted with water and sample buffer 6x to a final volume of 6µl and loaded in the gel, which was run at 100 volts for 40 minutes.

5.1.19 Exchange mutagenesis.

The pSB890-FliC-Nissle ΔD4 (henceforth suicide vector) carries the sequence encoding the FliC ΔD4 (EcN) construct flanked by 1 000 bp from the EcN chromosome on the 5´and 3´. The suicide vector encodes the enzyme SacB, which makes the bacteria carrying the vector sensible to sucrose. As a note, when sucrose is present, those bacteria still having suicide vector in their chromosome generate a toxic product that kill the bacteria.

Then the suicide vector was transformed into the chemo competent *E. coli* CC118 λ pir+ and plated on 50 µg/ml tetracycline LB-agar plates. The positive clones were screened and verified through sequencing. The positive colony was used to inoculate a culture of LB-medium with 50 µg/ml tetracycline and incubated for 37° C during 16 hours. The vector was then isolated and electroporated into a strain able to mate, *E. coli* β2168Δnic35. Then plated in LB-agar-tetracycline-DAP plates. A volume of 900 µl from an overnight culture of the chosen *E. coli* β2168Δnic35 was mixed with the same volume of an overnight culture from EcN. The mixture was centrifuged for 2 minutes at 6 000 g, the supernatant was discarded and the pellet was washed two times with a solution of 2 ml of LB and 40 µl DAP. At the end, 30 µl of the pellet were carefully pipetted out and placed gently in the center of a LB-DAP plate, which was then incubated overnight. The next day merodiploids were recovered, hence a fraction of the bacteria grown overnight were washed with 1 ml LB-medium and plated on LB and tetracycline overnight at 37°C. Next day, the colonies were used to inoculate 5 ml of LB-medium and let grow overnight at 37°C. Next day, 100 µl of the culture were used to inoculate an LB-sucrose plate which was incubated overnight at 30° C. Next day, each colony was inoculated on an LB-agar plate and LB-agar tetracycline plate, both

plates were incubated overnight at 37 °C and the colonies grown next day on LB-agar were selected for screening.

### 5.1.20 Mammalian cell culture assay.

The mTLR5/mTLR4-HEK293 cells used here were human embryonic kidney cells stably overexpressing mouse TLR5 or TLR4 (InvivoGen). Cells were cultured with DMEM, 10% FCS (previously inactivated during 30 minutes at 50 °C) and 1% penicillin/streptomycin. mTLR5/mTLR4-HEK293 cells were maintained with 100 µg/ml of hygromycin and 50 µg/ml normocin and incubated at 37°C and 5% $CO_2$. For stimulation, 2 x $10^5$ cells were seeded in each condition assayed as triplicate. The cells were incubated during 24 hours at 37 °C and 5 % $CO_2$.

After stimulation, the supernatant was carefully removed in order to measure the secretion of IL-8 trough enzyme-linked immunoassay (ELISA)

### 5.1.21 ELISA

The assay was performed in a flat bottom 96-wells plate (Maxisorb). Each well of the plate was incubated overnight at 4° C with 50 µl ELISA-binding solution. Next day the plate was washed four times with ELISA-washing buffer and incubated with 200 µl of ELISA-blocking buffer per well at room temperature for 2 hours. Next the plate was washed as before. Then 100 µl of sample (as triplicate) and 100 µl of IL-8 standard (as duplicate) are placed into the plated. IL-8 standard (BD) were previously serially diluted, ranging a concentration from 800 to 12,5 pg/ml. The plate was then incubated at room temperature for 2 hours. Next the plate was washed as before and then 100 µl of ELISA-detection buffer was added per well and incubated at room temperature for 1 hour. Then plate was washed as before and 100 µl per well of ELISA-conjugate buffer was added and incubated for 1 hour at room temperature. Next the plate was washed as before and 50 µl of ELISA-substrate buffer was added per well and incubated in the dark during 1 hour at 37 °C. Then, the measurement was performed with the ELISA reader under a length wave of 405 nm and a reference wavelength of 490 nm. Statistical analysis was performed using one-way-ANOVA.

### 5.1.22 Limulus amebocyte lysate (LAL) test.

The test was performed using the Pierce LAL chromogenic Endotoxin Quantification Kit from Thermo scientific (catalogue number 88282). The protocol was followed as indicated by the manufacturer. Briefly here, LPS activate a chromogenic substrate on

a concentration dependent manner. LPS standards and the samples reacted with the chromogenic substrate during 10 minutes, after that the reaction was stopped with 25 % acetic acid. The readout results from the detection of color at 405-410 nm. Statistical analysis was performed using a parametric Student *t* test.

5.1.23 DSS colitis experiment.

All the treatments (either living bacteria, recombinant flagellin or flagella enriched preparation) were given to mice three days before the administration of 3,5 % (w/v) DSS in the drinking water. At this point is set “Day 0” of the experiment and body weight determined. Next measurement was performed on day 3 and daily until day 7 (details on page 110).

5.1.24 Flagellin sequence analysis from *E. coli* from IBD and non-IBD conditions.

The flagellin amino acid sequence was extracted from the sequenced genome from *E. coli* isolates (same as from manuscript included in this monography). Bacterial isolation, DNA isolation, sequencing, assembly and annotation were performed as shown in section 3.1.4.1) . The software ARTEMIS (Sanger Institute) was used to look for annotated FliC on each of the genomes. A multi sequence alignment (MSA) was performed utilizing the tool MUSCLE (MUltiple Sequence Comparison by Log-Expectation) under standard settings and as part of the suite UGENE v 1.32.0 . The MSA was used as input data for the calculation of Shannon diversity. The calculation and plotting was performed using R studio [143] with the library *vegan*.

5.1.25 Shannon diversity calculation.

$$H' = -\sum_{i=1}^{s} p_i * \log p_i$$

$H'$ = Shannon diversity index

$p_i$ = In a defined position, the number

$\log p_i$ = natural logarithm of $p_i$

As example from section 3.2.1 :

Example 1.

It is desired to calculate the SDI for the first position of 10 proteins.

Since the first position has always methionine, then

$p_i$ = $\frac{10 \text{ methionine}}{10 \text{ residues in the first position}}$ = 1

log (1)= 0

The formula considers then the addition of each $pi * \log pi$ product, then SDI is going to be zero.

Example 2.

It is desired to calculate the SDI for the second position of the same 10 proteins. In this case 2 proteins have valine and 8 proteins have serine, then.

For valine:

$p_i$ = $\frac{\text{2 valine}}{\text{10 residues in the second position}}$ = 0,2

log (0,2) = -1,609

$p_i * \log p_i$ = -0,2 * -1,609 = - 0,321

For serine:

log (0,8) = -0,223

$p_i * \log p_i$ = 0,8 * -0,223 = - 0,178

Finally, the values are tabulated together

$$H' = -(-0{,}321 + -0{,}178)$$

$$H' = 0{,}499$$

## 5.2 Materials

### 5.2.1 Media

| MEDIA | COMPONENT | MANUFACTURER |
|---|---|---|
| Agar | Agar | Difco |
| Agarose-Gel | Agarose (1%) | Lonza |
| Cell culture Media | DMEM (500ml), L-Glutamin,Glucose (4,5g/l) | Gibco |
| | FCS (10%) | Sigma-Aldrich |
| | Glucose (4,5g/l) | Gibco |
| Endo Agar | Peptone 1%, lactose 1%, $K_2HPO_4$ 0,35%, $Na_2SO_3$ 0,25 %, basic fuchsin 0, 05% | IMIT |
| Luria-Bertani (LB)-Medium | Tryptone 1%, NaCl 1%, yeast extract 0,5%. | IMIT |
| Motility Agar | Bacto-tryptona 1%, NaCl 0,3 %, Bacto agar 0,3% | IMIT |
| Sheep Blood Agar | Agar, sheep blood 5% | Thermofisher |
| Sucrose Agar | Tryptone 1%, tryptone extract 0,5% , agar 1,5%, sucrose 1 % | IMIT |

### 5.2.2 Chemicals and Reagents

| CHEMICAL | MANUFACTURER |
|---|---|
| DNA Polymerase (2 U/µL) Phusion™ High-Fidelity | Thermofisher |
| p-nitrophenyl phosphate (PNPP) | Sigma-Aldrich |
| Benzonase® Nuclease HC, Purity > 90% | Merk |
| Blasticidin | Invivogen |
| Diaminopimelic acid DAP | Merck |
| DpnI | Thermofisher |
| Dulbecco′s Phosphate Buffered Saline (PBS) | Sigma-Aldrich |
| Hygromicin B | Invivogen |
| Imidazole | Applichem |
| IPTG (Isopropyl-β-D-thiogalactopyranoside) | WVR |
| Protease Inhibitor Cocktail (DMSO solution) | Merck |
| Streptavidin-AP | Roche |
| Tetracycline | IMIT |
| Trizma-Base | Merck |

### 5.2.3 Buffers

| BUFFER | COMPONENT |
|---|---|
| TBE-Buffer | Tris-borate-EDTA |
| ELISA-binding solution | 0,1 M $Na_2HPO_4$; pH= 9,0, 3 µg/ml anti-human IL-8 |
| ELISA-blocking buffer | PBS, 10% FCS |
| ELISA-detection buffer | ELISA-blocking-tween buffer, 1 µg/ml of biotin mouse anti-human IL-8 |
| ELISA-blocking-tween buffer | Elisa-blocking buffer, 0,05% Tween-20 |
| ELISA-washing buffer | PBS, 0,05 % FCS |
| ELISA-substrate buffer | 0,05 M $Na_2CO_3$ , 0,05 M $NaHCO_3$, 1 mM $MgCl_2$ ; pH = 9,8 , 1mg/ ml of p-NPP |
| ELISA-conjugate buffer | ELISA-blocking-tween buffer, 0,1 % Streptavidin-AP |

### 5.2.4 Machines and disposables.

| MATERIAL | MANUFACTURER |
|---|---|
| 250 Sonifier | Branson |
| ÄKTAprime plus | Cytiva |
| Amicon Ultra-15, 10 kDa | Merck |
| Centrifuge | Eppendorf |
| Commasie blue Instant Blue | Abcam |
| Dry Dialysis Tubing 3.5 kD | Spectra/Por |
| ELISA Reader | Tecan |
| ELISA Washer | Tecan |
| HiLoad Superdex 200 pg prepacked column | Cytiva |
| HiLoad Superdex 75 pg prepacked column | Cytiva |
| Laemmli Sample Buffer 4x | Biorad |
| Ni Sepharose column, HisTrap | Cytiva |
| Photometer | Eppendorf |
| Polyacrylamide 4–15% precast gel | Biorad |
| Thermocycler | Biorad |
| Thermocycler T3 | Biometra |
| Ultra centrifuge | Beckman coulter |

### 5.2.5 Software

| SOFTWARE, use | MANUFACTURER |
|---|---|
| https://www.bioinformatics.org/sms/rev_comp.html, cloning design | Scilico, LLC |
| Amplif X 2, primer design. | CRN2M |
| https://www.bioinformatics.org/sms2/pcr_products.html, in sillico PCR | Scilico, LLC |
| Lasergene 12, cloning design | DNASTAR |
| https://web.expasy.org/translate/ | SIB Swiss Institute of Bioinformatics |
| Unicorn 7, Protein purification | Cytiva |

### 5.2.6 Strains

| STRAIN | DESCRIPTION | ORIGIN |
|---|---|---|
| *E. coli* DH5 α | *lacZ*ΔM15 | NEB® |
| *E. coli* NiCo21(DE3) | protein expresion. OmpT and Lon deficient. | NEB® |
| *E. coli* CC118 pir+ | propagation suidice vector kanR; ampR | AG Wagner, Tübingen (IMIT); |
| *E. coli* β2168Δnic35 | Δ*asd* (DAP dependent) | AG Wagner, Tübingen (IMIT); |
| *Escherichia coli Nissle* 1917 (EcN) | wild type | Ardeypharm, Herdecke; Deutsche Stammsammlung (DSMZ) 6601 |
| *E. coli* Nissle Δ*fliC* | kanR | AG Oelschlaeger, Würzburg (IMBI) |
| *E. coli* Nissle ΔD4 | clean deletion of the FliC D4 domain | This Work |
| *E. coli* Nissle Hyperflagellated | positive control for swarming | Schubert/Wieser, München (MPK) |

### 5.2.7 Vectors

| VECTOR | DESCRIPTION | ORIGIN |
|---|---|---|
| pET-19b | T7 promoter, ampR, Enterokinase cleavage site, 10xHIS tag | Novagen (Merck) |
| pET-19b-FliC | EcN *fliC* cloned into pET-19b | This work |
| pET-19b-TEV | pET-19b with TEV instead of Enterokinase site | This work |
| pET-19b–TEV-FliC | EcN *fliC* cloned into pET-19b-TEV | This work |
| pET-19b–TEV-FliC ΔD0D1 | EcN *fliC* ΔD0D1 cloned into pET-19b-TEV | This work |
| pET-19b–TEV-ΔD0ΔHVR | EcN *fliC* ΔD0ΔHVR cloned into pET-19b-TEV | This work |

| VECTOR | DESCRIPTION | ORIGIN |
|---|---|---|
| pET-19b–TEV-FliC ΔD4 | EcN *fliC* ΔD4 cloned into pET-19b-TEV | This work |
| pET-19b–TEV-FliC ΔD4 long | EcN *fliC* ΔD4 long cloned into pET-19b-TEV | This work |
| Pet-19b-tev-FliC ΔDHVR | EcN *fliC* ΔHVR cloned into pET-19b-TEV | This work |
| pSB890-FliC-Nissle ΔD4 | EcN *fliC* ΔD4 flanked on 5´and 3´ by 1 kb of the EcN chromosome | This work |
| pSB890-FliC-Nissle +/- 1 kb | EcN *fliC* and flanked on 5´and 3´ by 1 kb of the EcN chromosome | AG Wagner, Tübingen (IMIT) |
| pET-19b-FliC_ΔD0 | EcN *fliC* ΔD0 cloned in pET-19b-TEV | AG Stehle , Tübingen (IFIB) |

### 5.2.8 Constructs

| CONSTRUCT | AMINO ACID SEQUENCE |
|---|---|
| FliC(EcN) | MAQVINTNSLSLITQNNINKNQSALSSSIERLSSGLRINSAKDDAAGQ AIANRFTSNIKGLTQAARNANDGISVAQTTEGALSEINNNLQRIRELT VQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAKDGS MKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGTIANKAATISDLT AAKMDAATNTITTTNNALTASKALDQLKDGDTVTIKADAAQTATVYT YNASAGNFSFSNVSNNTSAKAGDVAASLLPPAGQTASGVYKAASG EVNFDVDANGKITIGGQEAYLTSDGNLTTNDAGGATAATLDGLFKKA GDGQSIGFNKTASVTMGGTTYNFKTGADAGAATANAGVSFTDTASK ETVLNKVATAKQGTAVAANGDTSATITYKSGVQTYQAVFAAGDGTA SAKYADNTDVSNATATYTDADGEMTTIGSYTTKYSIDANNGKVTVDS GTGTGKYAPKVGAEVYVSANGTLTTDATSEGTVTKDPLKALDEAISS IDKFRSSLGAIQNRLDSAVTNLNNTTTNLSEAQSRIQDADYATEVSN MSKAQIIQQAGNSVLAKANQVPQQVLSLLQG |
| FliC ΔD0D1(EcN) | MNGSGTIANKAATISDLTAAKMDAATNTITTTNNALTASKALDQLKDG DTVTIKADAAQTATVYTYNASAGNFSFSNVSNNTSAKAGDVAASLLP PAGQTASGVYKAASGEVNFDVDANGKITIGGQEAYLTSDGNLTTND AGGATAATLDGLFKKAGDGQSIGFNKTASVTMGGTTYNFKTGADAG AATANAGVSFTDTASKETVLNKVATAKQGTAVAANGDTSATITYKSG VQTYQAVFAAGDGTASAKYADNTDVSNATATYTDADGEMTTIGSYT TKYSIDANNGKVTVDSGTGTGKYAPKVGAEVYVSANGTLTTDATSE |
| FliC ΔD0(EcN) | MQAIANRFTSNIKGLTQAARNANDGISVAQTTEGALSEINNNLQRIRE LTVQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAKD GSMKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGTIANKAATIS DLTAAKMDAATNTITTTNNALTASKALDQLKDGDTVTIKADAAQTATV YTYNASAGNFSFSNVSNNTSAKAGDVAASLLPPAGQTASGVYKAAS GEVNFDVDANGKITIGGQEAYLTSDGNLTTNDAGGATAATLDGLFKK AGDGQSIGFNKTASVTMGGTTYNFKTGADAGAATANAGVSFTDTAS KETVLNKVATAKQGTAVAANGDTSATITYKSGVQTYQAVFAAGDGT ASAKYADNTDVSNATATYTDADGEMTTIGSYTTKYSIDANNGKVTVD SGTGTGKYAPKVGAEVYVSANGTLTTDATSEGTVTKDPLKALDEAIS SIDKFRSSLGAIQNRLDSAVTNLNNTTTNLSEAQSRIQD |

| CONSTRUCT | AMINO ACID SEQUENCE |
|---|---|
| FliC ΔD0ΔHVR (EcN) | MQAIANRFTSNIKGLTQAARNANDGISVAQTTEGALSEINNNLQRIRE LTVQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAKD GSMKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGGTVTKDPLK ALDEAISSIDKFRSSLGAIQNRLDSAVTNLNNTTTNLSEAQSRIQD |
| FliC ΔD4 (EcN) | MAQVINTNSLSLITQNNININKNQSALSSSIERLSSGLRINSAKDDAAGQ AIANRFTSNIKGLTQAARNANDGISVAQTTEGALSEINNNLQRIRELT VQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAKDGS MKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGTIANKAATISDLT AAKMDAATNTITTTNNALTASKALDQLKDGDTVTIKADAAQTATVYT YNASAGNFSFSNVSNGSGSATDTASKETVLNKVATAKQGTAVAAN GDTSATITYKSGVQTYQAVFAAGDGTASAKYADNTDVSNATATYTD ADGEMTTIGSYTTKYSIDANNGKVTVDSGTGTGKYAPKVGAEVYVS ANGTLTTDATSEGTVTKDPLKALDEAISSIDKFRSSLGAIQNRLDSAV TNLNNTTTNLSEAQSRIQDADYATEVSNMSKAQIIQQAGNSVLAKAN QVPQQVLSLLQG |
| FliC ΔD4 long (EcN) | MAQVINTNSLSLITQNNININKNQSALSSSIERLSSGLRINSAKDDAAGQ AIANRFTSNIKGLTQAARNANDGISVAQTTEGALSEINNNLQRIRELT VQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAKDGS MKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGTIANKAATISDLT AAKMDAATNTITTTNNALTASKALDQLKDGDTVTIKADAAQTATVYT YNASAGNFSFSNTDTASKETVLNKVATAKQGTAVAANGDTSATITYK SGVQTYQAVFAAGDGTASAKYADNTDVSNATATYTDADGEMTTIGS YTTKYSIDANNGKVTVDSGTGTGKYAPKVGAEVYVSANGTLTTDAT SEGTVTKDPLKALDEAISSIDKFRSSLGAIQNRLDSAVTNLNNTTTNL SEAQSRIQDADYATEVSNMSKAQIIQQAGNSVLAKANQVPQQVLSLL QG |
| FliC ΔDHVR (EcN) | MAQVINTNSLSLITQNNININKNQSALSSSIERLSSGLRINSAKDDAAGQ AIANRFTSNIKGLTQAARNANDGISVAQTTEGALSEINNNLQRIRELT VQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAKDGS MKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGGTVTKDPLKAL DEAISSIDKFRSSLGAIQNRLDSAVTNLNNTTTNLSEAQSRIQDADYA TEVSNMSKAQIIQQAGNSVLAKANQVPQQVLSLLQG |

