Title

A conserved protein, BcmA, mediates motility, biofilm formation, and host colonisation in Adherent Invasive *Escherichia coli*

Short Title

BcmA is a novel modulator of motility in AIEC

Authors

Robert Cogger-Ward[1], Adam Collins[1], Denise McLean[1], Jacob Dehinsilu[2], Alan Huett[1*]

Affiliations

[1] School of Life Sciences, Medical School, University of Nottingham, Nottingham, UK

[2] Department of Biomedical Sciences, University of Sheffield, Sheffield, UK

* Corresponding author

E-mail: alan.huett@nottingham.ac.uk (A.H.)

Abstract

Adherent Invasive *Escherichia coli* (AIEC) is a non-diarrhoeagenic intestinal *E. coli* pathotype associated with Crohn's Disease. AIEC pathogenesis is characterised by biofilm formation, adhesion to and invasion of intestinal epithelial cells, and intracellular replication within epithelial cells and macrophages. Here, we identify and characterise a protein in the prototypical AIEC strain LF82 which is required for efficient biofilm formation and dispersal – LF82_p314. LF82 Δ*LF82_314* have defective swimming and swarming motility, indicating LF82_p314 is important for flagellar-mediated motility, and thus surface colonisation and biofilm dispersal. Flagellar morphology and chemotaxis in liquid appear unaffected by deletion of *LF82_314*, suggesting LF82_p314 does not elicit an effect on flagella biogenesis or environmental sensing. Flagellar motility has been implicated in AIEC virulence, therefore we assessed the role of LF82_p314 in host colonisation using a *Caenorhabditis elegans* model. We found that LF82 Δ*LF82_314* have an impaired ability to colonise the *C. elegans* compared to wild-type LF82. Phylogenetic analysis showed that *LF82_314* is conserved in several major enterobacterial pathogens, and suggests the gene may have been acquired horizontally in several genera. Our data suggests LF82_p314 may be a novel component in the flagellar motility pathway and is a novel determinant of AIEC colonisation. Our findings have potential implications not only for the pathogenesis of Crohn's Disease, but also for the course of infection in several major bacterial pathogens. We propose a new designation for *LF82_314*, *b*iofilm *c*oupled to *m*otility *A*, or *bcmA*.

Author summary

Adherent Invasive *Escherichia coli* (AIEC) are a group of bacteria implicated in the pathogenesis of Crohn's Disease, a chronic inflammatory bowel disease with no cure. Critical to the process of many bacterial infections is the ability of bacteria to swim towards and colonise the host surface using specialised, propeller-like appendages called flagella. In this paper, we describe a novel protein – LF82_p314 (BcmA) – which is required for efficient flagella-mediated motility and surface colonisation in AIEC. Using a nematode worm (*Caenorhabditis elegans*) infection model, we show that LF82_p314 enables effective colonisation of the *C. elegans* gut, suggesting a role for the protein during human infection.

These findings indicate BcmA is significant for initial colonisation of the human gut by AIEC, and therefore the onset of Crohn's Disease.

Introduction

Crohn's Disease (CD) is a chronic and relapsing inflammatory bowel disease presenting with frequent bloody diarrhoea, bowel obstruction, abdominal pain, and extraintestinal manifestations affecting the eyes, skin, joints, and liver (reviewed in [1–3]). CD is a complex syndrome which is understood as an unchecked and inappropriate inflammatory response to intestinal bacteria, potentiated by carriage of one or several of over 180 predisposing immune-related alleles [4–13] and their interaction with environmental risk factors such as smoking [14–16]; consumption of a "western" high-fat, low-fibre diet [17–20]; colonisation by a low-complexity, pro-inflammatory microbiome [21–23]; and carriage of CD-associated pathobionts *Mycobacterium avium* subsp. *paratuberculosis* [24,25] and Adherent Invasive *Escherichia coli* (AIEC) [26–34]. An increasing body of evidence suggests AIEC can act as a key aetiological component of CD. AIEC strains are found present in the ileal mucosa of up to 51.9% CD patients compared to 16.7% healthy controls [26,30,33], and have been shown to induce inflammation and colitis in mice carrying CD-associated *TLR5* deletions [35–37]; mice fed CD-associated "western" diets [19]; and mice with infection-associated intestinal inflammation and microbiome perturbations [37–39]. Indeed, the recent demonstration that AIEC alone can perturb simple microbiomes and instigate inflammation in a *TLR5*$^{-/-}$ mouse model [37] raises the possibility that – given a set of predisposing factors – AIEC infection may serve as a first step towards triggering the CD inflammatory cascade.