5.2.9 Primers.

| PRIMER | DNA SEQUENCE 5' TO 3' |
|---|---|
| *FI_pETb_OP_F* | catcacagcagcggccatatcgacgacgacgacaagcatatggcacaagtcat taataccaa |
| *FI_pETb_OP_R* | aactcagcttcctttcgggctttgttagcagccggatcctcgagttaaccctgcagc agagacagaac |
| *FI_pETb_GB_F* | ctcgaggatccggctgctaac |
| *FI_pETb_GB_R* | catatgcttgtcgtcgtcgatatg |
| *pETb__to_TEV_mega_F* | catcatcatcatcacagcagcggccatatcaccaccgaaaacctgtattttcaga gctaacatatgctcgaggatccggctgctaac |
| *pETb__to_TEV_mega_R* | gggttccgcgcacatttccccgaaaagt |

| PRIMER | DNA SEQUENCE 5' TO 3' |
|---|---|
| *Fl_TEV_OP_F* | taacatatgctcgaggatccggctgctaacaa |
| *Fl_TEV_OP_R* | gctctgaaaatacaggttttcggtggtgatatggccgctgctgtgatgatga |
| *Fl_TEV_GB_F* | gaaaacctgtattttcagagcatggcacaagtcattaatacc |
| *Fl_TEV_GB_R* | gttagcagccggatcctcgagcatatgttaaccctgcagcagagacagaacctg |
| *Fl_ΔD0D1_OP_F* | taacatatgctcgaggatccggctgctaacaa |
| *Fl_ΔD0D1_OP_R* | gctctgaaaatacaggttttcggtggtgatatggccgctgctgtgatgatga |
| *Fl_ΔD0D1_GB_F* | agcagcggccatatcaccaccgaaaacctgtattttcagagcaatggttccggtacgatagc |
| *Fl_ΔD0D1_GB_R* | ttgttagcagccggatcctcgagcatatgttattcgctagttgcatctgttgttaa |
| *Fl_ΔD4_long_OP_F* | actgatacagctagcaaagaaaccg |
| *Fl_ΔD4_long_OP_R* | ctcgacagcgcagactggttcttg |
| *Fl_ΔD4_long_GB_F* | tatcaacaagaaccagtctgcgctgtcgagttctatcgagcgtctgtcttctggc |
| *Fl_ΔD4_long_GB_R* | ccactttatttaaaacggtttctttgctagctgtatcagtattactgaatgagaagttac |
| *Fl_ΔD0ΔHVR_OP_F* | ggcacagtaacaaaagatccac |
| *Fl_ΔD0ΔHVR_OP_R* | accggaaccattcacg |
| *Fl_ΔD0ΔHVR_GB_F* | actggcaaaagacggttcgatg |
| *Fl_ΔD0ΔHVR_GB_R* | ttcatccagagctttcagtggatcttttgttactgtgccaccggaaccattcacgtta |
| *Fl_ΔD4_flex_OP_F* | cgaatggcagcggcagcgcgactgatacagctagcaaagaaaccgttttaaataaagtgg |
| *Fl_ΔD4_flex_OP_R* | cgcgctgccgctgccattcgatacattactgaatgag |
| *Fl_ΔD4_flex_GB_F* | actgatacagctagcaaagaaaccgttttaaataaagtgg |
| *Fl_ΔD4_flex_GB_R* | ccagggatgaacggaatttgtcgatg |
| *pET_SQ_F* | ccgcgaaattaatacgactcac |
| *pET_SQ_F* | cctcaagacccgtttagagg |
| *EcNΔD4_SQ_1_F* | gcggtgaggctatcaaggattgaa |
| *EcNΔD4_SQ_1_R* | gttgagaagcggtgtaagtg |
| *EcNΔD4_SQ_2_F* | ttgaccgtcagctgtgcattct |
| *EcNΔD4_SQ_2_R* | ccagggatgaacggaatttgtcgatg |

# 6 Addendum

## Flagellin hypervariable region determines symbiotic properties of commensal *Escherichia coli* strains

PLOS BIOLOGY

RESEARCH ARTICLE

# Flagellin hypervariable region determines symbiotic properties of commensal *Escherichia coli* strains

**Alex Steimle[1,2☯], Sarah Menz[1,2☯], Annika Bender[1,2], Brianna Ball[1,2], Alexander N. R. Weber[3], Thomas Hagemann[1,2], Anna Lange[1,2], Jan K. Maerz[1,2], Raphael Parusel[1,2], Lena Michaelis[1,2], Andrea Schäfer[1,2], Hans Yao[1,2], Hanna-Christine Löw[1,2], Sina Beier[4], Mehari Tesfazgi Mebrhatu[1,2], Kerstin Gronbach[1,2], Samuel Wagner[1,2], David Voehringer[5], Martin Schaller[6], Birgit Fehrenbacher[6], Ingo B. Autenrieth[1,2], Tobias A. Oelschlaeger[7], Julia-Stefanie Frick[1,2]***

**1** Institute of Medical Microbiology and Hygiene, University of Tübingen, Tübingen, Germany, **2** German Center for Infection Research, Partner Site Tübingen, Tübingen, Germany, **3** Department of Immunology, University of Tübingen, Tübingen, Germany, **4** Chair of Algorithms in Bioinformatics, Faculty of Computer Science, University of Tübingen, Tübingen, Germany, **5** Department of Infection Biology, University Hospital Erlangen, Erlangen, Germany, **6** Department of Dermatology, University Hospital Tübingen, Tübingen, Germany, **7** Institute for Molecular Infection Biology, University of Würzburg, Germany

☯ These authors contributed equally to this work.
* julia-stefanie.frick@med.uni-tuebingen.de



OPEN ACCESS

**Citation:** Steimle A, Menz S, Bender A, Ball B, Weber ANR, Hagemann T, et al. (2019) Flagellin hypervariable region determines symbiotic properties of commensal *Escherichia coli* strains. PLoS Biol 17(6): e3000334. https://doi.org/10.1371/journal.pbio.3000334

**Academic Editor:** Luis Teixeira, Instituto Gulbenkian de Ciencia, PORTUGAL

**Received:** October 16, 2018

**Accepted:** June 5, 2019

**Published:** June 17, 2019



**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. S1 Data is also available at https://figshare.com/s/2ae422600252771b58d9.

**Funding:** This work was supported by Deutsches Zentrum für Infektionsforschung (German Center for Infection Research, http://www.dzif.de/en/), Deutsche Forschungsgemeinschaft (DFG) GRK1708 (http://www.dfg.de/gefoerderte_projekte/programme_und_projekte/listen/projektdetails/index.jsp?id=174858087), Deutsche

## Abstract

*Escherichia coli* represents a classical intestinal gram-negative commensal. Despite this commensalism, different *E. coli* strains can mediate disparate immunogenic properties in a given host. Symbiotic *E. coli* strains such as *E. coli* Nissle 1917 (EcN) are attributed beneficial properties, e.g., promotion of intestinal homeostasis. Therefore, we aimed to identify molecular features derived from symbiotic bacteria that might help to develop innovative therapeutic alternatives for the treatment of intestinal immune disorders. This study was performed using the dextran sodium sulphate (DSS)-induced colitis mouse model, which is routinely used to evaluate potential therapeutics for the treatment of Inflammatory Bowel Diseases (IBDs). We focused on the analysis of flagellin structures of different *E. coli* strains. EcN flagellin was found to harbor a substantially longer hypervariable region (HVR) compared to other commensal *E. coli* strains, and this longer HVR mediated symbiotic properties through stronger activation of Toll-like receptor (TLR)5, thereby resulting in interleukin (IL)-22–mediated protection of mice against DSS-induced colitis. Furthermore, using bone-marrow–chimeric mice (BMCM), CD11c+ cells of the colonic lamina propria (LP) were identified as the main mediators of these flagellin-induced symbiotic effects. We propose flagellin from symbiotic *E. coli* strains as a potential therapeutic to restore intestinal immune homeostasis, e.g., for the treatment of IBD patients.

Forschungsgemeinschaft (DFG) SFB685 (http://gepris.dfg.de/gepris/projekt/13799719). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** BMCM, bone-marrow–chimeric mice; CD, C-terminal domain; CFU, colony-forming unit; cLP, colonic LP; DC, dendritic cell; DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; FEP, flagella-enriched preparation; fliC, flagellin; FSC, forward scatter; GM-CSF, granulocyte-macrophage colony-stimulating factor; HCS, histological colitis score; HE, hematoxylin–eosin; HVR, hypervariable region; IBD, Inflammatory Bowel Disease; IEC, intestinal epithelial cell; IFN, interferon; IL, interleukin; ILC, innate lymphoid cell; intDC, intestinal DC; LP, lamina propria; MAFFT, Multiple Alignment using Fast Fourier Transform; MAMP, microbe-associated molecular pattern; MCP, monocyte chemoattractant protein; MG1655, *E. coli* K12 MG1655; MOI, multiplicity of infection; MPK, *E. coli* mpk; mTLR5-HEK293 cell, mouse-TLR5–expressing human embryonic kidney 293 cell; MyD, myeloid differentiation primary response; NAIP, NLR family, apoptosis inhibitory protein 5; ND, N-terminal domain; NFκB, nuclear factor "kappa-light-chain–enhancer" of activated B-cells; NK, natural killer; NLRC, NLR Family CARD Domain Containing; PP, Peyer's Patch; PRR, pattern-recognition receptor; rfliC(EcN), recombinant flagellin from EcN; rfliC(K12), recombinant flagellin from MG1655; SPF, specific-pathogen–free; TCRγδ, T-cell receptor γδ; Th, T helper; TLR, Toll-like receptor; TNF, tumor necrosis factor; UC, ulcerative colitis; WT, wild type.

## Introduction

*E. coli* belongs to the phylum of gram-negative Proteobacteria. Besides certain pathogenic strains, *E. coli* represents a commensal member of the intestinal microbiota. However, distinct commensal *E. coli* strains can mediate substantially different immunological host responses. On the one hand, so-called "pathobionts" may induce severe pathological inflammatory reactions in a certain genetically predisposed or environmentally challenged host. On the other hand, "symbionts" generally provide beneficial effects and do not induce inflammatory responses at intestinal mucosal interfaces [1]. Although *Escherichia* is usually not among the most abundant bacterial genera within a healthy, balanced, and diversified intestinal microbiota [2,3], the impact of enhanced proportions of Proteobacteria in general and *E. coli* strains in particular on inflammatory processes in Inflammatory Bowel Disease (IBD) patients has widely been reported [4–7]. This observation leads to questions concerning mechanistic and structural differences between symbiotic and nonsymbiotic commensal *E. coli* strains and their subsequent impact on IBD pathology. One of the most intensely studied symbiotic *E. coli* strains is *E. coli* Nissle 1917 (EcN). EcN is generally classified as a probiotic and is the only bacterial symbiont that is successfully used to extend remission phases in IBD patients in clinical routine [8]. In this context, EcN mediates similar therapeutic effects as mesalamine, the gold standard therapeutic to extend remission time in ulcerative colitis (UC) patients [9–11]. EcN provides different beneficial properties such as (1) the formation of biofilms [12] leading to the production of defensins [12,13], (2) strengthening of tight junctions within the intestinal epithelium [14], (3) direct antimicrobial effects via secretion of bacteriocins and microcins [15], and (4) direct interaction with the host immune system [16].

However, most of these effects are thought to require viable EcN bacteria, and the recommended therapeutic EcN dose comprises extremely high numbers of vital cells [2]. In general, administration of viable bacteria involves the risk of proactive bacterial translocation across the epithelial barrier, particularly in inflamed intestinal tissue, which provides a disturbed barrier integrity. Thus, use of live bacteria for the treatment of ongoing inflammatory reactions bears considerable risks. Therefore, it appears desirable to identify bacterial factors that distinguish symbiotic from nonsymbiotic *E. coli* strains. Such identified symbiotic factors could then be used as novel therapeutic approaches to restore gut immune homeostasis and could offer a wide range of potential clinical applications.

The investigation of commensal bacteria-mediated symbiotic properties requires the use of adequate mouse models. One of these mouse models involves the application of dextran sodium sulphate (DSS). Administration of DSS leads to an IBD-resembling phenotype in mice of almost all genetic backgrounds [17]. The similarities between IBD in humans and DSS-induced colitis in mice include a similar inflammatory gene expression pattern [18], T-cell accumulation in the colon [19], and the development of a chronic pathology after initial acute inflammation involving a T helper (Th)1/Th2 cytokine secretion pattern [20]. As in IBD patients, DSS-induced colitis in mice leads to influx and transepithelial migration of neutrophils into the mucosal epithelium and lumen, leading to the formation of crypt abscesses [21–24]. Furthermore, DSS-induced colitis in mice provides comparable sensitivity toward therapeutics as in IBD [25], making this model appropriate for preclinical studies involving the evaluation of new potential treatments for IBD patients [26], e.g., commensal-derived molecular features such as surface-associated structures.

Various bacterial surface structures serve as so-called microbe-associated molecular patterns (MAMPs), molecules that are recognized by host pattern-recognition receptors (PRRs), e.g., Toll-like receptors (TLRs) [27]. PRR sensing leads to activation of the innate and modulation of the adaptive immune system. One of these MAMPs is flagellin, the constitutive protein

building up bacterial flagella. Flagellin typically consists of different structural domains: all flagellins contain N-terminal and C-terminal D0 and D1 domains that form the structurally highly homogenous so-called "conserved region" [28–30]. Some flagellins, such as most *E. coli* flagellins, additionally contain a "hypervariable region" (HVR) made up of C- and N-terminal D2 domains as well as of a central D3 domain [28]. Extracellular flagellin is recognized only as a monomer [29] by host TLR5 [30], leading to myeloid differentiation primary response (MyD)88-dependent activation of nuclear factor "kappa-light-chain–enhancer" of activated B-cells (NFκB) [31,32]. Interestingly, flagellin represents a major target antigen in human IBD patients as well as in experimental mouse models for colitis [33]. Additionally, certain *TLR5* SNPs are associated with higher incidence of UC [34] and colon cancer [35]. Furthermore, TLR5-deficient mice are prone to develop spontaneous intestinal inflammation [36], and intestinal TLR5 signaling was demonstrated to be crucial for preventing gut inflammation and metabolic syndrome in mice [37].

Taking these data together, flagellin recognition by intestinal TLR5-expressing host cells seems to be crucially involved in shaping host immunity, leading to maintenance of intestinal immune homeostasis. Therefore, we hypothesized that the precise flagellin structure could be a decisive factor that allows for a distinction between symbiotic and nonsymbiotic *E. coli* strains. Here, we demonstrate that EcN flagellin is sufficient to mediate a crucial part of EcN's symbiotic properties via TLR5 on intestinal $CD11c^+$ cells within the lamina propria (LP). Furthermore, we provide evidence that the structure of the flagellin HVR is a decisive factor that classifies EcN as a symbiont. In consequence, symbiotic HVR containing flagellin structures might be used as innovative therapeutic approaches for the treatment of IBD.

## Results

### Interaction of EcN flagella with host TLR5 is crucial for the beneficial effects of EcN during DSS-induced colitis

EcN is the only symbiont with verified symbiotic effects on the outcome and progression of UC in humans [8] as well as in DSS-induced colitis in mice [38–40]. For these reasons, we decided to use the DSS-induced colitis model for the elucidation of *E. coli*-mediated effects on colitis pathology. Administration of 3.5% DSS in the drinking water led to strong loss of body weight in wild-type (WT) C57BL/6 mice (Fig 1A), accompanied by severe tissue changes and leukocyte influx, as indicated by high histological colitis scores (HCSs) (Fig 1B). In agreement with previous studies [38–40], administration of viable symbiotic EcN provided beneficial effects on the outcome of DSS-induced colitis in WT BL/6 mice, characterized by drastically reduced weight loss and HCS (Fig 1A and 1B). Therefore, we were interested in how other commensal *E. coli* strains influence the progress and outcome of DSS-induced colitis and which structural features of these strains might mediate potentially observed differences.

Therefore, we administered two additional *E. coli* strains in the same concentration as EcN, which resulted in comparable *E. coli* colony-forming units (CFUs) in the feces of all groups (S1 Fig), and monitored disease progression and outcome compared to DSS-only–treated WT mice; specifically, the well-known commensal strains (1) *E. coli* K12 MG1655 (MG1655) and (2) *E. coli* mpk (MPK) [41–44] were used. Importantly, these strains do not mediate symbiotic properties in various mouse models for microbiota-influenced pathologies [41,43,45], and in contrast to treatment with EcN, administration of MPK and MG1655 did not provide any beneficial effects on the outcome of DSS-induced colitis since the HCS of both groups was comparable to the DSS-only control group (Fig 1B), and weight loss was only slightly reduced.

Given the direct effect of flagellin on TLR5-signaling and its impact on intestinal immune homeostasis [36,37], we hypothesized that the protective effect of EcN was at least partially

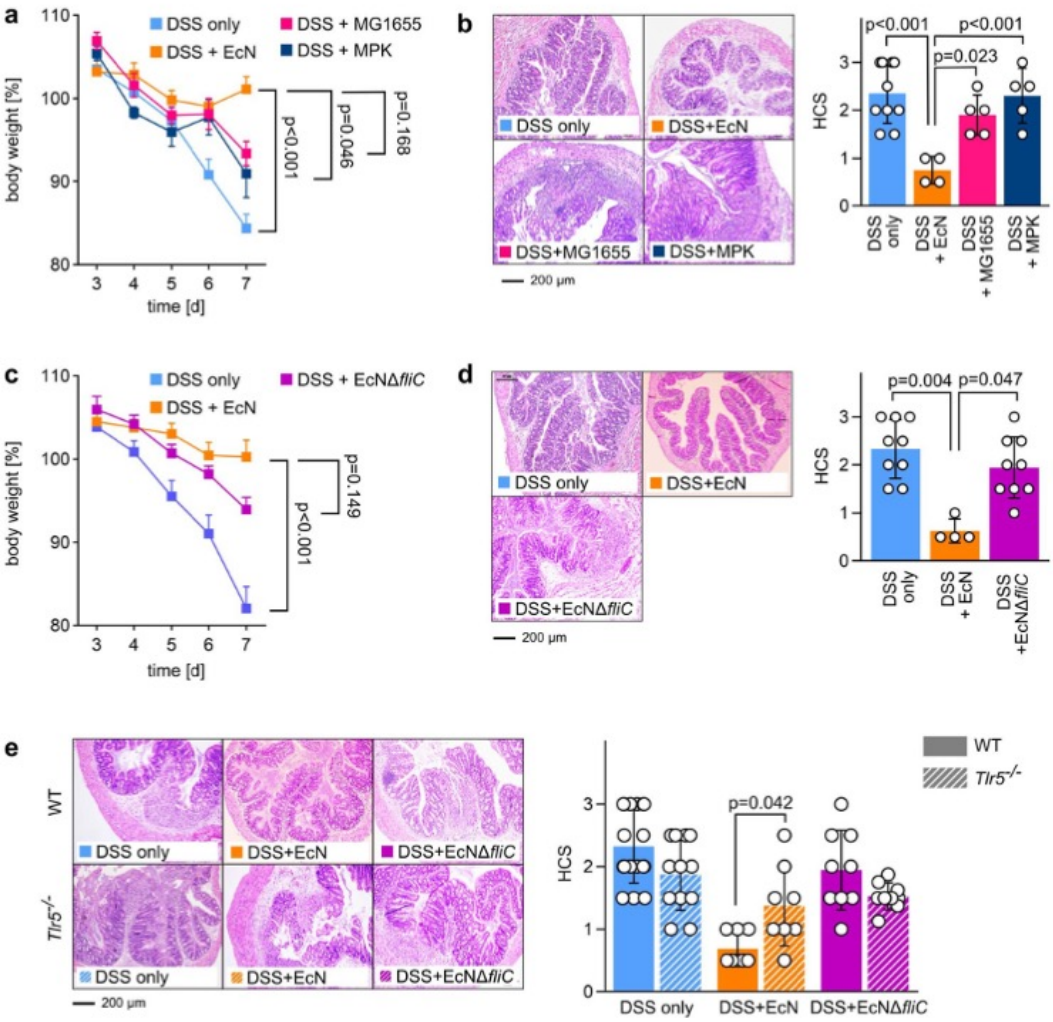



**Fig 1. Influence of different commensal *E. coli* strains on DSS-induced colitis in WT and *Tlr5*$^{-/-}$ mice.** (a + b) SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with EcN (DSS + EcN), MG1655 (DSS + MG1655), or MPK (DSS + MPK) resuspended in DSS-containing drinking water at $10^8$ bacteria mL$^{-1}$. (a) Change in body weight relative to start of DSS administration at day 0. (b) Left panel: HE-stained colonic sections at day 7 after start of DSS administration. Right panel: HCS at day 7. (c + d) SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with EcN (DSS + EcN) or an EcNΔ*fliC* deletion mutant (DSS + EcNΔ*fliC*) resuspended in DSS-containing drinking water at $10^8$ bacteria mL$^{-1}$. (c) Change in body weight relative to start of DSS administration at day 0. (d) Left panel: HE-stained colonic sections at day 7 after start of DSS administration. Right panel: HCS at day 7. (e) SPF C57BL/6 WT mice and SPF *Tlr5*$^{-/-}$ mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with EcN (DSS + EcN) or an EcNΔ*fliC* deletion mutant (DSS + EcNΔ*fliC*) resuspended in DSS-containing drinking water at $10^8$ bacteria mL$^{-1}$. Left panel: HE-stained colonic sections at day 7 after start of DSS administration. Right panel: HCS at day 7. Statistics: (a), (b), (c), (e), one-way ANOVA with Tukey multiple comparison test; (d) Kruskal–Wallis test with multiple comparisons. *p*-values $< 0.05$ are considered to represent statistical significance. (a–e) The data underlying this figure can be found in S1 Data. DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; *fliC*, flagellin; HCS, histological colitis score; HE, hematoxylin–eosin; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk; SPF, specific-pathogen–free; TLR, Toll-like receptor; WT, wild type.

https://doi.org/10.1371/journal.pbio.3000334.g001

mediated by its flagella. In order to test this hypothesis, an EcN mutant lacking the *fliC* gene encoding for flagellin (EcNΔ*fliC*) [46] was administered to DSS-treated mice. While EcN, MG1655, and MPK each harbor a functional flagellum (S2 Fig), EcNΔ*fliC* does not (S3 Fig). Fecal *E. coli* CFUs in all bacteria-treated groups were comparable (S1A Fig). Whereas weight loss was only partially rescued by administration of live EcNΔ*fliC* in the drinking water (Fig 1C), absence of flagellin in this strain completely abrogated the positive effect on histological damage observed for flagellin-expressing EcN and was comparable to DSS-only–treated control mice, providing strong signs of intestinal inflammation (Fig 1D). This indicated that the presence of flagella is crucial for mediation of protective properties of EcN in this mouse model. Since the flagellum and its constituent protein flagellin are recognized by host TLR5, we performed DSS-induced colitis experiments using TLR5-deficient ($Tlr5^{-/-}$) mice.

Dependent on the housing conditions, $Tlr5^{-/-}$ mice either spontaneously develop a chronic form of colitis and/or metabolic syndrome [37,47,48] or not [49]. Our specific-pathogen–free (SPF) housing conditions did not lead to spontaneous intestinal disorders and did not differently affect the outcome of DSS-induced colitis in $Tlr5^{-/-}$ mice compared to their equally treated WT counterparts (Fig 1E), though $Tlr5^{-/-}$ mice treated with EcN showed significantly increased histological damage compared to EcN-treated WT mice, resembling the disease phenotype that was observed in DSS-only–treated WT and $Tlr5^{-/-}$ mice (Fig 1E). Administration of EcNΔ*fliC* to DSS-treated $Tlr5^{-/-}$ mice did not positively affect the outcome of experimental colitis, as demonstrated by HCSs comparable to EcNΔ*fliC*-treated WT mice and to DSS-only–treated mice of both genotypes (Fig 1E; see also S1 Table for detailed statistical analysis). Importantly, fecal *E. coli* CFUs in all bacteria-treated groups in $Tlr5^{-/-}$ mice were comparable (S1B Fig). Therefore, we concluded that the beneficial effects of symbiotic EcN on the progress and outcome of DSS-induced colitis is mainly mediated by the interaction of host TLR5 with EcN flagella.

## Flagella-mediated inflammation-suppressing effects are specific for EcN flagella-enriched preparations and are not mediated by preparations from other commensal *E. coli* strains

Since all three *E. coli* strains used—EcN, MPK, and MG1655—expressed a functional flagellum, we hypothesized that the different effects on DSS-induced colitis mediated by distinct *E. coli* strains were not caused by the presence of a flagellum per se but rather were rooted in structural differences in their flagellin protein structures. To test this and to exclude the contribution of factors requiring viable bacteria, we generated flagella-enriched preparations (FEPs) from symbiotic EcN and nonsymbiotic MG1655 and MPK as described in the experimental procedures, and these FEPs were devoid of any viable bacteria. Fig 2A demonstrates that the EcN FEP recapitulated the protective effects on the outcome of DSS-induced colitis in WT mice observed for viable EcN. This effect was dose dependent, with the FEP obtained from $10^{10}$ EcN per mL drinking water providing the strongest inflammation-reducing effects, as indicated by absent weight loss (Fig 2A) and low HCS of DSS-treated animals (Fig 2B).

According to the results shown in Fig 2A and 2B, we performed all further experiments by default with the FEP obtained from $10^{10}$ bacteria per mL drinking water, which was thought to induce a strong TLR5-mediating signaling. Next, we compared the FEP from EcN with the FEPs obtained from MPK (MPK FEP) and MG1655 (MG1655 FEP). We verified that FEPs generated from each of the tested strains contained similar concentrations of endotoxin, overall protein, and flagellin (S4A–S4E Fig). Thus, DSS-treated WT mice were administered MPK FEP as well as MG1655 FEP and compared to DSS-treated WT mice, which were administered EcN FEP. As demonstrated in Fig 2C and 2D, neither MPK FEP nor MG1655 FEP provided

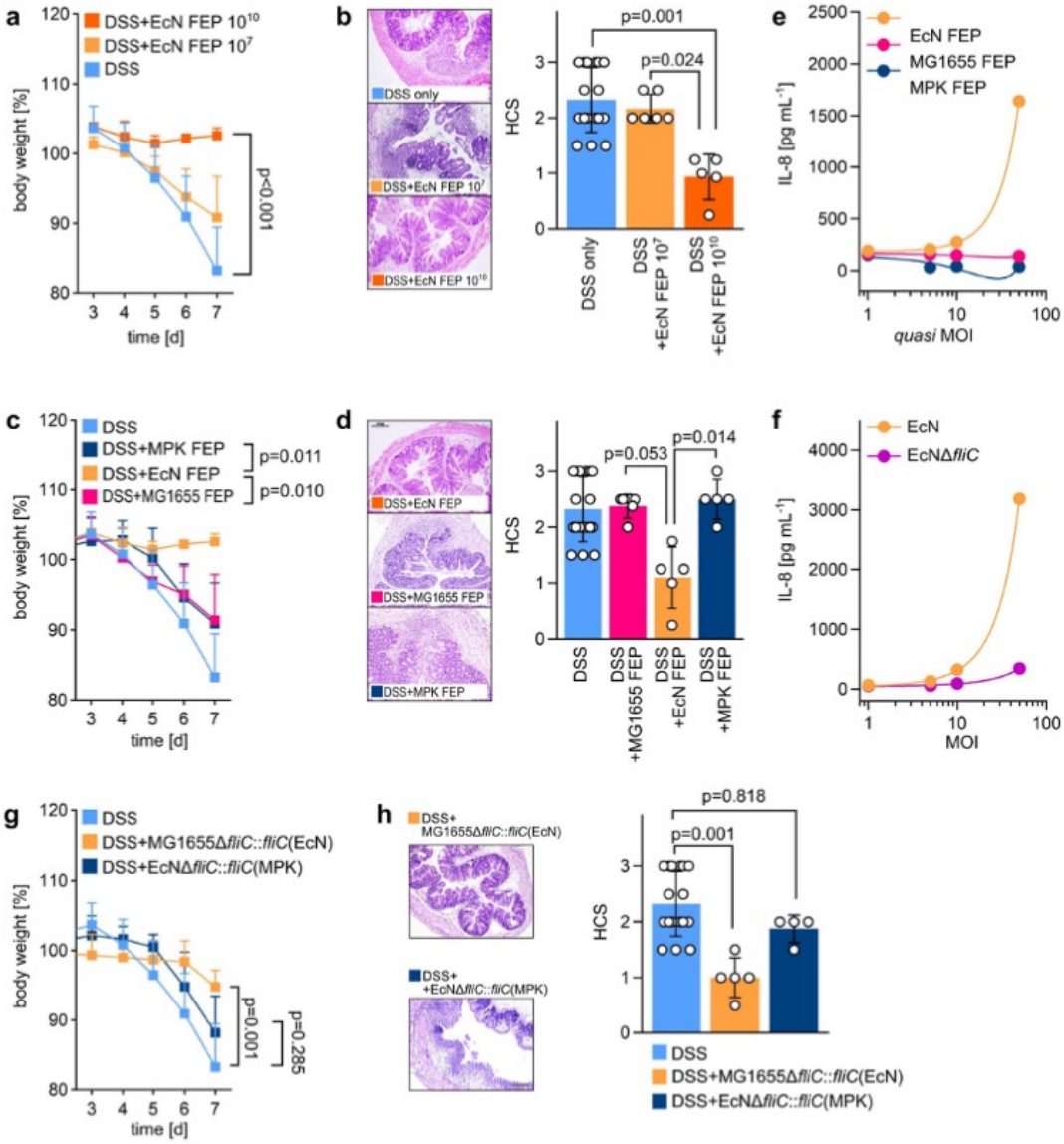



**Fig 2. Flagella-dependent influence of different commensal *E. coli* strains on DSS-induced colitis in WT mice.** (a + b) SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with FEPs obtained from $10^7$ (EcN FEP $10^7$) and $10^{10}$ (EcN FEP $10^{10}$) EcN per 100 mL drinking water. (a) Change in body weight relative to start of DSS administration at day 0. (b) Left panel: HE-stained colonic sections at day 7 after start of DSS administration. Right panel: HCS at day 7. (c + d) SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with FEPs obtained from $10^{10}$ EcN, MG1655, or MPK per 100 mL drinking water. (c) Change in body weight relative to start of DSS administration at day 0. (d) Left panel: HE-stained colonic sections at day 7 after start of DSS administration. Right panel: HCS at day 7. (e) mTLR5-HEK293 cells were stimulated with FEP obtained from EcN, MG1655, and MPK for 24 h. FEPs were generated from the number of bacteria corresponding to a certain MOI (quasi-MOI). Resulting IL-8 secretion into cell supernatant as a result of TLR5 receptor activation was detected by ELISA. (f) mTLR5-HEK293 cells were stimulated with EcN and EcNΔ*fliC* at different MOI for 24 h. Resulting IL-8 secretion into cell supernatant as a result of TLR5 receptor activation was detected by ELISA. (g + h) SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with $10^{10}$ viable bacteria of the indicated

complementation mutants per 100 mL DSS-containing drinking water. (g) Change in body weight relative to start of DSS administration at day 0. (h) Left panel: HE-stained colonic sections at day 7 after start of DSS administration. Right panel: HCS at day 7. Statistics: (a), (c), (g), one-way ANOVA with Tukey multiple comparison test; (b), (d), (h), Kruskal–Wallis test with multiple comparisons. $p$-values $< 0.05$ are considered to represent statistical significance. The data underlying this figure can be found in S1 Data. DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; FEP, flagella-enriched preparation; *fliC*, flagellin; HCS, histological colitis score; HE, hematoxylin–eosin;; IL, interleukin; MG1655, *E. coli* K12 MG1655; MOI, multiplicity of infection; MPK, *E. coli* mpk; mTLR5-HEK293 cell, mouse-TLR5–expressing human embryonic kidney 293 cell; SPF, specific-pathogen–free; TLR, Toll-like receptor; WT, wild type.

https://doi.org/10.1371/journal.pbio.3000334.g002

any beneficial effects concerning DSS-induced colitis in WT mice, indicating that the molecular nature of flagellin impacted on this process.

To assess whether this was reflected at the level of TLR5 activation, we used human embryonic kidney cells overexpressing mouse TLR5 (mTLR5-HEK293), stimulated them with either viable bacteria or FEPs, and measured the resulting interleukin (IL)-8 secretion to quantify TLR5-dependent NFκB activation. When comparing EcN FEP, MG1655 FEP, and MPK FEP, the latter FEPs from nonsymbiotic bacteria completely failed to activate the mouse TLR5 receptor in any tested concentration, while EcN FEP provided a concentration-dependent signaling intensity, as demonstrated by differential NFκB-activation–dependent IL-8 secretion from mTLR5-HEK293 cells (Fig 2E). Of note, a strong effect on NFκB activation was also observed for viable EcN but not EcNΔ*fliC* (Fig 2F). Thus, we assumed that flagellin of symbiotic EcN might feature a distinct property that is absent in nonsymbiotic *E. coli* strains. To explore this in detail, different complementation mutants were generated: (1) an MG1655Δ*fliC* mutant was complemented with the *fliC* gene from EcN (MG1655Δ*fliC*::*fliC*(EcN)), and (2) an EcNΔ*fliC* mutant was complemented with the *fliC* gene from MPK (EcNΔ*fliC*::*fliC*(MPK)) (see S5 Fig for details on complementation). Both mutants provided efficient flagella expression and adequate mobility as demonstrated by electron microscopy and bacterial swarming assays (S6 Fig), indicating that the exchange mutation did not negatively affect flagellum expression and function. Both complemented viable strains were administered to DSS-treated mice, and the disease outcome was monitored as before (Fig 2G and 2H). Although a slight reduction in body weight was observed in DSS-treated animals that were administered MG1655Δ*fliC*::*fliC*(EcN) (Fig 2G), the damage of colonic tissue was significantly lower compared to DSS-treated mice without additional bacterial administration (Fig 2H). However, administration of EcNΔ*fliC*::*fliC*(MPK) did not provide any beneficial effects on DSS-induced colitis, as indicated by strong weight loss and increased HCS (Fig 2G and 2H). Therefore, we concluded that the inflammation-reducing features of the EcN flagella are specific for this strain and are not present in the flagella of MG1655 or MPK. However, this raises the question of which structural features might account for the differences of the EcN FliC protein compared to FliC from nonsymbiotic *E. coli* strains.

## Insertions within the HVR of flagellin account for the symbiotic effects of EcN compared to nonsymbiotic *E. coli* strains

FliC proteins consist of a constant region comprising one N- and one C-terminal D0 and D1 domain (NTD0, NTD1, CTD0, and CTD1). Some bacteria—e.g., *Escherichia* and *Salmonella* strains—additionally contain an HVR composed of a C- and N-terminal D2 (NTD2, CTD2) domain as well as a central D3 domain (Fig 3A). As expected, sequence alignment comparisons between FliC of EcN, MG1655, and MPK revealed that in the constant region, the sequence similarity of the 4 domains exceeds 95% (Fig 3B and 3C). This constant region is thought to be the primary mediator of TLR5 activation [30,28], However, sequence similarity of the N- and C-terminal D2 domains within the HVR is 65% (NTD2) and 64% (CTD2) between EcN FliC and MG1655 FliC, as well as 61% (NTD2) and 64% (CTD2) in the case of

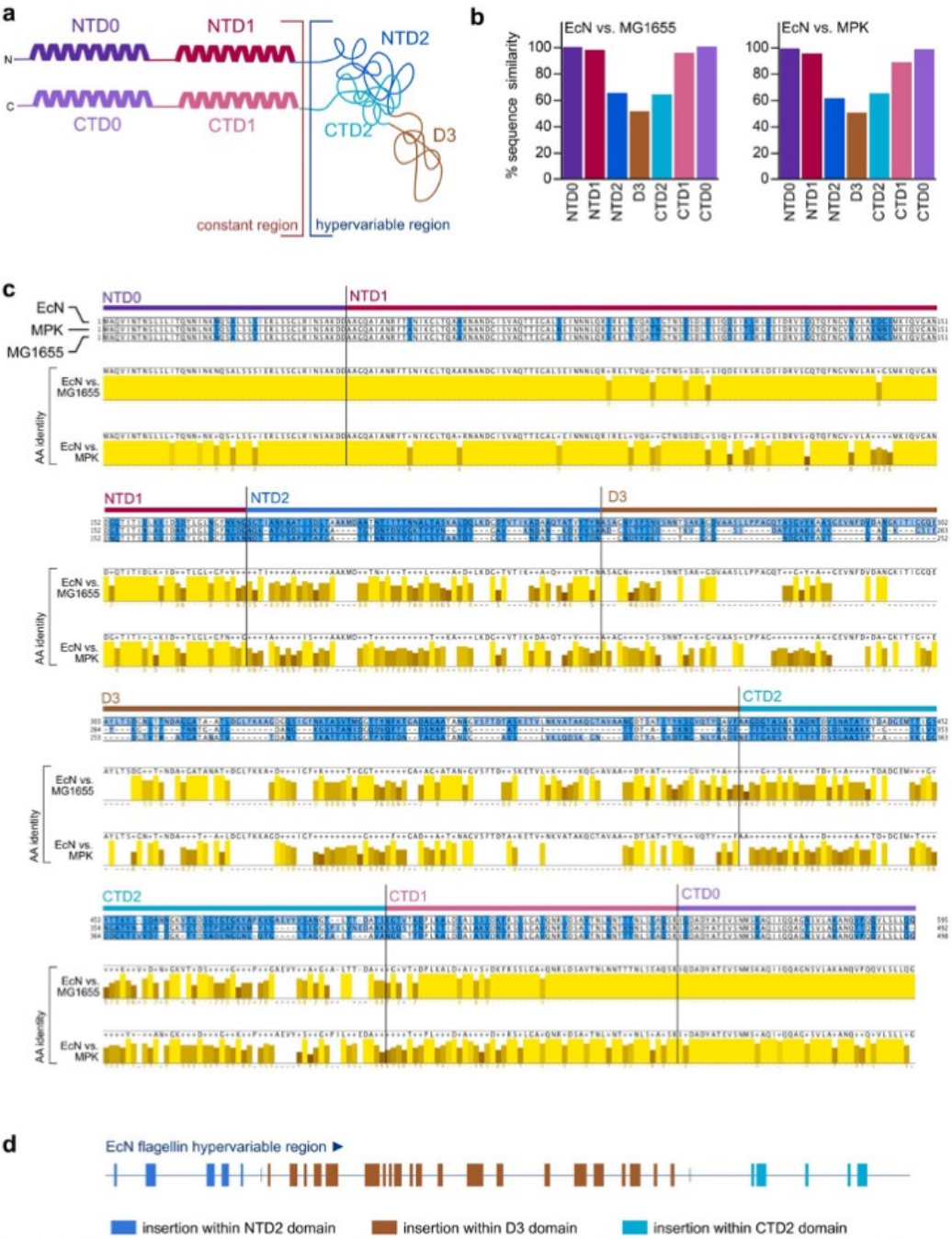



**Fig 3. Detailed comparison of the flagellin amino-acid sequences of different *E. coli* strains.** Protein alignment of FliC proteins from EcN (CCQ05465.1), MPK, and MG1655 (NP_416433.1) genomes. (a) Schematic structure of flagellin according to Yonekura and colleagues [50,51]. D0

and D1 comprise conserved N- and C-termini of *fliC*, packed into α-helical structures in the filament core. The NTD2/CTD2- and D3-domain–containing HVR is attached adjacent to the D2 domains, located at the outer surface of the filament. (b) Quantification of amino-acid sequence similarities of all 6 flagellin domains. Computation was performed using similarity indices depicted in 3c. (c) The alignment was generated using MAFFT. Amino acids were colored by overall conservation (white: fully conserved, light blue to dark blue: high to low conservation). Consensus sequence and sequence conservation for pairwise comparisons of EcN to MPK and MG1655, respectively, are shown in yellow. Darker shades of yellow represent lower sequence conservation. (d) Schematic overview of the differences of the EcN flagellin HVR compared to the HVRs of both MG1655 and MPK. Sequences (insertions) that are only present in EcN HVR and not in MPK HVR or MG1655 HVR are highlighted as colored squares. AA, amino acid; CTD, C-terminal domain; EcN, *E. coli* Nissle 1917; *fliC*, flagellin; HVR, hypervariable region; MAFFT, Multiple Alignment using Fast Fourier Transform; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk; NTD, N-terminal domain.

https://doi.org/10.1371/journal.pbio.3000334.g003

comparing EcN FliC with MPK FliC. The differences between D3 domain sequences is even lower than 50% sequence similarity for both comparisons (Fig 3B and 3C). These low numbers in relative sequence similarity are mainly due to the presence of numerous sets of amino-acid inserts within the HVR of EcN flagellin compared to the corresponding regions of MG1655 flagellin and MPK flagellin (Fig 3D), thus rendering the NTD2, D3, and CTD3 domains of EcN substantially longer. However, not only the length but also the amino-acid sequence of the HVR is highly different, both in comparison between EcN and MG1655 or EcN and MPK and also between MG1655 and MPK.

## Deletions within the HVR of FliC leads to loss of EcN inflammation-reducing properties

In order to check whether the different HVR structure is causative for the observed disparities in influencing the progress of DSS-induced colitis, we deleted large parts of the EcN *fliC* HVR, namely a short part of the NTD2 domain as well as large parts of D3 and the entire CTD2 domain (EcNΔ*fliC*(HVR)) (Fig 4A). This deletion resulted in a significantly shorter assembled flagellum compared to WT EcN, providing only about 10% of the length of WT EcN flagella (S7 Fig). We administered EcNΔ*fliC*(HVR) as a viable bacterium to DSS-treated mice and checked for changes in body weight, followed by histological analysis of the colonic tissue 7 days after the start of DSS treatment. As demonstrated in Fig 4B and 4C, there were no detectable inflammation-preventing effects when the EcNΔ*fliC*(HVR) deletion mutant was used, as indicated by strong weight loss (Fig 4B) and high HCSs (Fig 4C), which was in sharp contrast to WT EcN (Fig 1A–1D). In conclusion, these experiments show that insertions within the flagellin HVR are responsible for the disease-ameliorating properties of EcN flagellin. To gain insight into the effect of HVR modulation on systemic inflammation, we determined serum concentrations of 13 pro- and anti-inflammatory cytokines in DSS + EcN-treated, DSS + EcNΔ*fliC*(HVR)-treated, and DSS-only–treated mice to characterize the influence of the flagellin HVR on the cytokine-associated protective properties of EcN in this mouse model. Consistent with reports showing that DSS-treated mice provide increased serum levels of, e.g., tumor necrosis factor (TNF)α, IL-6, IL-1β, and IL-17 compared to healthy non-DSS–treated controls [52,53], colonic inflammation led to a systemic increase of proinflammatory cytokines, which translated into elevated serum concentrations. Fig 4E depicts a heat map of all detected cytokines in each mouse of the three respective groups. We detected significantly higher serum levels of the proinflammatory cytokines IL-1β, IL-1α, TNFα, IL-6, and IL-17A as well as of the anti-inflammatory cytokines IL-27, IL-10, and interferon (IFN)β in DSS-only– and DSS + EcNΔfliC(HVR)-treated mice compared to DSS + EcN-treated mice (Fig 4E, see also S1 Table for detailed statistical analysis). Therefore, we concluded that the presence of the full HVR is a precondition for the EcN-mediated symbiotic effects during DSS-induced colitis in WT mice, and its deletion results in loss of protection against intestinal inflammation.