AIEC pathogenesis is classically characterised by adherence to, invasion of, and replication within intestinal epithelial cells (IECs) and macrophages [40–43]. Despite its significance in CD aetiology, however, the molecular pathogenesis of AIEC infection is comparatively poorly understood. AIEC are thought to use flagella to swim through the mucus layer in the gut [44–46], and secrete the mucolytic Vat-AIEC protease [47] to gain access to the intestinal epithelial surface. AIEC bind the epithelial surface via long polar fimbriae [48] and interactions between type 1 pili and CEACAM6 [40,49–51], a host adhesin over-expressed by CD patient intestinal epithelial cells. Epithelium-associated AIEC may be transcytosed by microfold cells into Peyer's Patches to be phagocytosed by macrophages, may actively invade

IECs, or may alternatively form biofilms on the luminal surface of the gut. Invasion is mediated by microtubule polymerisation and actin recruitment [41], and is thought in part to be facilitated *via* uncharacterised effector delivery in outer-membrane vesicles (OMVs) [52,53], and by a putative oxidoreductase, *ibeA* [54]. However, inhibited OMV release and *ibeA* deletion do not fully abrogate invasion, suggesting other, unknown factors may be involved. The mechanisms of intracellular replication remain to be elucidated, with only one protein – the oxidoreductase *dsbA* – known to be required [55]. In addition to canonical adhesion and invasion traits explored in the first descriptions of AIEC, biofilm formation [56–58], motility [44–46,59], and the ability to utilise short-chain fatty acids (SCFA's) as carbon sources [59–61] are becoming understood as determinants of AIEC pathogenesis. The discovery of elevated mucosa-associated biofilms in CD patients [29] suggests biofilms may be of specific importance in AIEC pathogenesis, warranting further investigation.

We previously conducted a high-throughput heterologous expression screen to identify putative effectors in the prototypical AIEC strain, LF82 [62]. AIEC-specific putative virulence genes were selected by comparison of the LF82 genome to several pathogenic and commensal *E. coli* reference genomes, and expressed in a HeLa cell line as GFP fusions. Automated microscopy and image analysis of putative effector-GFP fusions expressing HeLa cells allowed identification of protein subcellular localisations and co-localisations. From this screen, we identified a conserved hypothetical protein of unknown function – *LF82_314* – which self-assembles into large filaments (Fig. S1). *LF82_314* is widely conserved, and bioinformatic analysis (Table S1) suggested the gene is co-inherited with components of the General Secretion Pathway (GSP). The GSP is required for secretion of extracellular proteins, including pili (reviewed in [63]). We therefore hypothesised that *LF82_314* may encode either a novel, self-assembling pilin, or an amyloid-like biofilm matrix component. Using established biofilm and motility assays, and a *Caenorhabditis elegans* infection model, we established that LF82_p314 is required for efficient biofilm formation, motility, and host colonisation in LF82. Furthermore, bioinformatic analysis reveals *LF82_314* is conserved in a range of enterobacterial genomes, many of which are human pathogens. Because of the roles *LF82_314* plays in infection, and the potential significance of this novel virulence factor in diverse enterobacterial pathogens, we propose a new designation for *LF82_314*, *b*iofilm *c*oupled to *m*otility *A*, or *bcmA*.