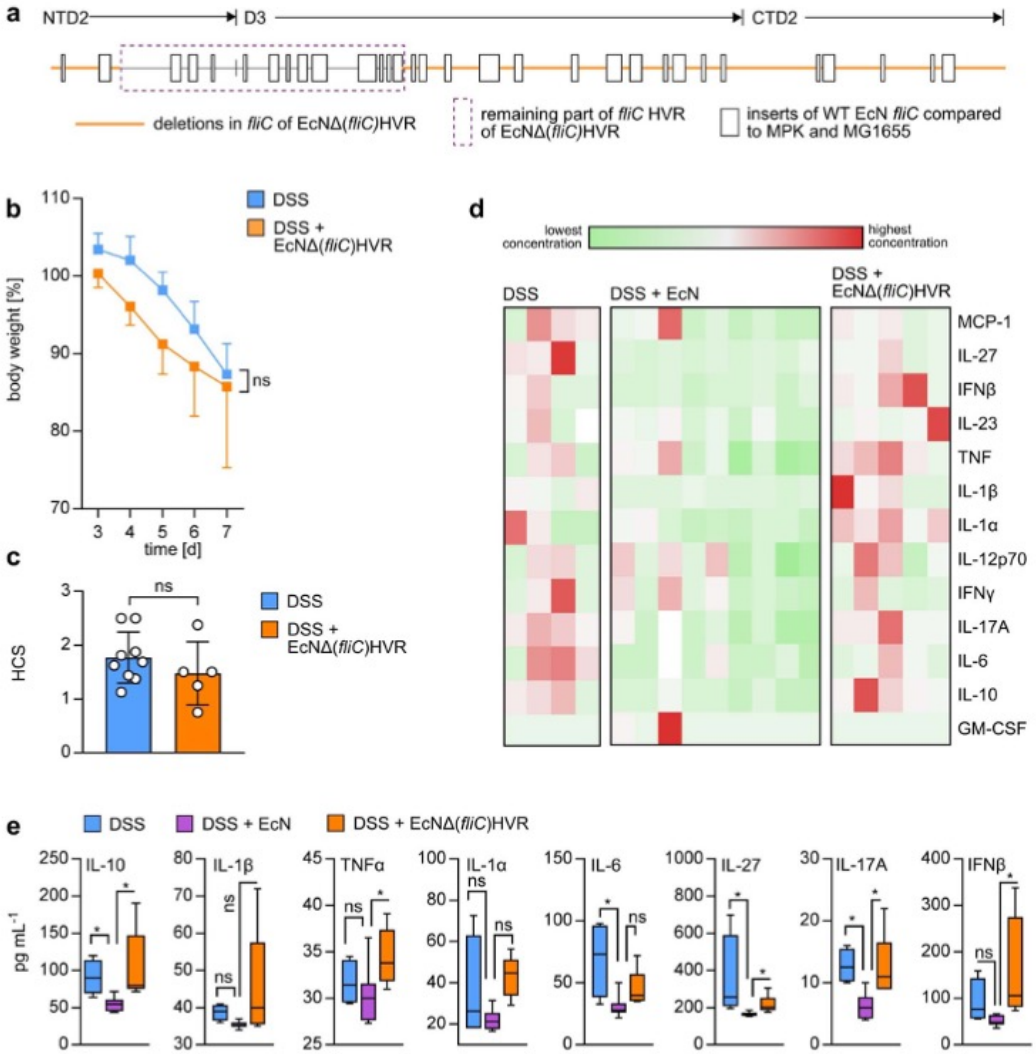



**Fig 4. Partial deletion of the EcN HVR leads to loss of EcN probiotic effects.** (a) Schematic display of the deleted parts in the EcNΔ(*fliC*)HVR mutant. The orange line depicts the parts of the EcN HVR that were deleted. (b + c) SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with $10^{10}$ viable bacteria of EcN or the EcNΔ(*fliC*)HVR mutant resuspended in 100 mL DSS-containing drinking water. (b) Change in body weight relative to start of DSS administration at day 0. (c) HCS at day 7. Statistical analysis was performed using Student *t* test. Error bars represent SD. White dots in column bars represent each biological replicate. (d) Heat map of cytokine concentrations in serum from DSS-treated mice. Each column represents a different individual. (e) Detailed analysis of cytokine concentrations from serum depicted in (d). Asterisks indicate statistical significance. All shown cytokines provide the strongest differences between DSS-treated and DSS + EcNΔ*fliC*(HVR)-treated as well as between DSS + EcN and DSS + EcNΔ*fliC*(HVR)-treated groups. See S1 Table for detailed statistical analysis using Kruskal–Wallis test with multiple comparisons. *p*-values < 0.05 are considered to represent statistical significance. (b + c + e) The data underlying this figure can be found in S1 Data. CTD, C-terminal domain; DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; *fliC*, flagellin; GM-CSF, granulocyte-macrophage colony-stimulating factor; HCS, histological colitis score; HVR, hypervariable region; IFN, interferon; IL, interleukin; MCP, monocyte chemoattractant protein; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk; NTD, N-terminal domain; ns, not significant; SPF, specific-pathogen–free; TNF, tumor necrosis factor; WT, wild type.

https://doi.org/10.1371/journal.pbio.3000334.g004

## EcN-flagellin–induced protective effects are mediated by host TLR5$^+$CD11c$^+$ cells in the colonic LP

So far, we have demonstrated that symbiotic EcN mediated at least a remarkable part of its symbiotic properties through its flagellin HVR. This effect was dependent on TLR5 expression in the host. Therefore, we wanted to further elucidate which TLR5-expressing host cell population was pivotal for the mediation of the EcN-flagellin–induced effects. The crucial contribution of TLR5 expression on intestinal epithelial cells (IECs) for maintenance of intestinal homeostasis has already been reported [37]. Colonic IECs do not only express TLR5 on their basolateral side [54]. A recent publication demonstrated that colonic IECs also express TLR5 on the luminal side, thus secreting antimicrobial peptides and cytokines in response to flagellin recognition [55]. However, TLR5-expressing intestinal LP dendritic cells (DCs) also contribute to flagellin-mediated immune responses [37,56]. Therefore, we were interested in which intestinal cell population mediated the observed symbiotic flagellin-induced inflammation-silencing properties. Since IECs belong to the group of stromal cells and antigen-presenting cells such as DCs are derived from hematopoietic stem cells, one can investigate the distinct influence of these two differentially originated cell types by using bone-marrow–chimeric mice (BMCM) (Fig 5A). Therefore, we generated different groups of BMCM: (1) WT BL/6 recipient mice transplanted with bone marrow from *Tlr5*$^{-/-}$ donor mice (*Tlr5*$^{-/-}$ → WT), (2) *Tlr5*$^{-/-}$ recipient mice transplanted with bone marrow from WT BL/6 donor mice (WT → *Tlr5*$^{-/-}$), and (3) as controls, WT BL/6 recipient mice transplanted with bone marrow from WT BL/6 donor mice (WT → WT). All these animals were SPF-housed and treated with DSS and EcN, DSS and EcN*ΔfliC*, or DSS only. As demonstrated in Fig 5B, WT → *Tlr5*$^{-/-}$ mice that were administered EcN provided low HCS, indicating low colonic inflammation, while the protective effect of EcN was completely abolished using *Tlr5*$^{-/-}$ → WT mice (Fig 5B). As seen in WT BL/6 mice (Fig 1), both groups of BMCM developed severe colitis symptoms when administered DSS and the EcN*ΔfliC* deletion mutant. Therefore, we concluded that the protective effect of EcN flagellin is mainly mediated by cells of hematopoietic origin. CD11c$^+$ cells such as DCs are one major cell type from the hematopoietic cell lineage mediating microbiota-derived anti-inflammatory processes in the intestine [41,57]. Therefore, we had a closer look at CD11c$^+$ cells in the colonic LP (cLP) and their relevance for EcN-mediated symbiotic effects during DSS-induced colitis.

We generated BMCM restricting TLR5 deficiency largely to CD11c$^+$ cells with a substantial amount of other hematopoietic cells still expressing TLR5 (*Tlr5*$^{-/-}$ + ΔCD11c → WT). By using *Tlr5*$^{-/-}$ and ΔCD11c donor mice for bone marrow transplantation, we ensured that at least half of hematopoietic cells other than CD11c$^+$ cells do express TLR5. Generation of these BMCM was strictly validated (S8 Fig). As demonstrated in Fig 5C, this group provided significantly higher inflammation when treated with DSS and EcN compared to WT → WT control animals that were treated equally. This finding strongly supports the idea that TLR5$^+$CD11c$^+$ cells in the cLP are the main mediators of the symbiotic effects caused by EcN flagellin.

## Administration of symbiotic recombinant flagellin prevents intestinal inflammation

According to our results, we proposed flagellin from a symbiotic *E. coli* strain to be a suitable agent in order to prevent pathological intestinal inflammation that was supposed to be rooted in its increased TLR5 activation capacity. To finally provide evidence for this hypothesis, we aimed to administer recombinant flagellin from symbiotic and nonsymbiotic *E. coli* strains to DSS-treated WT mice. Therefore, we used recombinant flagellin from symbiotic EcN (rfliC (EcN)) and from nonsymbiotic MG1655 (rfliC(K12)). While rfliC(K12) was commercially

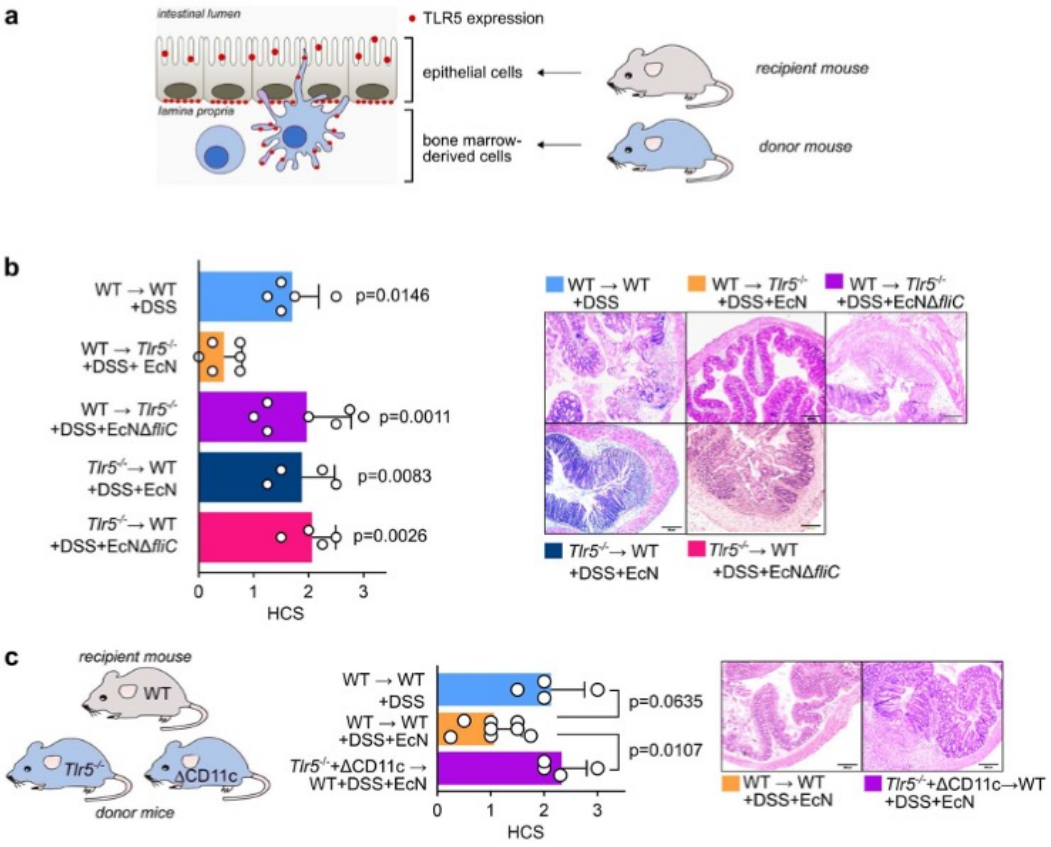



**Fig 5. EcN-flagellin–induced protective effects are mediated by host TLR5$^{+}$CD11c$^{+}$ cells in the cLP.** (a) Schematic depiction of the generation of BMCM as described in the experimental procedures. Red dots represent the typical sites of TLR5 expression in the mouse intestine. (b) Different SPF BMCM were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with $10^{10}$ viable bacteria of EcN or the EcNΔ(*fliC*) HVR mutant resuspended in 100 mL DSS-containing drinking water. See text for further information on the nomenclature of distinct BMCM groups. Indicated *p*-values refer to the comparison of the respective data set with the WT → *Tlr5*$^{-/-}$ + DSS + EcN group. Upper panel: HCS at day 7. Lower panel: HE-stained colonic sections at day 7 after start of DSS administration. (d) Irradiated WT recipient mice were transplanted with bone marrow from *Tlr5*$^{-/-}$ and ΔCD11c donor mice in a 1:1 ratio. Mice were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with $10^{10}$ viable bacteria of EcN resuspended in 100 mL DSS-containing drinking water and compared to control groups. See text for further information on the BMCM groups nomenclature. Middle panel: HCS at day 7. Right panel: HE-stained colonic sections at day 7 after start of DSS administration. Statistics: (b) One-way ANOVA with Tukey multiple comparison test, (c) Mann–Whitney test, (d) Kruskal–Wallis test with multiple comparisons, *p*-values < 0.05 are considered to represent statistical significance. (b + c) The data underlying this figure can be found in S1 Data. BMCM, bone-marrow–chimeric mice; cLP, colonic LP; DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; *fliC*, flagellin; HCS, histological colitis score; HE, hematoxylin–eosin; HVR, hypervariable region; LP, lamina propria; SPF, specific-pathogen–free; TLR, Toll-like receptor; WT, wild type.

https://doi.org/10.1371/journal.pbio.3000334.g005

available, rfliC(EcN) was generated and quality-checked as described in the supplementary material (S9 and S10 Figs). First, we checked for their ability to induce TLR5 receptor activation. As seen in Fig 6A, both recombinant flagellins induced IL-8 secretion in mTLR5-HEK293 cells in a concentration-dependent manner. However, rfliC(EcN) induced higher IL-8 secretion compared to rfliC(K12) at a medium-range concentration of 100 ng mL$^{-1}$, thus indicating stronger TLR5 receptor activation. Next, we aimed to test the inflammation-preventing properties of both recombinant flagellins during DSS-induced colitis in mice. Therefore, we used

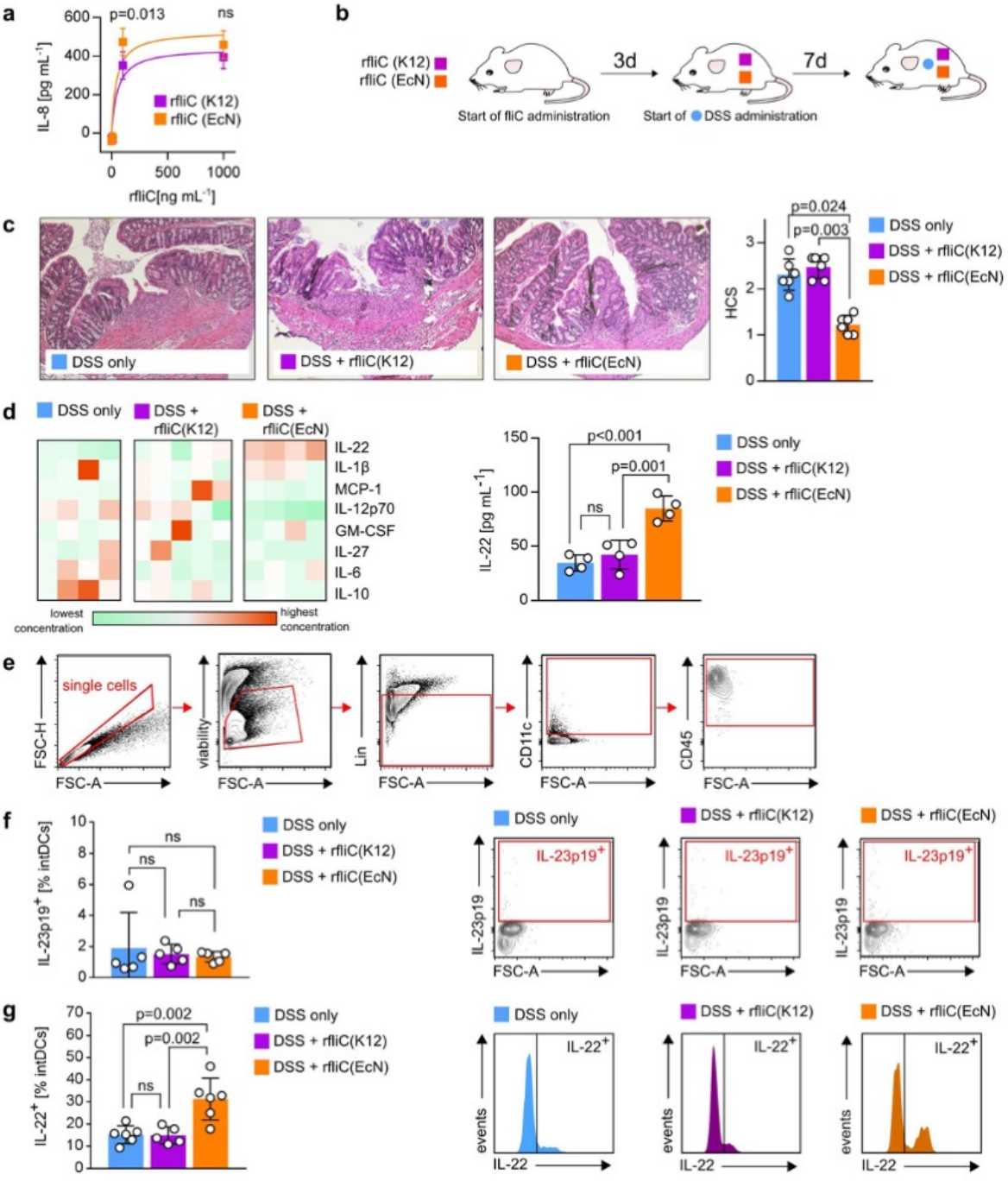

a
p=0.013
ns
IL-8 [pg mL-1]
rfliC (K12)
rfliC (EcN)
rfliC[ng mL-1]
b
rfliC (K12)
rfliC (EcN)
3d
7d
Start of fliC administration
Start of DSS administration
c
DSS only
DSS + rfliC(K12)
DSS + rfliC(EcN)
p=0.024
p=0.003
HCS
d
IL-22
IL-1β
MCP-1
IL-12p70
GM-CSF
IL-27
IL-6
IL-10
lowest concentration
highest concentration
p<0.001
p=0.001
ns
IL-22 [pg mL-1]
e
single cells
FSC-H
FSC-A
viability
Lin
CD11c
CD45
f
IL-23p19+ [% intDCs]
IL-23p19+
IL-23p19
g
IL-22+ [% intDCs]
p=0.002
IL-22+
events
IL-22

**Fig 6. Recombinant FliC protects against DSS-induced colitis.** (a) mTLR5-HEK293 cells were stimulated with rfliC(K12) or rfliC(EcN) for 24 h. Resulting IL-8 secretion into cell supernatant as a result of TLR5 receptor activation was detected by ELISA. (b) Experimental setup: SPF C57BL/6 WT mice aged 6 to 8 weeks were administered 2 μg recombinant flagellin daily via intragastral gavage. 3 days after start of flagellin administration, 3.5% DSS was added to the drinking water. Progress of DSS-induced colitis was monitored for additional 7 days. (c) HCS and representative HE-stained colonic sections at day 7. Statistical analysis was performed using the Kruskal–Wallis test. Error bars represent SD. White dots in column bars represent each biological replicate. (d) Heat map of cytokine concentrations in serum from DSS-treated mice as shown in (b). Each column represents a different individual. Right panel: IL-22 concentration in blood serum. Statistical analysis was performed using one-way ANOVA. Error bars represent SD. White dots in column bars represent each biological replicate. (e) Gating strategy to define the population of intDCs from the cLP. Lin = Ly6G/C, CD45R, CD64. (f + g) Proportion of IL-23$^+$ (f) or IL-22$^+$ (g) intDCs from the experiment shown in (b). Representative histograms or contour blots are shown. Statistical analysis was performed using one-way ANOVA. Error bars represent SD. White dots in column bars represent each biological replicate. *p*-values $< 0.05$ are considered to represent statistical significance. (a + c + d + f + g) The data underlying this figure can be found in S1 Data. cLP, colonic LP; DC, dendritic cell; DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; *fliC*, flagellin; FSC, forward scatter; HCS, histological colitis score; HE, hematoxylin–eosin; IL, interleukin; intDC, intestinal DC; LP, lamina propria; MG1655, *E. coli* K12 MG1655; mTLR5-HEK293 cell, mouse-TLR5–overexpressing human embryonic kidney 293 cell; ns, not significant; rfliC(EcN), recombinant flagellin from EcN; rfliC(K12), recombinant flagellin from MG1655; SPF, specific-pathogen–free; TLR, Toll-like receptor.

https://doi.org/10.1371/journal.pbio.3000334.g006

WT SPF C57BL/6 mice and started daily intragastral gavage of 2 μg recombinant flagellin 3 days prior to start of DSS administration. DSS was administered as 3.5% solution in drinking water, and flagellin administration was continued daily until the end of the experiment, 7 days after initial DSS exposure (Fig 6B). The non-flagellin–administered DSS-only–treated control group was gavaged daily with 100 μL sterile PBS, the solvent flagellin was reconstituted in. As demonstrated in Fig 6C, administration of rfliC(EcN) resulted in significantly lower tissue damage in the colon compared to rfliC(K12)-treated and DSS-only–treated mice, therefore indicating prevention of intestinal inflammation. Lower HCSs in rfliC(EcN)-treated mice were associated with decreased concentrations of various proinflammatory cytokines in the blood serum. However, we detected increased serum levels of IL-22 in rfliC(EcN)-treated mice compared to both other groups (Fig 6D). Since IL-22 contributes to maintenance of intestinal immune homeostasis by strengthening the intestinal epithelial barrier, we assumed that EcN-flagellin–induced stronger TLR5-activation was associated with increased IL-22 expression. The main sources of intestinal IL-22 are reported to be innate lymphoid cells (ILC3s) and CD4$^+$ T cells, which secrete IL-22 in response to IL-23 [58,59]. Since CD11c is expressed on various different cell types in the mouse intestine, we further focused on intestinal DCs. Intestinal DCs (intDCs) were defined as being Ly6G/Ly6C$^{neg}$, CD45R/B220$^{neg}$, CD64$^{neg}$, CD45$^{pos}$, and CD11c$^{pos}$ (Fig 6E). Focusing on intDCs in the cLP, however, did not reveal any differences in IL-23 expression between all compared groups (Fig 6F). This prompted us to check whether intDCs themselves might be the source of IL-22. In fact, we detected significantly increased proportions of IL-22$^+$ cells in rfliC(EcN)-treated mice compared to both other groups (Fig 6F), which correlated with a lower HCS (Fig 6C). Therefore, we assumed that enhanced IL-22 expression by these cells contributed to symbiotic-flagellin–mediated intestinal immune homeostasis maintenance in DSS-induced colitis.