Results

LF82_p314 promotes biofilm formation

To identify a role for LF82_p314 in biofilm formation, we created a clean, markerless deletion in *LF82_314*, LF82 Δ*LF82_314*. Using a microtitre plate-based crystal violet assay, we established that LF82 Δ*LF82_314* has a marked biofilm formation defect with incomplete dispersal upon biofilm maturation, when compared to wild-type LF82 (Fig. 1A). Episomal expression of LF82_p314 from p*LF82_314* complemented *LF82_314* deletion. Microscopic analysis of LF82 biofilms formed on glass cover slips revealed LF82 Δ*LF82_314* form patchier, less complete biofilms than wild-type LF82, a defect which can be also complemented by LF82_p314 expression (Fig. 1B). We theorised that the biofilm formation defect may be due to defective initial surface attachment, intercellular adhesion, or altered extracellular matrix architecture. If intercellular adhesion or extracellular matrix formation is altered by *LF82_314* deletion, LF82 Δ*LF82_314* biofilm formation may be complemented in *trans* by co-culture with wild-type LF82. We therefore conducted a *trans*-complementation assay, in which the biofilm formation of wild-type LF82 and LF82 Δ*LF82_314* mixed in a 1:1 ratio was assessed. We found that the LF82:LF82 Δ*LF82_314* mix formed biofilms of intermediate mass when compared to LF82 and LF82 Δ*LF82_314* biofilms (Fig 1C). To characterise the architecture of these mixed biofilms, LF82 and LF82 Δ*LF82_314* strains expressing sGPF2 and mScarlet-I, respectively, were generated for fluorescence microscopy of biofilms (Fig. 1D). LF82-sGFP2 and LF82 Δ*LF82_314*-mScarlet-I biofilms appear similar in extant and structure to those generated by non-fluorescent, parental strains. When mixed in a 1:1 ratio, LF82-sGFP2 and LF82 Δ*LF82_314*-mScarlet-I form biofilms composed of distinct, strain-exclusive islands, suggesting initial attachment and biofilm growth of LF82 and LF82 Δ*LF82_314* are independent of one another. Taken together with Fig. 1C, this demonstrates that the biofilm formation defect observed in LF82 Δ*LF82_314* cannot be complemented in *trans*, suggesting that LF82_p314 is unlikely to have a direct role in intercellular adhesion or biofilm matrix architecture.

To define whether LF82_p314 is likely to function as a pilin, adhesin required for initial attachment, or extracellular matrix component, we characterised the cellular localisation of LF82_p314 in biofilms and planktonic cells. Fluorescence microscopy of biofilms formed by

LF82 expressing an LF82_p314-mEmerald fusion protein showed LF82_p314 localises as cell-associated filaments – as in HeLa cells (Fig. 1E) – which align with the long axis of the bacterium. To assess the subcellular localisation of the LF82_p314 filaments, we stained live, planktonic LF82 p*LF82_314-mEmerald* with an amine-reactive succinimidyl ester dye conjugate (CF™ 633, Sigma Aldrich) to define the outer membrane, fixed the dyed cells, and them imaged by fluorescence microscopy. Pearson correlation analysis of pixel intensities was performed using CellProfiler and demonstrated a very weak correlation between CF633 and LF82_p314-mEmerald fluorescence (Pearson correlation coefficient, mean $r = 0.135 \pm 0.032$ (95% CI)). Furthermore, cross-sectional analysis of fluorescence intensity (Fig. 1F) shows two peaks of CF633 intensity, representing the cell membranes, and one peak of LF82_p314-mEmerald intensity between these peaks, suggesting that LF82_p314 filaments localise intracellularly. We also note that in cells imaged 1 hour post induction, LF82_p314 filaments are shorter than in biofilms imaged at 16 h, and localise near the cell pole, suggesting interactions with intracellular, pole-localised proteins. LF82_p314 is therefore unlikely to be a pilin or extracellular matrix component, and the biofilm defect observed in LF82 Δ*LF82_314* is not due to aberrant pilin-mediated attachment or extracellular matrix architecture.

LF82_p314 modulates flagella-mediated motility *via* an uncharacterised mechanism

In the absence of evidence for an adhesin or extracellular matrix function for LF82_p314, we reasoned that a motility defect might confer a surface colonisation defect, manifesting in an apparent biofilm formation defect, as has been shown elsewhere [64]. Accordingly, we used established soft agar motility assay methods to analyse the swimming and swarming behaviour of LF82, LF82 Δ*LF82_314*, and LF82 Δ*LF82_314* p*LF82_314*. We found that LF82 Δ*LF82_314* has notable swimming (Fig. 2A) and swarming (Fig. 2B) defects (One-way ANOVA with multiple comparisons to wild-type LF82; swim, LF82 vs LF28 Δ*LF82_314*, $p = 0.0004$; swarm, LF82 vs LF28 Δ*LF82_314*, $p = 0.0001$) when compared to wild-type LF82 at 10 and 24h post-inoculation, respectively. No significant difference was observed between LF82 and the *LF82_314*-expressing strain, LF82 Δ*LF82_314* + p*LF82_314*, demonstrating these defects are fully complemented by *LF82_314* expression. We noted that at 24h post-inoculation, both LF82 and LF82 Δ*LF82_314* on swimming plates had reached the edge of the plate; however, these plates lack the characteristic chemotactic rings observed on wild-type LF82 and LF82 Δ*LF82_314* + p*LF82_314* plates (Fig. 2A), and also often showed swarming behaviour in the

centre. We therefore theorised that the motility and/or chemotaxis systems may be defective in LF82 Δ*LF82_314*, leading to a slower rate of swimming, and/or an inappropriate response to wetness conditions. To test the chemotactic response of LF82 Δ*LF82_314*, we conducted a simple capillary-based chemotaxis assay using media with or without glucose as a chemoattractant (Fig. 2C). We found both LF82 and LF82 Δ*LF82_314* are more enriched in capillaries containing glucose than without, and no statistically significant difference was observed, suggesting chemotaxis is intact in LF82 Δ*LF82_314*. We also assessed whether the number per cell or morphology of flagella was affected by deletion of *LF82_314*, using Kodaka staining [65] and transmission electron microscopy (TEM). Kodaka staining (Fig. 2D) confirmed the presence of flagella on both wild-type LF82 and LF82 Δ*LF82_314*. Negative-stain TEM demonstrated no gross morphological differences in flagella between wild-type LF82 and LF82 Δ*LF82_314* flagella (Fig. 2E). Flagella counts from 30 TEM micrographs (Fig. 2F) revealed no difference between the numbers of flagella per flagellated cell. These data demonstrate *LF82_314* is required for efficient flagella-mediated motility; however, gross behavioural and morphological traits such as in-liquid chemotaxis and flagella biosynthesis are intact in LF82 Δ*LF82_314*, suggesting LF82_p314 elicits its effect via a more subtle, uncharacterised mechanism.

LF82_p314 is required for optimal *C. elegans* gut colonisation

Non-motile AIEC have significantly reduced virulence in *in vivo* models [45,46,66], and host-adapted AIEC are hyper-motile [59], suggesting flagella motility is critical in AIEC virulence. We therefore assayed the *in vivo* virulence of LF82 and LF82 Δ*LF82_314* using an established *C. elegans* survival assay [66]. The *C. elegans* food source strain *E. coli* OP50 was used as a negative control. We found LF82 was capable of "slow killing" *C. elegans*, and that deletion of *LF82_314* does not improve or abrogate the survival of infected *C. elegans* (Fig. 3A), suggesting *LF82_314* does not directly contribute to *C. elegans* killing by AIEC in this model. We noted, however, that worms fed wild-type LF82 consistently begin to die 1-2 days before those fed LF82 Δ*LF82_314*, and reasoned that this may be due to less efficient colonisation of the *C. elegans* gut by the less motile LF82 Δ*LF82_314*, prolonging the time required to fully establish infection. We therefore chose to assess the number of bacteria stably colonising the *C. elegans* gut at daily intervals. In worms fed on lawns containing exclusively LF82 or LF82 Δ*LF82_314*, we found no statistically significant deviation between wild-type and mutant CFU

recovered per worm gut (Fig. 3B), although mean LF82 CFU per worm gut was higher at 3 and 4 days post-infection (d.p.i.).