## Discussion

*E. coli* represents one of the most intensely studied intestinal commensals. Colonization of the human gut with *E. coli* starts early after birth even before colonization with anaerobes, which usually make up the largest part of the human intestinal microbiota [60]. However, distinct commensal *E. coli* strains can provide completely different immunogenic properties in a certain host's intestine. While pathobiotic commensal *E. coli* strains may mediate pathophysiological immunological processes in a predisposed host, symbiotic commensal *E. coli* strains may exhibit even strong probiotic features. Symbionts with beneficial features are widely used to restore gut homeostasis in human patients and animal models [61]. One of the most prominent and therefore most extensively studied symbiotic *E. coli* strains is EcN. However, its symbiotic properties have not yet been linked to a certain molecular or structural feature.

Interestingly, studies using outer-membrane vesicles from EcN have revealed that these vesicles are able to mimic the properties of viable EcN, thus indicating that any surface-associated component might be crucial for the mediation of its symbiotic capacities [62]. An important surface-associated molecule of EcN is flagellin, the constitutive protein of the bacterial flagellum. Most clinically relevant *E. coli* strains express flagella on the cell surface [63], thus rendering bacteria motile and facilitating bacterial attachment to the intestinal mucus [46]. Since flagellin represents one of the most important antigens in IBD patients [33], we assumed that the definite flagellin structure might contribute to the symbiotic properties of certain *E. coli* strains and is therefore involved in modulation of IBD pathogenesis.

With this study, we found the HVR of flagellin to strongly influence the immunogenic properties of the tested commensal *E. coli* strains. Unexpectedly, we observed that symbiotic EcN harbors a substantially longer HVR, comprised of various small inserts, compared to nonsymbiotic commensal *E. coli* strains. Previous work on structure-function relationship for flagellin–TLR5 interactions have focused on the so-called N- and C-terminal D0 and D1 domains of flagellin that make up the protein's constant region and that are highly conserved among all flagellins across different bacterial phyla [50,51]. This constant region was also characterized as being crucial for interaction with TLR5 and therefore for TLR5-mediated intracellular signaling in host cells [29,30]. Given the high similarity of EcN, MPK, and MG1655 in this region, it was therefore surprising to detect a significant difference between symbiotic and nonsymbiotic strains in the ability to induce TLR5-mediated signaling. Flagellin of EcN, harboring the substantially longer HVR, induced significantly stronger TLR5-mediated signaling compared to the tested nonsymbiotic *E. coli* strains. This indicated an unexpected contribution of the HVR D2 and D3 domains to TLR5 activation.

Andersen-Niessen and colleagues reported that flagellin's N-terminal D1 domain predominantly determined its TLR5-stimulatory activity. However, this required additional contribution from the HVR D2/D3 and the CD1 domain [64]. This might indicate that HVR domains generally have a stronger impact on TLR5 activation than could have been gleaned from only the crystal structure of TLR5 and flagellin [30]. This structure shows zebrafish TLR5 in complex with *Salmonella* FliC, out of which 100 amino acids in the HVR of *Salmonella* are not structurally resolved. This indicates that parts of the HVR might be highly flexible and/or do not stably interact with TLR5, at least with zebrafish TLR5. The situation may be different for the murine TLR5 that was investigated here.

Since the D3 domain of *Salmonella* flagellin contributes to the stability of flagellin monomers [65], which is a prerequisite for TLR5 receptor activation, it might be possible that the HVR of symbiotic *E. coli* strains also positively regulates monomer stability, whereas mucus attachment, another function of flagellin, has been shown to be independent of the D3 region [46] and thus probably not relevant here. Although the precise molecular mechanism by which the HVR modulates TLR5 signaling thus remains to be established, our functional data warrant a more thorough exploration of this region, both functionally and structurally.

While extracellular flagellin is sensed by TLR5, leading to MyD88-dependent NFκB activation, flagellin present in the cytosol results in activation of the NLR Family CARD Domain Containing 4 (NLRC4) inflammasome [32]. NLRC4 activation leads to IL-1β and IL-18 secretion, and the NLRC4 inflammasome helps to discriminate harmful pathogens from beneficial commensals [66]. However, we assume that potential NLRC4-inflammasome–mediated effects play a neglectable role concerning the discrimination of symbiotic from nonsymbiotic *E. coli* strains because of two reasons: first, flagellin recognition by the NLRC4 inflammasome component NLR family, apoptosis inhibitory protein 5 (NAIP5) is crucially mediated by 35 amino acids within the flagellin CD0 domain [67], which is identical in all tested strains. Secondly, we demonstrated that flagellin-HVR–mediated symbiotic effects were completely abolished in

TLR5-deficient mice and FEPs were sufficient to mediate symbiotic properties, therefore indicating that TLR5-dependent sensing of extracellular flagellin is far more important than a potential intracellular NLRC4 inflammasome activation in our system.

But how can it be explained that a stronger intestinal TLR5 signaling, induced by a symbiotic commensal, correlates with beneficial effects during DSS-induced colitis accompanied by lower systemic cytokine levels? Systemic loss of TLR5 signaling was shown to entail an overgrowth of flagellated members of the intestinal microbiota, promoting inflammatory conditions [48,56]. Furthermore, certain SNPs in the human *TLR5* gene are associated with higher incidence of IBDs [68,69]. Therefore, TLR5 expression and signaling seem to be necessary to maintain a balanced homeostatic microbiota composition and intestinal immune homeostasis. This effect has been mostly traced back to TLR5 signaling in IECs, which is suspected to lead to immune cell recruitment, which mediates clearance from bacteria breaching the mucus barrier [37]. However, using BMCM, we have identified intestinal LP $CD11c^+$ cells as crucially contributing to the EcN-flagellin–mediated symbiotic effects. This observation is in line with findings that TLR5 on hematopoietic cells is involved in flagellin sensing [70] and flagellin-dependent activation of Th17 cells [56]. Furthermore, SNP-mediated differences in TLR5 signaling were observed to affect immune cells rather than epithelial cells.

Besides IECs, $CD11c_+$ cells such as DCs are the most important TLR5-expressing cells at intestinal mucosal interfaces [71]. We demonstrated that symbiotic-flagellin–mediated beneficial effects on the progress of DSS-induced colitis was associated with increased serum levels of IL-22 as well as with a higher proportion of $IL\text{-}22^+$ intDCs. In general, IL-22 contributes to restoration of intestinal homeostasis, promotes regeneration of damaged intestinal epithelium [72], modulates epithelial cell fucosylation [73], and induces the secretion of antimicrobial peptides [74,75] and mucins [76]. Therefore, it represents a pivotal cytokine in modulation of intestinal tissue responses during inflammatory processes. In intestinal tissue, IL-22 is mainly produced by ILC3 cells [59], $CD4^+$ T cells [77], and, to a lesser extent, $CD8^+$ T cells, T-cell receptor (TCR)γδ T cells, neutrophils, and natural killer (NK) cells [78,79]. ILC3 and $CD4^+$ T cells secrete IL-22 mainly in response to IL-23 [58,59,77]. In this connection, intDCs play a decisive role in shaping the LP cytokine milieu in response to intestinal microbiota sensing. However, even intDCs themselves emerged as sources of gastrointestinal IL-22 [75,80,81] which is in line with our observations. Therefore, we hypothesize that symbiotic flagellin, bearing a longer HVR, leads to stronger TLR5 signaling in intDCs, resulting in enhanced IL-22 expression and thus contributing to maintenance of the intestinal barrier. Since disruption of the intestinal barrier provides a characteristic of DSS-induced colitis, IL-22 was demonstrated to play a pivotal role in counteracting inflammatory processes in this disease model [81–83]. However, and to date, we cannot completely exclude a merely indirect contribution of intDCs by modulating IL-22 secretion through ILC3 cells or T cells. This indirect influence might occur via secretion of ILC3- and T-cell–activating cytokines by $CD11c^+$ cells other than intDCs. Nevertheless, our results indicate intDCs to be, at least partially, direct sources of IL-22 in response to different flagellins, which, in turn, might account for the observed distinct DSS-induced colitis phenotypes mediated by administration of different flagellins. Thus, paradoxically, a symbiotic-flagellin–mediated stronger TLR5 signaling based on its flagellin HVR structure might contribute to the observed lower inflammatory status of symbiotic-flagellin–administered DSS-treated mice.

Interestingly, flagellin-dependent TLR5 signaling was also demonstrated to be involved in the mediation of symbiotic properties of another gut commensal, *Roseburia hominis* [84], and may thus be a more common phenomenon than previously thought. However, the role of certain cell subtypes within the intestinal tissue mediating the decisive beneficial effects in response to flagellin still remains to be further elucidated. While we propose intDCs from the

cLP to be the most important cell type to promote inflammation-preventing events during DSS-induced colitis, Uematsu and colleagues demonstrated $CD11c^+$ cells from Peyer's Patches (PPs) in the small intestine to be responsible for flagellin-induced IL-10 secretion, which might, in turn, contribute to homeostasis, while LP $CD11c^+$ cells instead promote secretion of proinflammatory cytokines [85]. Therefore, both studies accentuate the contribution of intestinal $CD11c^+$ cells on host immune homeostasis by sensing flagellin from luminal microbes, even though different $CD11c^+$ cell populations emerged as crucial. Since our study focused on DSS-induced colitis, we did not further investigate cells from noncolonic gastrointestinal tissues. Uematsu and colleagues, on the other hand, did not investigate homeostasis-promoting mechanisms other than IL-10 secretion. Thus, both results are not necessarily contradictory. Nevertheless, more studies have to be conducted to uncover the role of CD11c+ cells all across the gastrointestinal tract in terms of flagellin-mediated immunological reactions.

In addition to local intestinal effects, the *E. coli*-flagellin-HVR–mediated impact on the progress of DSS-induced colitis in mice was accompanied by regulation of systemic expression of various pro- and anti-inflammatory cytokines. The most important pathology-promoting cytokines in IBD patients are IL-1β, IL-6, TNFα, IFNγ, and IL-17 [86]. We demonstrated that the EcN-administration–associated decrease of IL-1β, IL-17, IL-6, and TNFα serum levels in DSS-treated mice is abolished when an EcN deletion mutant is used that lacks large parts of its flagellin HVR. Therefore, the *E. coli* flagellin HVR seems to be involved in the regulation of systemic expression of such IBD-promoting cytokines. However, we think that this is instead a secondary effect, probably rooted in IL-22–mediated increased epithelial barrier integrity, thus preventing inflammation-driving translocation of luminal content.

Interestingly, Rakoff-Nahoum and colleagues demonstrated in 2004 that proper TLR2- and TLR4-signaling protects from DSS-induced mortality in microbiota-depleted mice, which is assumed to be necessary for ligand-dependent steady-state induction of protective factors [87]. Our data indicate that intestinal TLR5 signaling might also contribute to these fundamental homeostasis-preserving mechanisms.

Taken together, we propose the flagellin HVR structure may be a distinguishing marker for the classification of *E. coli* strains as either nonsymbiotic or symbiotic. Furthermore, we demonstrated that flagellin-mediated symbiotic effects were originated in the structure of the flagellin HVR, which, in turn, influences TLR5-mediated intracellular signaling in intestinal $CD11c^+$ cells, accompanied by regulation of IBD-promoting cytokines. These insights might therefore offer new possibilities for drug development involving existing or custom-designed flagellin structures, especially for intestinal inflammatory disorders such as IBD.

## Materials and methods

### Ethics statement

In this study, we use WT C57BL/6 and TLR5-deficient animals as well as BMCM as described in the manuscript. This study was carried out in accordance with the principles of the Basel Declaration. Protocols and experiments involving mice were reviewed and approved by the responsible Institutional Review Committee and the local authorities (permit numbers: H3/18, H9/11, H5/10). All mice used in the experiments were killed by $CO_2$ fumigation, as demanded by the responsible authorities. Mice were weighed daily, and a general checkup of the animals' constitution was performed daily. A weight loss of more than 20% compared to the start of the experiment is an indicator for unacceptable suffering, leading to immediate euthanization of the respective animal. Additionally, the general constitution of the mouse was evaluated according to a score sheet that was approved by the responsible authorities. Exceeding a certain score requires immediate euthanization of the mouse. However, no mouse used for the

experiments in the submitted manuscript was required to be euthanized before the projected end of each experiment.

## Mice

C57BL/6N (WT) and *Tlr5*$^{-/-}$ mice were bred and raised under SPF conditions in the animal facility at the University of Tübingen, Germany and did not provide any signs of spontaneous colitis. CD11cCre/R-DTA (ΔCD11c) [88] mice were bred under SPF conditions at the animal facility of the University of Erlangen. All animal experiments were reviewed and approved by the responsible authorities.

## Generation of BMCM

Mice were irradiated in a gamma cell (GammaCell 1000 Elite; Nordion International, Ottawa, ON, Canada) with 900 cGy (female) or 950 cGy (male). Six hours after irradiation, mice were injected intravenously with freshly isolated bone marrow cells ($1 \times 10^7$ cells/100 μL PBS) from donor mice. Bone marrow cells were isolated as described previously [41] with minor modifications. Mice were administered Cotrim E (26 mg/100 mL sterile drinking water, purchased from Ratiopharm, Ulm, Germany) for the first two weeks after irradiation. Six weeks after irradiation, blood was analyzed for successful reconstitution of the transplanted bone marrow by flow cytometry. BMCM were used for induction of DSS-induced colitis as described below.

## Bacteria

Mice were administered EcN, the isogenic mutant EcNΔ*fliC* [46], the nonsymbiotic MG1655 [89], and pathobiotic MPK [41,42,44]. Furthermore, an MG1655Δ*fliC* mutant was complemented with the *fliC* gene from EcN (MG1655Δ*fliC*::*fliC*(EcN)) and an EcNΔ*fliC* mutant was complemented with the *fliC* gene from MPK (EcNΔ*fliC*::*fliC*(MPK)), and both were administered to mice. The EcNΔ*fliC*(HVR) mutant lacks large parts of the HVR within the *fliC* gene. All strains were grown at 37˚C in LB broth under aerobic conditions.

## Generation of FEPs

Bacteria were grown in LB broth as described above. Bacteria were centrifuged at 4,000 rpm, and the cell pellet was resuspended in proteinase inhibitor cocktail (Roche, Basel, Switzerland) and shaken vigorously for 10 min. The suspension was centrifuged for 15 min (4,000 rpm, 4˚C), and the supernatant was used as an FEP.

## Administration of live bacteria, FEPs, or recombinant flagellin to mice

We observed the progress of DSS-induced colitis in mice in response to treatment with viable bacteria, FEPs, or recombinant flagellin. Treatment with viable bacteria started 3 days prior to start of DSS administration by one-time intragastral gavage of a total of $10^8$ bacteria. Simultaneously, autoclaved drinking water was supplemented with viable bacteria at a final concentration of $10^8$ mL$^{-1}$. 3 days after initial intragastral gavage, DSS was added to drinking water suspensions at a final concentration of 3.5% (w/v). Drinking water containing bacteria and DSS were renewed every 2 days until the end of the experiment. Colonization with viable bacteria was assessed by determination of CFUs, as described in the supplementary material. Administration of FEPs to DSS-treated mice was comparable: FEPs obtained from $10^7$ or $10^{10}$ bacteria were resuspended in 100 μL autoclaved drinking water and administered once by intragastral gavage 3 days prior to the start of DSS administration. Subsequent treatment with FEPs for the next 10 days was performed by generating drinking water suspensions containing

FEPs from $10^7$ or $10^{10}$ in 100 mL drinking water. DSS was added to the drinking water suspension 3 days after initial FEP administration, and suspensions were renewed every 2 days. Recombinant flagellin was also started to be administered 3 d before start of DSS treatment. A total of 2 μg recombinant flagellin was administered daily by intragastral gavage for a total of 10 d. DSS treatment started 3 d after initial flagellin administration, and DSS solutions were renewed every 2 d.

## Generation of pure recombinant EcN flagellin preparations

Chemocompetent NiCo BL21 (D3) *E. coli* were transformed by heatshock with a pET19-b expression vector harboring the His-tagged sequence for EcN flagellin (FliC(EcN)). After recovery time, transformed bacteria were inoculated into ampicillin-containing LB medium and incubated at 37˚C O/N. An aliquot of the O/N culture was grown until OD = 0.6. 200 μM. IPTG was added to induce lac-operon controlled protein expression of rfliC(EcN). Bacterial culture was grown O/N at 18˚C and 90 rpm, and cells were harvested by centrifugation at 17,000 rpm and 4˚C. The pellet was lysed for 15 min on ice using protease-inhibitor–and benzonase-containing lysis buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 5 mM imidazole) at a ratio of 10 mL buffer per gram pellet followed by sonification on a Branson 250 sonifier (Branson Ultrasonics, Danbury, CT, USA). Lysates were centrifuged at 35,000 rpm for 45 min at 4˚C. The 1 mL HisTrap Column (GE Healthcare Life Sciences, Marlborough, MA, USA) was equilibrated with equilibration buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0]), and bacterial lysate was loaded on the column afterwards. Protein binding to the HisTrap Column was performed O/N. Next, columns were washed with various concentrations of elution buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 500 mM imidazole), ranging from 3% to 100%. At 20% elution buffer, His-tagged rfliC(EcN) started to be eluted, and elute fractions were collected. Fractions were analyzed via SDS-PAGE and α-fliC western blotting as demonstrated in the supplementary material (S9 and S10 Figs).

## Isolation of LP cells and staining for flow cytometry

LP cells from the colon were isolated as published previously [41,57] with minor modifications. Cells were stained for viability using ViabilityStain (eBioscience, Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instruction. $2 \times 10^6$ cells were incubated in DMEM (Gibco, Gaithersburg, MD, USA) supplemented with 10% FCS, 1% HEPES, 1% nonessential amino acids, 1% sodium pyruvate, 0.5% penicillin/streptomycin, 0.5% β-mercaptoethanol, and 2 μL leukocyte activation cocktail (BD Biosciences, San Jose, CA, USA) for 4 h at 37˚C. Cells were washed and fixed with Cytofix/Cytoperm (BD Biosciences). Cells were washed in PBS/FCS containing 0.1% saponin and treated with Cytofix/Cytoperm (BD Biosciences) for 10 min at RT. Antibodies were diluted by factor 100 in PBS/FCS + 0.1% saponin and incubated with cells for 30 min at 4˚C. Cells were washed twice, and flow cytometrical detection was performed subsequently.

## Antibodies, chemicals, and reagents

For flow cytometry, the following antibodies were used: α-mouse IL-22 (1H8PWSR; eBioscience), α-mouse IL-23p19 (N71-1183; BD Biosciences), α-mouse CD11c (HL3; BD Biosciences), α-mouse CD64 (X54-5/7.1; BD Biosciences), α-mouse CD45 (30-F11; BD Biosciences), α-mouse CD45R (RA3-6B2; BD Biosciences), and α-mouse Ly6G/C (GR-1/RB-68C5; BD Biosciences). Recombinant flagellin from MG1655 was obtained from MyBioSource (#MBS1265520; San Diego, CA, USA).

### DSS-induced colitis

Mice were administered live bacteria, recombinant flagellin, or FEP 3 days prior to challenging with 3.5% (w/v) DSS in drinking water and during the whole course of the experiment as described above. Body weight was determined on day 0 (start of DSS administration) as well as on days 3 to 7.

### Statistical analysis

For comparisons of two groups, a parametric Student *t* test was used for normally distributed values and nonparametric Mann–Whitney test elsewhere. For multiple comparison of more than two groups, one-way ANOVA was used for normally distributed values, and nonparametric Kruskal–Wallis test was used elsewhere. *p*-values are indicated in the figures. *p*-values $< 0.05$ were considered to be significant.

See Supplementary material S1 Text for additional information on experimental procedures.

## Supporting information

**S1 Fig. Determination of *E. coli* CFUs in the feces of DSS-treated mice.** SPF C57BL/6 WT mice (a) and *Tlr5*$^{-/-}$ mice (b) aged 6 to 8 weeks were administered 3.5% DSS in drinking water at day 0. Mice were additionally treated with EcN (DSS + EcN), MG1655 (DSS + MG1655), MPK (DSS + MPK), or the EcN Δ*fliC* deletion mutant (EcNΔ*fliC*) resuspended in DSS-containing drinking water at $10^8$ bacteria mL$^{-1}$. At day 7 after start of DSS administration, feces were plated on Enterobacteriaceae-specific agar in serial dilutions and CFUs were determined by counting dark red colonies specific for *E. coli* strains. (a + b) The data underlying this figure can be found in S1 Data. CFU, colony-forming unit; DSS, dextran sodium sulphate; EcN, *E. coli* Nissle 1917; *fliC*, flagellin; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk; SPF, specific-pathogen–free; TLR, Toll-like receptor; WT, wild type.
(PNG)

**S2 Fig. EcN, MPK, and MG1655 express functional flagella.** Right column: overnight cultures of MPK, EcN, and MG1655. Overnight bacterial culture was seeded in the middle of a swarming culture medium and incubated for 24 h. The inoculation spot is indicated by a red circle, and the borders of the swarming area are highlighted with a white scattered line. Left column: electron microscopy pictures (negative staining) of EcN (upper panel), MPK (middle panel), and MG1655 (lower panel) highlighting the respective flagellum (red arrows). MPK lost its flagella during the staining procedure and could be detected as the shed structure. The insert in the respective picture (left column, middle panel) shows an MPK bacterium. EcN, *E. coli* Nissle 1917; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk.
(PNG)

**S3 Fig. EcNΔ*fliC* does not express a functional flagellum.** Right: overnight bacterial culture of EcNΔfliC was seeded in the middle of a swarming culture medium and incubated for 24 h. The inoculation spot is indicated by a red circle, and the borders of the swarming area are highlighted with a white scattered line. Left column: electron microscopy pictures (negative staining) of EcNΔfliC (highlighting the absence of flagella). EcN, *E. coli* Nissle 1917; *fliC*, flagellin.
(PNG)

**S4 Fig. Characterization of FEPs.** EcN, MPK, and MG1655 were grown to OD100, and FEPs were generated as described in the main manuscript. (a) Silver staining of 10 μL FEP on an

8%–15% gradient SDS gel. (b) LAL test to determine endotoxin levels in FEPs. (c) Determination of overall protein concentration in FEPs using a bicinchoninic acid assisted assay. (d) Western blot of FliC of 10 μL FEPs using anti-flagellin antibody (ab93713; Abcam, Cambridge, UK). (e) Quantification of FliC concentrations in FEPs. Band intensities of FliC bands in western blots depicted in (d) were quantified. A standard curve of recombinant EcN FliC was generated and visualized with the same antibodies on the same blots. FliC concentrations were computed using the determined FliC band intensities in relation to a linear regression of the band intensities of FliC standard curve. (b + c + e) The data underlying this figure can be found in S1 Data. EcN, *E. coli* Nissle 1917; FEP, flagella-enriched preparation; *fliC*, flagellin; LAL, limulus amebocyte lysate; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk; OD, optical density.
(PNG)

**S5 Fig. Generation of *fliC* exchange mutant strains.** Chromosomal exchange of *fliC* alleles was done by allelic exchange as described previously [90]. Upper panel: suicide plasmids were constructed by Gibson assembly according to standard protocols [91]. Lower panel: primers and plasmids for allelic exchange as well as resulting strains. *fliC*, flagellin.
(PNG)

**S6 Fig. MG1655ΔfliC::fliC(EcN) and EcNΔfliC::fliC(MPK) express a functional flagella.** Right column: overnight bacterial culture of MG1655ΔfliC::fliC(EcN) and EcNΔfliC::fliC (MPK) exchange mutants were seeded in the middle of a swarming culture medium and incubated for 24 h. The inoculation spot is indicated by a red circle, and the borders of the swarming area are highlighted with a white scattered line. Left column: electron microscopy pictures of MG1655ΔfliC::fliC(EcN) and EcNΔfliC::fliC(MPK) highlighting the respective flagella (red arrow). EcN, *E. coli* Nissle 1917; *fliC*, flagellin; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk.
(PNG)

**S7 Fig. EcNΔflic(HVR) expresses a shorter flagella compared to WT EcN.** Left panel: EM pictures of EcNΔfliC(HVR) deletion mutants highlighting the flagella (red arrow). Right panel: EM-assisted determination of flagella lengths. Each white dot represents one detected flagellum in EM pictures. The data underlying this figure can be found in S1 Data. EcN, *E. coli* Nissle 1917; EM, electron microscopy; *fliC*, flagellin; HVR, hypervariable region; WT, wild type.
(PNG)

**S8 Fig. Schematic illustration of generation of BMCM.** (A) $Tlr5^{-/-}$ → WT mice by irradiation of C57BL/6×WT-CD45.1–expressing mice transplanted with bone marrow of C57BL/6×$Tlr5^{-/-}$-CD45.2–expressing mice and (B) WT → $Tlr5^{-/-}$ mice by irradiation of C57BL/6×$Tlr5^{-/-}$-CD45.2–expressing mice transplanted with bone marrow of C57BL/6×WT-CD45.1–expressing mice. (C) Irradiated C57BL/6-CD45.2 mice transplanted with C57BL/6-CD45.1 bone marrow (WT → WT), (D) $Tlr5^{-/-}$CD45.2 mice transplanted with $Tlr5^{-/-}$-CD45.2 bone marrow ($Tlr5^{-/-}$ → $Tlr5^{-/-}$). Successful transplantation was monitored by flow cytometry analysis of blood samples stained with antibodies against CD45.1 and CD45.2. Figures show means ± SD of 4 to 9 mice per experiment. BMCM, bone-marrow–chimeric mice; TLR, Toll-like receptor; WT, wild type.
(PNG)

**S9 Fig. Validation of rFliC(EcN) purity.** Coomassi-stained 4%–15% gradient gel of all collected eluted fractions after elution from HisTrap columns using elution buffer (300 mM NaCl, 50 mM Tris-HCl [pH 8.0], 500 mM imidazole) at concentrations from 20% to 100%.

EcN, *E. coli* Nissle 1917; *fliC*, flagellin; rfliC(EcN), recombinant flagellin from EcN.
(PNG)

**S10 Fig. Western blot of rFliC(EcN).** Western blots against the His-tags of rFliC(EcN) (left panel) and FliC (right panel) were performed to verify the proper expression of the recombinant protein. Cell lysates before IPTG-assisted induction of protein expression (preinduction), after IPTG-assisted induction (postinduction), the collected elutes from the HisTrap column (Elute) (see S9 Fig), and a previously purified MS-controlled rFliC(EcN) as positive control were loaded on a 4%–15% gradient gel, and western blots were performed as described. EcN, *E. coli* Nissle 1917; *fliC*, flagellin; MS, mass spectrometry; rfliC(EcN), recombinant flagellin from EcN.
(PNG)

**S1 Table. Detailed statistical analysis of the values depicted in Fig 1E as determined by one-way ANOVA.**
(DOCX)

**S2 Table. Statistical analysis of cytokine serum levels in DSS-treated mice.** *p*-values were computed using nonparametric Kruskal–Wallis test. DSS, dextran sodium sulphate.
(DOCX)

**S1 Data. Raw data underlying the following figures: Fig 1, Fig 2, Fig 4, Fig 5, Fig 6, S1 Fig, S4 Fig, S7 Fig.**
(XLSX)

**S1 Text. Supplementary methods.**
(DOCX)

## Author Contributions

**Conceptualization:** Alex Steimle, Sarah Menz, Alexander N. R. Weber, Samuel Wagner, David Voehringer, Julia-Stefanie Frick.

**Data curation:** Alex Steimle, Sarah Menz, Annika Bender, Brianna Ball, Thomas Hagemann, Anna Lange, Jan K. Maerz, Raphael Parusel, Lena Michaelis, Andrea Schäfer, Hans Yao, Hanna-Christine Löw, Sina Beier, Mehari Tesfazgi Mebrhatu, Kerstin Gronbach, David Voehringer, Birgit Fehrenbacher.

**Formal analysis:** Alex Steimle, Annika Bender, Brianna Ball, Sina Beier, Mehari Tesfazgi Mebrhatu, Kerstin Gronbach, Birgit Fehrenbacher, Julia-Stefanie Frick.

**Investigation:** Alex Steimle, Sarah Menz, Annika Bender, Brianna Ball, Alexander N. R. Weber, Anna Lange, David Voehringer, Birgit Fehrenbacher.

**Methodology:** Alex Steimle, Sarah Menz, Annika Bender, Brianna Ball, Birgit Fehrenbacher, Julia-Stefanie Frick.

**Project administration:** Alex Steimle, Julia-Stefanie Frick.

**Resources:** Mehari Tesfazgi Mebrhatu, David Voehringer, Martin Schaller.

**Supervision:** Alex Steimle, Samuel Wagner, Ingo B. Autenrieth, Tobias A. Oelschlaeger.

**Validation:** Alex Steimle, Sarah Menz, Annika Bender, Anna Lange, Sina Beier.

**Visualization:** Alex Steimle, Sarah Menz, Sina Beier.

**Writing – original draft:** Alex Steimle, Alexander N. R. Weber, Ingo B. Autenrieth, Julia-Stefanie Frick.

## References

1. Kamada N, Seo SU, Chen GY, Nunez G. Role of the gut microbiota in immunity and inflammatory disease. Nat Rev Immunol. 2013; 13: 321–335. https://doi.org/10.1038/nri3430 PMID: 23618829
2. Wassenaar TM. Insights from 100 Years of Research with Probiotic E. Coli. Eur J Microbiol Immunol (Bp). 2016; 6: 147–161.
3. Hill MJ, Drasar BS. The normal colonic bacterial flora. Gut. 1975; 16: 318–323. https://doi.org/10.1136/gut.16.4.318 PMID: 1093952
4. Khan AA, Khan Z, Malik A, Kalam MA, Cash P, et al. Colorectal cancer-inflammatory bowel disease nexus and felony of Escherichia coli. Life Sci. 2017; 180: 60–67. https://doi.org/10.1016/j.lfs.2017.05.016 PMID: 28506682
5. Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. ISME J. 2007; 1: 403–418. https://doi.org/10.1038/ismej.2007.52 PMID: 18043660
6. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007; 104: 13780–13785. https://doi.org/10.1073/pnas.0706625104 PMID: 17699621
7. Lepage P, Hasler R, Spehlmann ME, Rehman A, Zvirbliene A, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology. 2011; 141: 227–236. https://doi.org/10.1053/j.gastro.2011.04.011 PMID: 21621540
8. Scaldaferri F, Gerardi V, Mangiola F, Lopetuso LR, Pizzoferrato M, et al. Role and mechanisms of action of Escherichia coli Nissle 1917 in the maintenance of remission in ulcerative colitis patients: An update. World J Gastroenterol. 2016; 22: 5505–5511. https://doi.org/10.3748/wjg.v22.i24.5505 PMID: 27350728
9. Kruis W, Schutz E, Fric P, Fixa B, Judmaier G, et al. Double-blind comparison of an oral Escherichia coli preparation and mesalazine in maintaining remission of ulcerative colitis. Aliment Pharmacol Ther. 1997; 11: 853–858. PMID: 9354192
10. Rembacken BJ, Snelling AM, Hawkey PM, Chalmers DM, Axon AT. Non-pathogenic Escherichia coli versus mesalazine for the treatment of ulcerative colitis: a randomised trial. Lancet. 1999; 354: 635–639. https://doi.org/10.1016/s0140-6736(98)06343-0 PMID: 10466665
11. Henker J, Muller S, Laass MW, Schreiner A, Schulze J. Probiotic Escherichia coli Nissle 1917 (EcN) for successful remission maintenance of ulcerative colitis in children and adolescents: an open-label pilot study. Z Gastroenterol. 2008; 46: 874–875. https://doi.org/10.1055/s-2008-1027463 PMID: 18810672
12. Lasaro MA, Salinger N, Zhang J, Wang Y, Zhong Z, et al. F1C fimbriae play an important role in biofilm formation and intestinal colonization by the Escherichia coli commensal strain Nissle 1917. Appl Environ Microbiol. 2009; 75: 246–251. https://doi.org/10.1128/AEM.01144-08 PMID: 18997018
13. Wehkamp J, Harder J, Wehkamp K, Wehkamp-von Meissner B, Schlee M, et al. NF-kappaB- and AP-1-mediated induction of human beta defensin-2 in intestinal epithelial cells by Escherichia coli Nissle 1917: a novel effect of a probiotic bacterium. Infect Immun. 2004; 72: 5750–5758. https://doi.org/10.1128/IAI.72.10.5750-5758.2004 PMID: 15385474
14. Zyrek AA, Cichon C, Helms S, Enders C, Sonnenborn U, et al. Molecular mechanisms underlying the probiotic effects of Escherichia coli Nissle 1917 involve ZO-2 and PKCzeta redistribution resulting in tight junction and epithelial barrier repair. Cell Microbiol. 2007; 9: 804–816. https://doi.org/10.1111/j.1462-5822.2006.00836.x PMID: 17087734
15. Sassone-Corsi M, Nuccio SP, Liu H, Hernandez D, Vu CT, et al. Microcins mediate competition among Enterobacteriaceae in the inflamed gut. Nature. 2016; 540: 280–283. https://doi.org/10.1038/nature20557 PMID: 27798599
16. Sturm A, Rilling K, Baumgart DC, Gargas K, Abou-Ghazale T, et al. Escherichia coli Nissle 1917 distinctively modulates T-cell cycling and expansion via toll-like receptor 2 signaling. Infect Immun. 2005; 73: 1452–1465. https://doi.org/10.1128/IAI.73.3.1452-1465.2005 PMID: 15731043
17. Jiminez JA, Uwiera TC, Douglas Inglis G, Uwiera RR. Animal models to study acute and chronic intestinal inflammation in mammals. Gut Pathog. 2015; 7: 29. https://doi.org/10.1186/s13099-015-0076-y PMID: 26561503

18. Fang K, Bruce M, Pattillo CB, Zhang S, Stone R 2nd, et al. Temporal genomewide expression profiling of DSS colitis reveals novel inflammatory and angiogenesis genes similar to ulcerative colitis. Physiol Genomics. 2011; 43: 43–56. https://doi.org/10.1152/physiolgenomics.00138.2010 PMID: 20923862

19. Kanwar B, Gao DW, Hwang AB, Grenert JP, Williams SP, et al. In vivo imaging of mucosal CD4+ T cells using single photon emission computed tomography in a murine model of colitis. J Immunol Methods. 2008; 329: 21–30. https://doi.org/10.1016/j.jim.2007.09.008 PMID: 17964595

20. Dieleman LA, Palmen MJ, Akol H, Bloemena E, Pena AS, et al. Chronic experimental colitis induced by dextran sulphate sodium (DSS) is characterized by Th1 and Th2 cytokines. Clin Exp Immunol. 1998; 114: 385–391. https://doi.org/10.1046/j.1365-2249.1998.00728.x PMID: 9844047

21. Kim TW, Seo JN, Suh YH, Park HJ, Kim JH, et al. Involvement of lymphocytes in dextran sulfate sodium-induced experimental colitis. World J Gastroenterol. 2006; 12: 302–305. https://doi.org/10.3748/wjg.v12.i2.302 PMID: 16482634

22. Teahon K, Smethurst P, Pearson M, Levi AJ, Bjarnason I. The effect of elemental diet on intestinal permeability and inflammation in Crohn's disease. Gastroenterology. 1991; 101: 84–89. PMID: 1904381

23. Keshavarzian A, Price YE, Peters AM, Lavender JP, Wright NA, et al. Specificity of indium-111 granulocyte scanning and fecal excretion measurement in inflammatory bowel disease—an autoradiographic study. Dig Dis Sci. 1985; 30: 1156–1160. PMID: 4064866

24. Costa F, Mumolo MG, Ceccarelli L, Bellini M, Romano MR, et al. Calprotectin is a stronger predictive marker of relapse in ulcerative colitis than in Crohn's disease. Gut. 2005; 54: 364–368. https://doi.org/10.1136/gut.2004.043406 PMID: 15710984

25. Melgar S, Karlsson L, Rehnstrom E, Karlsson A, Utkovic H, et al. Validation of murine dextran sulfate sodium-induced colitis using four therapeutic agents for human inflammatory bowel disease. Int Immunopharmacol. 2008; 8: 836–844. https://doi.org/10.1016/j.intimp.2008.01.036 PMID: 18442787

26. Claes IJ, De Keersmaecker SC, Vanderleyden J, Lebeer S. Lessons from probiotic-host interaction studies in murine models of experimental colitis. Mol Nutr Food Res. 2011; 55: 1441–1453. https://doi.org/10.1002/mnfr.201100139 PMID: 21796777

27. Kawasaki T, Kawai T. Toll-like receptor signaling pathways. Front Immunol. 2014; 5: 461. https://doi.org/10.3389/fimmu.2014.00461 PMID: 25309543

28. Song WS, Jeon YJ, Namgung B, Hong M, Yoon SI. A conserved TLR5 binding and activation hot spot on flagellin. Sci Rep. 2017; 7: 40878. https://doi.org/10.1038/srep40878 PMID: 28106112

29. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, et al. Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. Nat Immunol. 2003; 4: 1247–1253. https://doi.org/10.1038/ni1011 PMID: 14625549

30. Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, et al. Structural basis of TLR5-flagellin recognition and signaling. Science. 2012; 335: 859–864. https://doi.org/10.1126/science.1215584 PMID: 22344444

31. Ramos HC, Rumbo M, Sirard JC. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. 2004; 12: 509–517. https://doi.org/10.1016/j.tim.2004.09.002 PMID: 15488392

32. Zhao Y, Yang J, Shi J, Gong YN, Lu Q, et al. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. Nature. 2011; 477: 596–600. https://doi.org/10.1038/nature10510 PMID: 21918512

33. Lodes MJ, Cong Y, Elson CO, Mohamath R, Landers CJ, et al. Bacterial flagellin is a dominant antigen in Crohn disease. J Clin Invest. 2004; 113: 1296–1306. https://doi.org/10.1172/JCI20295 PMID: 15124021

34. Meena NK, Ahuja V, Meena K, Paul J. Association of TLR5 gene polymorphisms in ulcerative colitis patients of north India and their role in cytokine homeostasis. PLoS ONE. 2015; 10: e0120697. https://doi.org/10.1371/journal.pone.0120697 PMID: 25789623

35. Klimosch SN, Forsti A, Eckert J, Knezevic J, Bevier M, et al. Functional TLR5 genetic variants affect human colorectal cancer survival. Cancer Res. 2013; 73: 7232–7242. https://doi.org/10.1158/0008-5472.CAN-13-1746 PMID: 24154872

36. Singh V, Yeoh BS, Carvalho F, Gewirtz AT, Vijay-Kumar M. Proneness of TLR5 deficient mice to develop colitis is microbiota dependent. Gut Microbes. 2015; 6: 279–283. https://doi.org/10.1080/19490976.2015.1060390 PMID: 26067589

37. Chassaing B, Ley RE, Gewirtz AT. Intestinal epithelial cell toll-like receptor 5 regulates the intestinal microbiota to prevent low-grade inflammation and metabolic syndrome in mice. Gastroenterology. 2014; 147: 1363–1377 e1317. https://doi.org/10.1053/j.gastro.2014.08.033 PMID: 25172014

38. Grabig A, Paclik D, Guzy C, Dankof A, Baumgart DC, et al. Escherichia coli strain Nissle 1917 ameliorates experimental colitis via toll-like receptor 2- and toll-like receptor 4-dependent pathways. Infect Immun. 2006; 74: 4075–4082. https://doi.org/10.1128/IAI.01449-05 PMID: 16790781

39. Garrido-Mesa N, Utrilla P, Comalada M, Zorrilla P, Garrido-Mesa J, et al. The association of minocycline and the probiotic Escherichia coli Nissle 1917 results in an additive beneficial effect in a DSS model of reactivated colitis in mice. Biochem Pharmacol. 2011; 82: 1891–1900. https://doi.org/10.1016/j.bcp.2011.09.004 PMID: 21930116

40. Olier M, Marcq I, Salvador-Cartier C, Secher T, Dobrindt U, et al. Genotoxicity of Escherichia coli Nissle 1917 strain cannot be dissociated from its probiotic activity. Gut Microbes. 2012; 3: 501–509. https://doi.org/10.4161/gmic.21737 PMID: 22895085

41. Steimle A, Gronbach K, Beifuss B, Schafer A, Harmening R, et al. Symbiotic gut commensal bacteria act as host cathepsin S activity regulators. J Autoimmun. 2016; 75: 82–95. https://doi.org/10.1016/j.jaut.2016.07.009 PMID: 27484364

42. Waidmann M, Bechtold O, Frick JS, Lehr HA, Schubert S, et al. *Bacteroides vulgatus* protects against *Escherichia coli*-induced colitis in gnotobiotic interleukin-2-deficient mice. Gastroenterology. 2003; 125: 162–177. PMID: 12851881

43. Muller M, Fink K, Geisel J, Kahl F, Jilge B, et al. Intestinal colonization of IL-2 deficient mice with non-colitogenic B. vulgatus prevents DC maturation and T-cell polarization. PLoS ONE. 2008; 3: e2376. https://doi.org/10.1371/journal.pone.0002376 PMID: 18545662

44. Frick JS, Zahir N, Muller M, Kahl F, Bechtold O, et al. Colitogenic and non-colitogenic commensal bacteria differentially trigger DC maturation and Th cell polarization: an important role for IL-6. Eur J Immunol. 2006; 36: 1537–1547. https://doi.org/10.1002/eji.200635840 PMID: 16708404

45. Secher T, Kassem S, Benamar M, Bernard I, Boury M, et al. Oral Administration of the Probiotic Strain Escherichia coli Nissle 1917 Reduces Susceptibility to Neuroinflammation and Repairs Experimental Autoimmune Encephalomyelitis-Induced Intestinal Barrier Dysfunction. Front Immunol. 2017; 8: 1096. https://doi.org/10.3389/fimmu.2017.01096 PMID: 28959254

46. Troge A, Scheppach W, Schroeder BO, Rund SA, Heuner K, et al. More than a marine propeller—the flagellum of the probiotic Escherichia coli strain Nissle 1917 is the major adhesin mediating binding to human mucus. Int J Med Microbiol. 2012; 302: 304–314. https://doi.org/10.1016/j.ijmm.2012.09.004 PMID: 23131416

47. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science. 2010; 328: 228–231. https://doi.org/10.1126/science.1179721 PMID: 20203013

48. Vijay-Kumar M, Sanders CJ, Taylor RT, Kumar A, Aitken JD, et al. Deletion of TLR5 results in spontaneous colitis in mice. J Clin Invest. 2007; 117: 3909–3921. https://doi.org/10.1172/JCI33084 PMID: 18008007

49. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, et al. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. J Exp Med. 2012; 209: 1445–1456. https://doi.org/10.1084/jem.20120504 PMID: 22826298

50. Yonekura K, Maki-Yonekura S, Namba K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature. 2003; 424: 643–650. https://doi.org/10.1038/nature01830 PMID: 12904785