Reasoning that an assay in which *C. elegans* are continuously fed up to 10^{11} CFU/mL of one bacterial strain may not represent a realistic infection scenario, and that this could mask a colonisation defect, we conducted competition assays to detect whether *LF82_314* deletion impacts fitness against wild-type LF82. LF82 and LF82 Δ*LF82_314* carrying the Kanamycin-resistant pBAD18 (LF82-Kan^R and LF82 Δ*LF82_314*-Kan^R) or Chloramphenicol-resistant pBAD33 (LF82-Cm^R and LF82 Δ*LF82_314*-Cm^R) were used to allow differential selection of CFU recovered from *C. elegans*. Worms were fed on 1:1 mixes of LF82-Kan^R:LF82 Δ*LF82_314*-Cm^R or LF82-Cm^R:LF82 Δ*LF82_314*-Kan^R as above. The Competitive Indices (CI) for LF82 Δ*LF82_314* in this assay (Fig. 3C) are significantly below a "no-disadvantage" CI ratio of 1 throughout infection (one-tailed Wilcoxon match-pairs signed rank test, 1 d.p.i., $p = 0.0195$; day 3 d.p.i., $p = 0.0004$; day 6 d.p.i., $p = 0.0011$), showing LF82 Δ*LF82_314* has a gut colonisation disadvantage to wild-type LF82. To visualise the infection process, LF82-sGFP2 and LF82 Δ*LF82_314*-mScarlet-I were used to infect worms either alone, or mixed in a 1:1 ratio as above. Fluorescence microscopy of infected worms (Fig. 3D) shows that at 1 d.p.i., both LF82-sGFP2 and LF82 Δ*LF82_314*-mScarlet-I colonise the worm mouth. Fluorescence in worms fed a LF82-sGFP2:LF82 Δ*LF82_314*-mScarlet-I mix was below background levels. At 3 d.p.i. and 6 d.p.i, LF82-sGFP2 and LF82 Δ*LF82_314*-mScarlet-I successfully colonise the head and gut of worm in both mono- and co-feeding conditions, with penetration into the pseudocoelom at 6 d.p.i.; however, in concordance with Fig. 3C, LF82-sGFP2 appears to outcompete LF82 Δ*LF82_314*-mScarlet-I when co-fed to *C. elegans*. Taken together, our data suggests that although deletion of *LF82_314* does not attenuate "slow killing" of *C. elegans* in mono-feeding conditions, *LF82_314* is required for efficient colonisation of the *C. elegans* gut.

***LF82_314* is a widely-conserved, and may be horizontally transmissible**

LF82_314 is encoded by 468 DNA bases, annotated as encoding the 155 residue protein LF82_p314 [67]. All results returned by blastp and JACKHMMER searches designate *LF82_314* as a "hypothetical protein," "conserved hypothetical protein," "MULTISPECIES: hypothetical protein," "uncharacterised protein," or "conserved uncharacterised protein". *LF82_314* is located proximal to a tRNA site (*asnV*) in a region of the LF82 genome (Fig. 4A) which encodes predicted transposases (*LF82_309* and *yhhI*), integrases (*LF82_309* and *LF82_311*), a toxin-

antitoxin addiction module (*LF82_312* and *LF82_313*), a transcription factor (*LF82_774*), an endonuclease (*LF82_317*), and a helicase (*LF82_318*). The putative components of this genome neighbourhood and its proximity to a common transposable element insertion site (tRNA) led us to theorise that *LF82_314* may be encoded on an active or former mobile genetic element (MGE). MGEs are significant sources of horizontally acquired virulence factors, notable examples of which include the Shiga toxin – which is transmissible among *E. coli* strains by the *stx* bacteriophage, generating highly virulent Shiga Toxin-producing *E. coli* (STEC; reviewed in [68]) – and the *Salmonella* Typhi pathogenicity island, SPI-7 – a mosaic of conjugative elements and temperate bacteriophage insertions which encodes genes for Vi capsule synthesis and the Type III Secretion System effector, *sopE* [69,70].

We therefore sought to assess the distribution of *LF82_314* homologues in related phyla. We harvested the top 100 DNA sequences of *LF82_314* homologues returned by blastn discontiguous megablast (Table S3), and curated this list to remove strains for which 16S rRNA sequences were not readily available. This produced a list of 77 *LF82_314* homologues encoded in 68 enterobacterial genomes. Many of the strains returned by our search strategy are human pathogens (see Fig 4C). Of note are several *E. coli* strains which belong to an emergent clonal, pandemic urinary tract infection (UTI) -associated Extraintestinal Pathogenic *E. coli* (ExPEC) clade, ST131 [71]. Interestingly, blast search strategies excluding the order enterobacteriales did not return any significant results, suggesting *LF82_314* homologues are restricted to this order. We obtained 16S rRNA sequences from the 68 selected strains from SILVA, and used these to build a Maximum Likelihood phylogenetic tree. Comparison of this 16S rRNA tree (Fig. S2) with a Maximum Likelihood tree created from *LF82_314* homologues showed marked differences (Figure 4C). For example, in the *LF82_314* homologue tree, *Salmonella* sp. have a fragmented phylogeny rather than clustering as a distinct phylogenetic group as in the 16S rRNA tree. Similarly, *Shigella boydii* strains, and AIEC LF82 and NRG 857c, cluster together away from the *E. coli* ST131 clade in the *LF82_314* tree, however in the 16S rRNA tree one *E. coli* group is formed. These data suggest that *LF82_314* may have been introduced into these strains horizontally, raising the possibility that in some conditions, *LF82_314* may become a transmissible virulence factor.

Discussion

Flagella-mediated motility is a critical virulence factor in a wide variety of gram-negative bacterial pathogens. In AIEC LF82 and the closely related strain NRG 857c, flagella motility has been shown to be required for host colonisation, cell invasion, and persistence in *in vivo* models [45,46,59,66]. Flagellar biosynthesis in AIEC is regulated in a canonical fashion by the master flagellar regulators, *flhCD* and *fliA* [44,72]. The *E. coli* quorum sensing system, QseBC, is involved in regulation of *flhCD* function [73], and a novel transcription factor, NrdR, has been implicated in flagellar biosynthesis and regulation of chemotaxis gene expression [45]. The AIEC LF82 genome encodes a full complement of chemotaxis genes [67], and typical chemotactic responses have been observed in LF82 in the literature and this study, suggesting environmental sensing and motile responses are conserved.

We describe in this study a conserved hypothetical gene – *LF82_314* – which has novel functions in biofilm formation and host colonisation, which are mediated by a role in flagellar motility. Although we initially theorised LF82_p314 may function as a self-assembling pilin or biofilm extracellular matrix protein, we have found that LF82_p314 localises within bacterial cells, and is therefore likely a cytoplasmic or periplasmic protein. We have demonstrated that *LF82_314* is required for efficient swimming and swarming in soft agar motility assays; we note, however, in a soft agar swimming assay, that both LF82 and LF82 Δ*LF82_314* reach the edge of the plate at 24h post-inoculation. When considered with the defect observed at 8h post-inoculation, this observation suggests that LF82 Δ*LF82_314* swims and swarms more slowly than wild-type LF82, a phenotype which may be mediated by defective chemotaxis or flagella. LF82 Δ*LF82_314* swim plates lack chemotactic rings, suggesting aberrant chemotaxis may be responsible for the observed defect. However, the in-liquid chemotactic response to glucose, and flagella biosynthesis, are indistinguishable from wild-type LF82 in LF82 Δ*LF82_314*, suggesting neither gross morphological differences nor defective chemotactic signalling can account for the observed motility defect. LF82_p314 must therefore elicit a more subtle effect which nevertheless manifests as a notable motility defect in soft agar. Currently no model of LF82_p314 function exists. We hypothesise that LF82_p314 may be involved in modulating bacterial velocity in high-viscosity environments, or may be involved in surface sensing and transitioning in-liquid motility to surface-associated motility and adhesion. Further study towards a molecular understanding of LF82_p314's function –