51. Yonekura K, Maki-Yonekura S, Namba K. Structure analysis of the flagellar cap-filament complex by electron cryomicroscopy and single-particle image analysis. J Struct Biol. 2001; 133: 246–253. https://doi.org/10.1006/jsbi.2000.4345 PMID: 11472095

52. Jeengar MK, Thummuri D, Magnusson M, Naidu VGM, Uppugunduri S. Uridine Ameliorates Dextran Sulfate Sodium (DSS)-Induced Colitis in Mice. Sci Rep. 2017; 7: 3924. https://doi.org/10.1038/s41598-017-04041-9 PMID: 28634361

53. Alex P, Zachos NC, Nguyen T, Gonzales L, Chen TE, et al. Distinct cytokine patterns identified from multiplex profiles of murine DSS and TNBS-induced colitis. Inflamm Bowel Dis. 2009; 15: 341–352. https://doi.org/10.1002/ibd.20753 PMID: 18942757

54. Marques R, Boneca IG. Expression and functional importance of innate immune receptors by intestinal epithelial cells. Cell Mol Life Sci. 2011; 68: 3661–3673. https://doi.org/10.1007/s00018-011-0829-9 PMID: 21984599

55. Price AE, Shamardani K, Lugo KA, Deguine J, Roberts AW, et al. A Map of Toll-like Receptor Expression in the Intestinal Epithelium Reveals Distinct Spatial, Cell Type-Specific, and Temporal Patterns. Immunity. 2018; 49: 560–575 e566. https://doi.org/10.1016/j.immuni.2018.07.016 PMID: 30170812

56. Liu H, Chen F, Wu W, Cao AT, Xue X, et al. TLR5 mediates CD172alpha(+) intestinal lamina propria dendritic cell induction of Th17 cells. Sci Rep. 2016; 6: 22040. https://doi.org/10.1038/srep22040 PMID: 26907705

57. Gronbach K, Flade I, Holst O, Lindner B, Ruscheweyh HJ, et al. Endotoxicity of lipopolysaccharide as a determinant of T-cell-mediated colitis induction in mice. Gastroenterology. 2014; 146: 765–775. https://doi.org/10.1053/j.gastro.2013.11.033 PMID: 24269927

58. Takatori H, Kanno Y, Watford WT, Tato CM, Weiss G, et al. Lymphoid tissue inducer-like cells are an innate source of IL-17 and IL-22. J Exp Med. 2009; 206: 35–41. https://doi.org/10.1084/jem.20072713 PMID: 19114665

59. Cella M, Fuchs A, Vermi W, Facchetti F, Otero K, et al. A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. Nature. 2009; 457: 722–725. https://doi.org/10.1038/nature07537 PMID: 18978771

60. Scholtens PA, Oozeer R, Martin R, Amor KB, Knol J. The early settlers: intestinal microbiology in early life. Annu Rev Food Sci Technol. 2012; 3: 425–447. https://doi.org/10.1146/annurev-food-022811-101120 PMID: 22224552

61. Pagnini C, Saeed R, Bamias G, Arseneau KO, Pizarro TT, et al. Probiotics promote gut health through stimulation of epithelial innate immunity. Proc Natl Acad Sci U S A. 2010; 107: 454–459. https://doi.org/10.1073/pnas.0910307107 PMID: 20018654

62. Fabrega MJ, Rodriguez-Nogales A, Garrido-Mesa J, Algieri F, Badia J, et al. Intestinal Anti-inflammatory Effects of Outer Membrane Vesicles from Escherichia coli Nissle 1917 in DSS-Experimental Colitis in Mice. Front Microbiol. 2017; 8: 1274. https://doi.org/10.3389/fmicb.2017.01274 PMID: 28744268

63. Cheng K, Sloan A, Peterson L, McCorrister S, Robinson A, et al. Comparative study of traditional flagellum serotyping and liquid chromatography-tandem mass spectrometry-based flagellum typing with clinical Escherichia coli isolates. J Clin Microbiol. 2014; 52: 2275–2278. https://doi.org/10.1128/JCM.00174-14 PMID: 24671787

64. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SL, Cookson BT, et al. Evasion of Toll-like receptor 5 by flagellated bacteria. Proc Natl Acad Sci U S A. 2005; 102: 9247–9252. https://doi.org/10.1073/pnas.0502040102 PMID: 15956202

65. Muskotal A, Seregelyes C, Sebestyen A, Vonderviszt F. Structural basis for stabilization of the hypervariable D3 domain of Salmonella flagellin upon filament formation. J Mol Biol. 2010; 403: 607–615. https://doi.org/10.1016/j.jmb.2010.09.024 PMID: 20868693

66. Franchi L, Kamada N, Nakamura Y, Burberry A, Kuffa P, et al. NLRC4-driven production of IL-1beta discriminates between pathogenic and commensal bacteria and promotes host intestinal defense. Nat Immunol. 2012; 13: 449–456. https://doi.org/10.1038/ni.2263 PMID: 22484733

67. Lightfield KL, Persson J, Brubaker SW, Witte CE, von Moltke J, et al. Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. Nat Immunol. 2008; 9: 1171–1178. https://doi.org/10.1038/ni.1646 PMID: 18724372

68. Gewirtz AT, Vijay-Kumar M, Brant SR, Duerr RH, Nicolae DL, et al. Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease. Am J Physiol Gastrointest Liver Physiol. 2006; 290: G1157–1163. https://doi.org/10.1152/ajpgi.00544.2005 PMID: 16439468

69. Leifer CA, McConkey C, Li S, Chassaing B, Gewirtz AT, et al. Linking genetic variation in human Toll-like receptor 5 genes to the gut microbiome's potential to cause inflammation. Immunol Lett. 2014; 162: 3–9. https://doi.org/10.1016/j.imlet.2014.07.017 PMID: 25284610

70. Sanders CJ, Moore DA 3rd, Williams IR, Gewirtz AT. Both radioresistant and hemopoietic cells promote innate and adaptive immune responses to flagellin. J Immunol. 2008; 180: 7184–7192. https://doi.org/10.4049/jimmunol.180.11.7184 PMID: 18490717

71. Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, et al. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature. 2001; 410: 1099–1103. https://doi.org/10.1038/35074106 PMID: 11323673

72. Dudakov JA, Hanash AM, van den Brink MR. Interleukin-22: immunobiology and pathology. Annu Rev Immunol. 2015; 33: 747–785. https://doi.org/10.1146/annurev-immunol-032414-112123 PMID: 25706098

73. Goto Y, Obata T, Kunisawa J, Sato S, Ivanov II, et al. Innate lymphoid cells regulate intestinal epithelial cell glycosylation. Science. 2014; 345: 1254009. https://doi.org/10.1126/science.1254009 PMID: 25214634

74. Wolk K, Kunz S, Witte E, Friedrich M, Asadullah K, et al. IL-22 increases the innate immunity of tissues. Immunity. 2004; 21: 241–254. https://doi.org/10.1016/j.immuni.2004.07.007 PMID: 15308104

75. Zheng Y, Valdez PA, Danilenko DM, Hu Y, Sa SM, et al. Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. Nat Med. 2008; 14: 282–289. https://doi.org/10.1038/nm1720 PMID: 18264109

76. Sugimoto K, Ogawa A, Mizoguchi E, Shimomura Y, Andoh A, et al. IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. J Clin Invest. 2008; 118: 534–544. https://doi.org/10.1172/JCI33194 PMID: 18172556

77. Liang SC, Tan XY, Luxenberg DP, Karim R, Dunussi-Joannopoulos K, et al. Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. J Exp Med. 2006; 203: 2271–2279. https://doi.org/10.1084/jem.20061308 PMID: 16982811

78. Zhou G, Yu L, Fang L, Yang W, Yu T, et al. CD177(+) neutrophils as functionally activated neutrophils negatively regulate IBD. Gut. 2018; 67: 1052–1063. https://doi.org/10.1136/gutjnl-2016-313535 PMID: 28468761

79. Sonnenberg GF, Fouser LA, Artis D. Border patrol: regulation of immunity, inflammation and tissue homeostasis at barrier surfaces by IL-22. Nat Immunol. 2011; 12: 383–390. https://doi.org/10.1038/ni.2025 PMID: 21502992

80. Mann ER, Bernardo D, Ng SC, Rigby RJ, Al-Hassi HO, et al. Human gut dendritic cells drive aberrant gut-specific t-cell responses in ulcerative colitis, characterized by increased IL-4 production and loss of IL-22 and IFNgamma. Inflamm Bowel Dis. 2014; 20: 2299–2307. https://doi.org/10.1097/MIB.0000000000000223 PMID: 25397892

81. Pickert G, Neufert C, Leppkes M, Zheng Y, Wittkopf N, et al. STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. J Exp Med. 2009; 206: 1465–1472. https://doi.org/10.1084/jem.20082683 PMID: 19564350

82. Zenewicz LA, Yancopoulos GD, Valenzuela DM, Murphy AJ, Stevens S, et al. Innate and adaptive interleukin-22 protects mice from inflammatory bowel disease. Immunity. 2008; 29: 947–957. https://doi.org/10.1016/j.immuni.2008.11.003 PMID: 19100701

83. Macho-Fernandez E, Koroleva EP, Spencer CM, Tighe M, Torrado E, et al. Lymphotoxin beta receptor signaling limits mucosal damage through driving IL-23 production by epithelial cells. Mucosal Immunol. 2015; 8: 403–413. https://doi.org/10.1038/mi.2014.78 PMID: 25183367

84. Patterson AM, Mulder IE, Travis AJ, Lan A, Cerf-Bensussan N, et al. Human Gut Symbiont Roseburia hominis Promotes and Regulates Innate Immunity. Front Immunol. 2017; 8: 1166. https://doi.org/10.3389/fimmu.2017.01166 PMID: 29018440

85. Uematsu S, Jang MH, Chevrier N, Guo Z, Kumagai Y, et al. Detection of pathogenic intestinal bacteria by Toll-like receptor 5 on intestinal CD11c+ lamina propria cells. Nat Immunol. 2006; 7: 868–874. https://doi.org/10.1038/ni1362 PMID: 16829963

86. Strober W, Fuss IJ. Proinflammatory cytokines in the pathogenesis of inflammatory bowel diseases. Gastroenterology. 2011; 140: 1756–1767. https://doi.org/10.1053/j.gastro.2011.02.016 PMID: 21530742

87. Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, Medzhitov R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell. 2004; 118: 229–241. https://doi.org/10.1016/j.cell.2004.07.002 PMID: 15260992

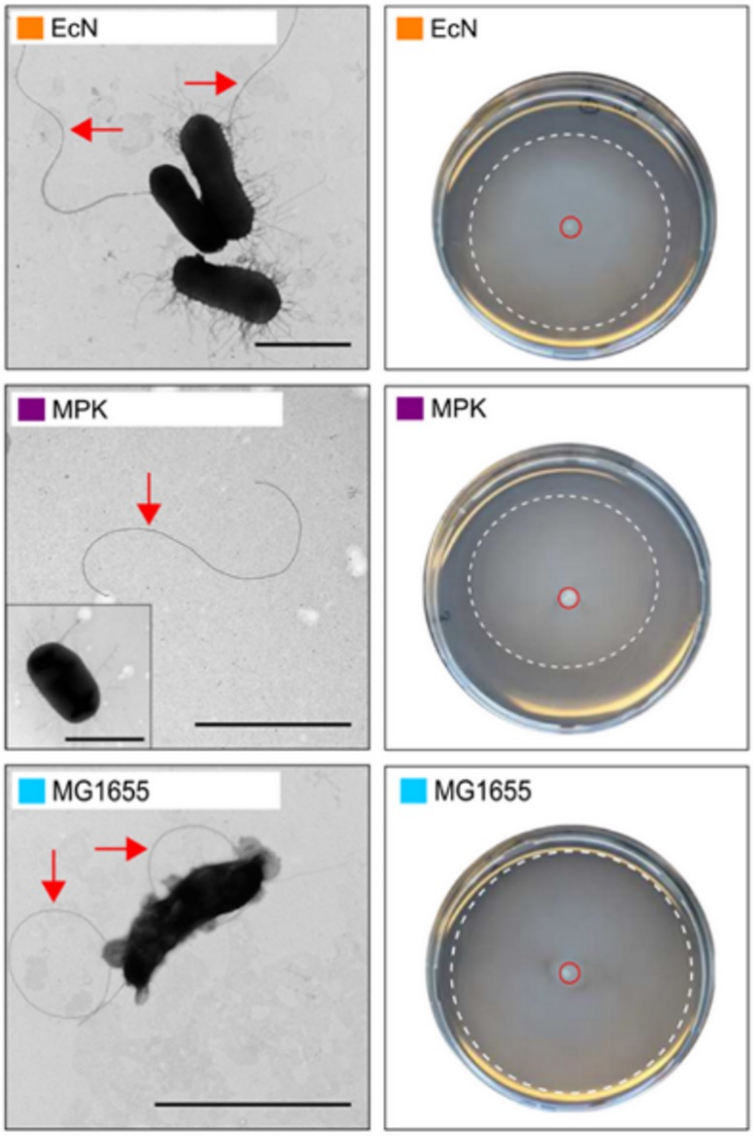
88. Ohnmacht C, Pullner A, King SB, Drexler I, Meier S, et al. Constitutive ablation of dendritic cells breaks self-tolerance of CD4 T cells and results in spontaneous fatal autoimmunity. J Exp Med. 2009; 206: 549–559. https://doi.org/10.1084/jem.20082394 PMID: 19237601

89. Smith SN, Hagan EC, Lane MC, Mobley HL. Dissemination and systemic colonization of uropathogenic Escherichia coli in a murine model of bacteremia. MBio. 2010; 1(5): e00262–10. https://doi.org/10.1128/mBio.00262-10 PMID: 21116344

90. Kaniga K, Bossio JC, Galan JE. The Salmonella typhimurium invasion genes invF and invG encode homologues of the AraC and PulD family of proteins. Mol Microbiol. 1994; 13:555–68. PMID: 7997169

91. Gibson DG, Young L, Chuang RY, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–5. https://doi.org/10.1038/nmeth.1318 PMID: 19363495

### 6.1.1 Supplementary information.




S1 Fig. EcN, MPK, and MG1655 express functional flagella.

Right column: overnight cultures of MPK, EcN, and MG1655. Overnight bacterial culture was seeded in the middle of a swarming culture medium and incubated for 24 h. The inoculation spot is indicated by a red circle, and the borders of the swarming area are highlighted with a white scattered line. Left column: electron microscopy pictures (negative staining) of EcN (upper panel), MPK (middle panel), and MG1655 (lower panel) highlighting the respective flagellum (red arrows). MPK lost its flagella during the staining procedure and could be detected as the shed structure. The insert in the respective picture (left column, middle panel) shows an MPK bacterium. EcN, *E. coli* Nissle 1917; MG1655, *E. coli* K12 MG1655; MPK, *E. coli* mpk.

# 7 Bibliography


1. Steimle, A., et al., *Flagellin hypervariable region determines symbiotic properties of commensal Escherichia coli strains.* PLoS Biol, 2019. **17**(6): p. e3000334.
2. Petagna, L., et al., *Pathophysiology of Crohn's disease inflammation and recurrence.* Biol Direct, 2020. **15**(1): p. 23.
3. Juillerat, P., et al., *Extraintestinal manifestations of Crohn's disease.* Digestion, 2007. **76**(2): p. 141-8.
4. Malik, T.F. and D.M. Aurelio, *Extraintestinal Manifestations of Inflammatory Bowel Disease*, in *StatPearls*. 2021: Treasure Island (FL).
5. Loos, R.J.F., *15 years of genome-wide association studies and no signs of slowing down.* Nat Commun, 2020. **11**(1): p. 5900.
6. Jess, T., C. Rungoe, and L. Peyrin-Biroulet, *Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies.* Clin Gastroenterol Hepatol, 2012. **10**(6): p. 639-45.
7. Ng, S.C., et al., *Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies.* Lancet, 2017. **390**(10114): p. 2769-2778.
8. Hold, G.L., *Western lifestyle: a 'master' manipulator of the intestinal microbiota?* Gut, 2014. **63**(1): p. 5-6.
9. Stiemsma, L.T., et al., *The hygiene hypothesis: current perspectives and future therapies.* Immunotargets Ther, 2015. **4**: p. 143-57.
10. de Lange, K.M. and J.C. Barrett, *Understanding inflammatory bowel disease via immunogenetics.* J Autoimmun, 2015. **64**: p. 91-100.
11. Mohan, S., S. Mok, and T. Judge, *Identification of Novel Therapeutic Molecular Targets in Inflammatory Bowel Disease by Using Genetic Databases.* Clin Exp Gastroenterol, 2020. **13**: p. 467-473.
12. Zhen, Y. and H. Zhang, *NLRP3 Inflammasome and Inflammatory Bowel Disease.* Front Immunol, 2019. **10**: p. 276.
13. Thursby, E. and N. Juge, *Introduction to the human gut microbiota.* Biochem J, 2017. **474**(11): p. 1823-1836.
14. Yang, J., et al., *Species-Level Analysis of Human Gut Microbiota With Metataxonomics.* Front Microbiol, 2020. **11**: p. 2029.
15. Surawicz, C.M., et al., *Guidelines for diagnosis, treatment, and prevention of Clostridium difficile infections.* Am J Gastroenterol, 2013. **108**(4): p. 478-98; quiz 499.
16. van Nood, E., et al., *Duodenal infusion of donor feces for recurrent Clostridium difficile.* N Engl J Med, 2013. **368**(5): p. 407-15.
17. LeBlanc, J.G., et al., *Bacteria as vitamin suppliers to their host: a gut microbiota perspective.* Curr Opin Biotechnol, 2013. **24**(2): p. 160-8.
18. Castro, F. and H.S.P. de Souza, *Dietary Composition and Effects in Inflammatory Bowel Disease.* Nutrients, 2019. **11**(6).
19. Vangoitsenhoven, R. and G.A.M. Cresci, *Role of Microbiome and Antibiotics in Autoimmune Diseases.* Nutr Clin Pract, 2020. **35**(3): p. 406-416.
20. Zheng, D., T. Liwinski, and E. Elinav, *Interaction between microbiota and immunity in health and disease.* Cell Res, 2020. **30**(6): p. 492-506.
21. Nguyen, Q.N., et al., *The Impact of the Gut Microbiota on Humoral Immunity to Pathogens and Vaccination in Early Infancy.* PLoS Pathog, 2016. **12**(12): p. e1005997.
22. Chen, G.Y. and T.S. Stappenbeck, *Mucus, it is not just a static barrier.* Sci Signal, 2014. **7**(323): p. pe11.
23. Shan, M., et al., *Mucus enhances gut homeostasis and oral tolerance by delivering immunoregulatory signals.* Science, 2013. **342**(6157): p. 447-53.

24. Bansal, T., et al., *The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 228-33.
25. Tezuka, H. and T. Ohteki, *Regulation of IgA Production by Intestinal Dendritic Cells and Related Cells.* Front Immunol, 2019. **10**: p. 1891.
26. Elemam, N.M., S. Hannawi, and A.A. Maghazachi, *Innate Lymphoid Cells (ILCs) as Mediators of Inflammation, Release of Cytokines and Lytic Molecules.* Toxins (Basel), 2017. **9**(12).
27. Maerz, J.K., et al., *Bacterial Immunogenicity Is Critical for the Induction of Regulatory B Cells in Suppressing Inflammatory Immune Responses.* Front Immunol, 2019. **10**: p. 3093.
28. Tanoue, T., et al., *A defined commensal consortium elicits CD8 T cells and anti-cancer immunity.* Nature, 2019. **565**(7741): p. 600-605.
29. Kawamoto, S., et al., *The inhibitory receptor PD-1 regulates IgA selection and bacterial composition in the gut.* Science, 2012. **336**(6080): p. 485-9.
30. Jochum, L. and B. Stecher, *Label or Concept - What Is a Pathobiont?* Trends Microbiol, 2020. **28**(10): p. 789-792.
31. Dethlefsen, L., M. McFall-Ngai, and D.A. Relman, *An ecological and evolutionary perspective on human-microbe mutualism and disease.* Nature, 2007. **449**(7164): p. 811-8.
32. George Kerry, R., et al., *Benefaction of probiotics for human health: A review.* J Food Drug Anal, 2018. **26**(3): p. 927-939.
33. Korpela, K., et al., *Probiotic supplementation restores normal microbiota composition and function in antibiotic-treated and in caesarean-born infants.* Microbiome, 2018. **6**(1): p. 182.
34. Gong, D., et al., *Involvement of Reduced Microbial Diversity in Inflammatory Bowel Disease.* Gastroenterol Res Pract, 2016. **2016**: p. 6951091.
35. Mirsepasi-Lauridsen, H.C., et al., *Escherichia coli Pathobionts Associated with Inflammatory Bowel Disease.* Clin Microbiol Rev, 2019. **32**(2).
36. Waidmann, M., et al., *Bacteroides vulgatus protects against Escherichia coli-induced colitis in gnotobiotic interleukin-2-deficient mice.* Gastroenterology, 2003. **125**(1): p. 162-77.
37. Gronbach, K., et al., *Endotoxicity of lipopolysaccharide as a determinant of T-cell-mediated colitis induction in mice.* Gastroenterology, 2014. **146**(3): p. 765-75.
38. Willer, Y., B. Muller, and D. Bumann, *Intestinal inflammation responds to microbial tissue load independent of pathogen/non-pathogen discrimination.* PLoS One, 2012. **7**(5): p. e35992.
39. Yang, J., *Th17 Cells*, in *Translational Immunology*, S.-L. Tan, Editor. 2016, Academic Press. p. 133-163.
40. Maerz, J.K., et al., *Outer membrane vesicles blebbing contributes to B. vulgatus mpk-mediated immune response silencing.* Gut Microbes, 2018. **9**(1): p. 1-12.
41. Mazmanian, S.K., J.L. Round, and D.L. Kasper, *A microbial symbiosis factor prevents intestinal inflammatory disease.* Nature, 2008. **453**(7195): p. 620-625.
42. Ramakrishna, C., et al., *Bacteroides fragilis polysaccharide A induces IL-10 secreting B and T cells that prevent viral encephalitis.* Nature Communications, 2019. **10**(1): p. 2153.
43. Blandford, L.E., et al., *Promoter orientation of the immunomodulatory Bacteroides fragilis capsular polysaccharide A (PSA) is off in individuals with inflammatory bowel disease (IBD).* Gut Microbes, 2019. **10**(5): p. 569-577.
44. Franchi, L., et al., *Function of Nod-like receptors in microbial recognition and host defense.* Immunol Rev, 2009. **227**(1): p. 106-28.
45. Steimle, A., et al., *Weak Agonistic LPS Restores Intestinal Immune Homeostasis.* Mol Ther, 2019. **27**(11): p. 1974-1991.
46. Maaser, C., et al., *ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications.* J Crohns Colitis, 2019. **13**(2): p. 144-164.
47. Shaoul, R. and A.S. Day, *An Overview of Tools to Score Severity in Pediatric Inflammatory Bowel Disease.* Front Pediatr, 2021. **9**: p. 615216.

48. Marchal Bressenot, A., et al., *Review article: the histological assessment of disease activity in ulcerative colitis.* Aliment Pharmacol Ther, 2015. **42**(8): p. 957-67.
49. Foell, D., H. Wittkowski, and J. Roth, *Monitoring disease activity by stool analyses: from occult blood to molecular markers of intestinal inflammation and damage.* Gut, 2009. **58**(6): p. 859-68.
50. Dai, C., M. Jiang, and M.J. Sun, *Fecal markers in the management of inflammatory bowel disease.* Postgrad Med, 2018. **130**(7): p. 597-606.
51. Chen, P., et al., *Serum Biomarkers for Inflammatory Bowel Disease.* Front Med (Lausanne), 2020. **7**: p. 123.
52. Barbarino, J.M., et al., *PharmGKB summary: cyclosporine and tacrolimus pathways.* Pharmacogenet Genomics, 2013. **23**(10): p. 563-85.
53. Loy, L., et al., *Detection and management of early stage inflammatory bowel disease: an update for clinicians.* Expert Rev Gastroenterol Hepatol, 2019. **13**(6): p. 547-555.
54. Guo, S., et al., *A Simple Fecal Bacterial Marker Panel for the Diagnosis of Crohn's Disease.* Front Microbiol, 2019. **10**: p. 1306.
55. Darfeuille-Michaud, A., et al., *High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease.* Gastroenterology, 2004. **127**(2): p. 412-21.
56. Palmela, C., et al., *Adherent-invasive Escherichia coli in inflammatory bowel disease.* Gut, 2018. **67**(3): p. 574-587.
57. Horesh, G., et al., *A comprehensive and high-quality collection of Escherichia coli genomes and their genes.* Microb Genom, 2021. **7**(2).
58. Chaudhuri, R.R. and I.R. Henderson, *The evolution of the Escherichia coli phylogeny.* Infect Genet Evol, 2012. **12**(2): p. 214-26.
59. Dubinsky, M. and J. Braun, *Diagnostic and Prognostic Microbial Biomarkers in Inflammatory Bowel Diseases.* Gastroenterology, 2015. **149**(5): p. 1265-1274 e3.
60. Garcia, M., et al., *Pseudomonas aeruginosa flagellum is critical for invasion, cutaneous persistence and induction of inflammatory response of skin epidermis.* Virulence, 2018. **9**(1): p. 1163-1175.
61. Stephenson, H.N., et al., *Pseudaminic acid on Campylobacter jejuni flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction.* J Infect Dis, 2014. **210**(9): p. 1487-98.
62. McCarter, L.L., *Regulation of flagella.* Curr Opin Microbiol, 2006. **9**(2): p. 180-6.
63. Namba, K. and F. Vonderviszt, *Molecular architecture of bacterial flagellum.* Q Rev Biophys, 1997. **30**(1): p. 1-65.
64. Grognot, M. and K.M. Taute, *More than propellers: how flagella shape bacterial motility behaviors.* Curr Opin Microbiol, 2021. **61**: p. 73-81.
65. Yonekura, K., S. Maki-Yonekura, and K. Namba, *Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy.* Nature, 2003. **424**(6949): p. 643-50.
66. Yoon, S.I., et al., *Structural basis of TLR5-flagellin recognition and signaling.* Science, 2012. **335**(6070): p. 859-64.
67. Muskotal, A., et al., *Structural basis for stabilization of the hypervariable D3 domain of Salmonella flagellin upon filament formation.* J Mol Biol, 2010. **403**(4): p. 607-15.
68. Beatson, S.A., T. Minamino, and M.J. Pallen, *Variation in bacterial flagellins: from sequence to structure.* Trends Microbiol, 2006. **14**(4): p. 151-5.
69. Sebestyen, A., et al., *The hypervariable D3 domain of Salmonella flagellin is an autonomous folding unit.* Protein Pept Lett, 2008. **15**(1): p. 54-7.
70. Miao, E.A., et al., *TLR5 and Ipaf: dual sensors of bacterial flagellin in the innate immune system.* Semin Immunopathol, 2007. **29**(3): p. 275-88.
71. Adrian L. Smith, S.R.F., *Pattern recognition receptors,*, in *Avian Immunology (Third Edition),*, K.A.S. Bernd Kaspers, Thomas W. Göbel, Lonneke Vervelde,, Editor. 2022, Academic Press,. p. 231-248,.

72. Vijay-Kumar, M. and A.T. Gewirtz, *Flagellin: key target of mucosal innate immunity.* Mucosal Immunol, 2009. **2**(3): p. 197-205.
73. Ivicak-Kocjan, K., et al., *Determination of the physiological 2:2 TLR5:flagellin activation stoichiometry revealed by the activity of a fusion receptor.* Biochem Biophys Res Commun, 2013. **435**(1): p. 40-5.
74. Yu, Y., S. Sitaraman, and A.T. Gewirtz, *Intestinal epithelial cell regulation of mucosal inflammation.* Immunol Res, 2004. **29**(1-3): p. 55-68.
75. Hajam, I.A., et al., *Bacterial flagellin-a potent immunomodulatory agent.* Exp Mol Med, 2017. **49**(9): p. e373.
76. Sanders, C.J., et al., *Both radioresistant and hemopoietic cells promote innate and adaptive immune responses to flagellin.* J Immunol, 2008. **180**(11): p. 7184-92.
77. Vijay-Kumar, M., et al., *Deletion of TLR5 results in spontaneous colitis in mice.* J Clin Invest, 2007. **117**(12): p. 3909-21.
78. Rhee, S.H., et al., *Pathophysiological role of Toll-like receptor 5 engagement by bacterial flagellin in colonic inflammation.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13610-5.
79. Carvalho, F.A., et al., *Crohn's disease-associated Escherichia coli LF82 aggravates colitis in injured mouse colon via signaling by flagellin.* Inflamm Bowel Dis, 2008. **14**(8): p. 1051-60.
80. Hawn, T.R., et al., *A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease.* J Exp Med, 2003. **198**(10): p. 1563-72.
81. Gewirtz, A.T., et al., *Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease.* Am J Physiol Gastrointest Liver Physiol, 2006. **290**(6): p. G1157-63.
82. Wang, L., et al., *Species-wide variation in the Escherichia coli flagellin (H-antigen) gene.* J Bacteriol, 2003. **185**(9): p. 2936-43.
83. Smith, K.D., et al., *Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility.* Nat Immunol, 2003. **4**(12): p. 1247-53.
84. Lodes, M.J., et al., *Bacterial flagellin is a dominant antigen in Crohn disease.* J Clin Invest, 2004. **113**(9): p. 1296-306.
85. Shannon, C.E., *A Mathematical Theory of Communication.* Bell System Technical Journal, 1948. **27**(3): p. 379-423.
86. Ortiz-Burgos, S., *Shannon-Weaver Diversity Index*, in *Encyclopedia of Estuaries*, M.J. Kennish, Editor. 2016, Springer Netherlands: Dordrecht. p. 572-573.
87. Kruis, W., et al., *Maintaining remission of ulcerative colitis with the probiotic Escherichia coli Nissle 1917 is as effective as with standard mesalazine.* Gut, 2004. **53**(11): p. 1617-23.
88. Scaldaferri, F., et al., *Role and mechanisms of action of Escherichia coli Nissle 1917 in the maintenance of remission in ulcerative colitis patients: An update.* World J Gastroenterol, 2016. **22**(24): p. 5505-11.
89. Jiminez, J.A., et al., *Animal models to study acute and chronic intestinal inflammation in mammals.* Gut Pathog, 2015. **7**: p. 29.
90. Mori, J., et al., *Chimeric flagellin as the self-adjuvanting antigen for the activation of immune response against Helicobacter pylori.* Vaccine, 2012. **30**(40): p. 5856-63.
91. Hochuli, E., et al., *Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent.* Bio/Technology, 1988. **6**(11): p. 1321-1325.
92. Sleight, S.C. and R.E. Lenski, *Evolutionary adaptation to freeze-thaw-growth cycles in Escherichia coli.* Physiol Biochem Zool, 2007. **80**(4): p. 370-85.
93. Mehmood, Y., *What is Limulus amebocyte lysate (LAL) and its applicability in endotoxin quantification of pharma products*, in *Growing and handling of bacterial cultures*. 2019, IntechOpen.
94. Kunjithapatham, R., et al., *Role for the alpha-helix in aberrant protein aggregation.* Biochemistry, 2005. **44**(1): p. 149-56.

95. Song, W.S., et al., *A conserved TLR5 binding and activation hot spot on flagellin.* Sci Rep, 2017. **7**: p. 40878.
96. Kreutzberger, M.A.B., et al., *Flagellin outer domain dimerization modulates motility in pathogenic and soil bacteria from viscous environments.* Nat Commun, 2022. **13**(1): p. 1422.
97. McSorley, S.J., et al., *Bacterial flagellin is an effective adjuvant for CD4+ T cells in vivo.* J Immunol, 2002. **169**(7): p. 3914-9.
98. *Ulcerative colitis.* Nature Reviews Disease Primers, 2020. **6**(1): p. 73.
99. Tilg, H. and A.R. Moschen, *Food, immunity, and the microbiome.* Gastroenterology, 2015. **148**(6): p. 1107-19.
100. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.
101. Rinninella, E., et al., *What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases.* Microorganisms, 2019. **7**(1).
102. Zuo, T. and S.C. Ng, *The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory Bowel Disease.* Front Microbiol, 2018. **9**: p. 2247.
103. Levy, M., et al., *Dysbiosis and the immune system.* Nat Rev Immunol, 2017. **17**(4): p. 219-232.
104. Buret, A.G., et al., *Pathobiont release from dysbiotic gut microbiota biofilms in intestinal inflammatory diseases: a role for iron?* J Biomed Sci, 2019. **26**(1): p. 1.
105. Cahill, R.J., et al., *Inflammatory bowel disease: an immunity-mediated condition triggered by bacterial infection with Helicobacter hepaticus.* Infect Immun, 1997. **65**(8): p. 3126-31.
106. Gombosova, L., et al., *Genes of intestinal Escherichia coli and their relation to the inflammatory activity in patients with ulcerative colitis and Crohn's disease.* Folia Microbiol (Praha), 2011. **56**(5): p. 367-72.
107. Abdelhalim, K.A., A. Uzel, and N.G. Unal, *The role of major virulence factors and pathogenicity of adherent-invasive Escherichia coli in patients with Crohn's disease.* Prz Gastroenterol, 2020. **15**(4): p. 279-288.
108. Heras, B., M.J. Scanlon, and J.L. Martin, *Targeting virulence not viability in the search for future antibacterials.* Br J Clin Pharmacol, 2015. **79**(2): p. 208-15.
109. Bringer, M.A., et al., *The oxidoreductase DsbA plays a key role in the ability of the Crohn's disease-associated adherent-invasive Escherichia coli strain LF82 to resist macrophage killing.* J Bacteriol, 2007. **189**(13): p. 4860-71.
110. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
111. Sharon, B.M., et al., *Complete Genome Sequences of Seven Uropathogenic Escherichia coli Strains Isolated from Postmenopausal Women with Recurrent Urinary Tract Infection.* Microbiol Resour Announc, 2020. **9**(33).
112. Shala-Lawrence, A., et al., *The interaction of TraW and TrbC is required to facilitate conjugation in F-like plasmids.* Biochem Biophys Res Commun, 2018. **503**(4): p. 2386-2392.
113. Stecher, B., et al., *Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1269-74.
114. Elhenawy, W., et al., *High-throughput fitness screening and transcriptomics identify a role for a type IV secretion system in the pathogenesis of Crohn's disease-associated Escherichia coli.* Nat Commun, 2021. **12**(1): p. 2032.
115. Martinez-Medina, M., et al., *Biofilm formation as a novel phenotypic feature of adherent-invasive Escherichia coli (AIEC).* BMC Microbiol, 2009. **9**: p. 202.
116. Srivastava, A., et al., *Gut biofilm forming bacteria in inflammatory bowel disease.* Microb Pathog, 2017. **112**: p. 5-14.
117. M Chandler P, S., *Insertion Sequences*, in *Brenner's Encyclopedia of Genetics (Second Edition)*, K.H. Stanley Maloy, Editor. 2013, Academic Press.
118. Carvalho, R., et al., *Genomic Characterization of Multidrug-Resistant Escherichia coli BH100 Sub-strains.* Front Microbiol, 2020. **11**: p. 549254.
119. Petersen, L., et al., *Genes under positive selection in Escherichia coli.* Genome Res, 2007. **17**(9): p. 1336-43.

120. Hawkey, J., et al., *Impact of insertion sequences on convergent evolution of Shigella species.* PLoS Genet, 2020. **16**(7): p. e1008931.
121. Vogeleer, P., et al., *Life on the outside: role of biofilms in environmental persistence of Shiga-toxin producing Escherichia coli.* Front Microbiol, 2014. **5**: p. 317.
122. Mahadevan, S. and A. Wright, *A bacterial gene involved in transcription antitermination: regulation at a rho-independent terminator in the bgl operon of E. coli.* Cell, 1987. **50**(3): p. 485-94.
123. Gorke, B. and J. Stulke, *Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.* Nat Rev Microbiol, 2008. **6**(8): p. 613-24.
124. Wells, T.J., et al., *EhaA is a novel autotransporter protein of enterohemorrhagic Escherichia coli O157:H7 that contributes to adhesion and biofilm formation.* Environ Microbiol, 2008. **10**(3): p. 589-604.
125. Booker, T.R., B.C. Jackson, and P.D. Keightley, *Detecting positive selection in the genome.* BMC Biol, 2017. **15**(1): p. 98.
126. Anisimova, M., J.P. Bielawski, and Z. Yang, *Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution.* Mol Biol Evol, 2001. **18**(8): p. 1585-92.
127. Lynam, A.L., et al., *Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults.* Diagn Progn Res, 2020. **4**: p. 6.
128. Kittana, H., et al., *Commensal Escherichia coli Strains Can Promote Intestinal Inflammation via Differential Interleukin-6 Production.* Front Immunol, 2018. **9**: p. 2318.
129. Ramos, H.C., M. Rumbo, and J.C. Sirard, *Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa.* Trends Microbiol, 2004. **12**(11): p. 509-17.
130. Lee, S.E., et al., *Flagellin is a strong vaginal adjuvant of a therapeutic vaccine for genital cancer.* Oncoimmunology, 2016. **5**(2): p. e1081328.
131. Hayashi, F., et al., *The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5.* Nature, 2001. **410**(6832): p. 1099-103.
132. Lopez-Yglesias, A.H., et al., *FliC's Hypervariable D3 Domain Is Required for Robust Anti-Flagellin Primary Antibody Responses.* Immunohorizons, 2019. **3**(9): p. 422-432.
133. Nempont, C., et al., *Deletion of flagellin's hypervariable region abrogates antibody-mediated neutralization and systemic activation of TLR5-dependent immunity.* J Immunol, 2008. **181**(3): p. 2036-43.
134. Sonnenborn, U. and J. Schulze, *The non-pathogenic Escherichia coli strain Nissle 1917–features of a versatile probiotic.* Microbial Ecology in Health and Disease, 2009. **21**(3-4): p. 122-158.
135. Vejborg, R.M., et al., *A virulent parent with probiotic progeny: comparative genomics of Escherichia coli strains CFT073, Nissle 1917 and ABU 83972.* Mol Genet Genomics, 2010. **283**(5): p. 469-84.
136. Hancock, V., R.M. Vejborg, and P. Klemm, *Functional genomics of probiotic Escherichia coli Nissle 1917 and 83972, and UPEC strain CFT073: comparison of transcriptomes, growth and biofilm formation.* Mol Genet Genomics, 2010. **284**(6): p. 437-54.
137. Bhushan, S., et al., *Uropathogenic Escherichia coli block MyD88-dependent and activate MyD88-independent signaling pathways in rat testicular cells.* J Immunol, 2008. **180**(8): p. 5537-47.
138. Acharya, D., et al., *Rapid Bladder Interleukin-10 Synthesis in Response to Uropathogenic Escherichia coli Is Part of a Defense Strategy Triggered by the Major Bacterial Flagellar Filament FliC and Contingent on TLR5.* mSphere, 2019. **4**(6).
139. Hawn, T.R., et al., *Toll-like receptor polymorphisms and susceptibility to urinary tract infections in adult women.* PLoS One, 2009. **4**(6): p. e5990.
140. Pradhan, S. and A.A. Weiss, *Probiotic Properties of Escherichia coli Nissle in Human Intestinal Organoids.* mBio, 2020. **11**(4).

141. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
142. Phan, J., et al., *Structural basis for the substrate specificity of tobacco etch virus protease.* J Biol Chem, 2002. **277**(52): p. 50564-72.
143. R Core Team, *A language and environment for statistical computing*. 2021, Foundation for Statistical Computing.

# 8 Acknowledgments.

Bevor ich beginne möchte ich anmerken, dass es nicht genug Platz gibt, um alle Namen derjenigen zu nennen, die mich bei dieser Thesis unterstützt haben.
An erster Stelle möchte ich Prof. Dr. Julia-Stefanie Frick nennen. Danke das du meine Arbeit immer unterstützt hast und über deine Verpflichtungen hinaus meine Integration in Deutschland ermöglicht hast.
Ich danke Prof. Forchhammer für die Übernahme der Rolle des Zweitgutachters meiner Arbeit und für die positive Intensität seiner wissenschaftlichen Kritik während meiner Vorträge an der GRK.
Vielen Dank an Prof. Dr. Silke Peter und Prof. Dr. Samuel Wagner für ihre Unterstützung.
Ich danke Prof. Dr. Lisa Maier für ihre ständige Bereitschaft mir Hilfe anzubieten und mich in ihre Arbeitsgruppe zu „adoptieren“.
Ganz besonders möchte ich mich bei Lena Michaelis dafür bedanken, dass sie nie ihr Vertrauen in mich verloren hat, dass sie ihre geistige Schärfe geteilt hat und für ihre unendliche Geduld bei der Beantwortung meiner Fragen (meistens zum dritten oder vierten Mal!).
Ein besonderer Dank gilt auch Ulrich Schoppmeier, der diese besondere Kombination aus Genialität, Herzlichkeit und unerklärlichem Interesse besitzt, selbstlos zu helfen. Uli, ich sage es dir noch einmal: für mich bist du sowohl als Wissenschaftler als auch als Mensch ein Vorbild. Auch an Libera Lo Presti, die mir in der Endphase meiner Arbeit begegnet ist und doch maßgeblich zum Gelingen beigetragen hat herzlichsten Dank. Libera, deine Großzügigkeit mir gegenüber ist unerklärlich. Das Wort Danke deckt nur einen kleinen Teil der Dankbarkeit ab, die ich dir gegenüber habe.
Ein Danke auch an Michael Braun, der mich in die komplizierte Welt der Proteinstruktur eingeführt hat. Ich schätze Ihre Unterstützung, Ihre Wut, Ihre Geduld und Ihre Ungeduld. Auch an Georg Zocher vielen Dank für das Verständnis für meinen seltsamen Sinn für Humor.
Ich möchte auch Andrea Schäfer dafür danken, dass sie mir beigebracht hat, wie wichtig der gesunde Menschenverstand beim wissenschaftlichen Arbeiten ist. Andi, deine Wärme wird nur von deiner Intelligenz übertroffen. Ich danke auch Anna Lange für ihre Geduld und Ehrlichkeit. An Tinotenda Pesanai herzlichen Dank, da sie mir die wahre Dimension des Wortes Resilienz beigebracht hat und wie wichtig es ist, die Seele mit Musik und Rhythmus zu nähren. Tino, in deinem dritten Lebensjahrzehnt bist du schon weiser, als ich es je sein werde.
Ein abschließendes großes Dankeschön an alle meine Kollegen am Institut für Mikrobiologie, insbesondere an Fabian Rentschler, Iwan Grin, Andrea Eipper, Karo Birkle, Malte Schweers, Fabio Gratani, Stefan Lang, Ulrich Neu und natürlich an den mächtigen Udo.
Kathi, nur deine Liebe, Gesellschaft, Fürsorge und Zuneigung haben es möglich gemacht, dass dieser Prozess zum Tragen kam.
Zu guter Letzt meiner Familie, die trotz der Entfernung in meinem Herzen lebt.