297 including mapping LF82_p314 protein interactions, and studying the effects of LF82_p314 on
298 the AIEC transcriptome – are ongoing in our laboratory.

299 We demonstrate that *LF82_314* is required for efficient biofilm formation in LF82.
300 Biofilms play a role in Crohn's Disease pathology [29], and infection-associated biofilms are
301 often sources of persistence, antibiotic resistance, and tolerance [74,75]. Antibiotic therapy
302 is routinely used as an intervention in CD, and is known to temporarily ameliorate symptoms
303 in the majority of patients [76]. However, relapse during treatment is common and reportedly
304 universal [77] when treatment is halted, suggesting inflammation in relapsing CD may be due
305 to outgrowth of antibiotic resistant or surviving, tolerant bacteria, such as those in mucosa-
306 associated AIEC biofilms. LF82_p314 may therefore be of some interest as a potential anti-
307 virulence target which might potentiate more successful antibiotic treatment in CD, by
308 breaking down or inhibiting formation of drug-tolerant AIEC biofilms.

309 Of particular significance, our work demonstrates *LF82_314* is required for effective
310 colonisation of the gut in a *C. elegans* infection model. We did not observe decreased
311 virulence or colonisation by LF82 Δ*LF82_314* when worms were fed on one strain exclusively,
312 but were able to detect a clear defect when LF82 Δ*LF82_314* was in competition with wild-
313 type LF82. However, this does not imply that *LF82_314* has only a marginal effect on LF82 gut
314 colonisation. In the assay we have adapted from [66], worms are constantly fed on plates
315 prepared with bacterial concentrations between 10^{10} and 10^{11} CFU per ml of culture. In such
316 mono-feeding experiments, it is likely that bacteria at this density saturate the worm gut,
317 bringing equal numbers of wild-type and mutant bacterial cells in contact with the gut surface,
318 thus masking colonisation defects. Indeed, the colonisation defect observed in the
319 competition assay, which still saturates the gut with bacteria, suggests that in more
320 biologically relevant scenarios – such as a substantially reduced total infectious dose of LF82
321 competing against an established microbiome – LF82 Δ*LF82_314* may have a marked
322 colonisation defect. Further study to establish the role of *LF82_314* in AIEC colonisation in
323 complex polymicrobial contexts is required to test this hypothesis; however, our data
324 provides strong evidence to suggest that LF82_p314-mediated motility plays an important
325 role in host colonisation.

326 Finally, we report that *LF82_314* is widely distributed throughout the
327 Enterobacteriaceae, including several significant human pathogens, and that the gene

appears to be laterally inherited. Our analysis was limited to the top 100 blastn results; however, the wide distribution of closely related *LF82_314* homologues presented in our analysis suggests that this novel virulence factor is likely to be present in an even greater range of enterobacterial pathogens. Although our analysis does not show that the putative *LF82_314* mobile genetic element can be mobilised in the strains we have analysed, it is conceivable that this element may be transmissible from a strain not included in our analysis. This is of particular interest in the context of some of the strains we analysed, such as those belonging to the *E. coli* ST131 clade. ST131 is a clade of ExPEC associated with antibiotic-resistant recurrent UTIs, which was first identified in 2008 [71]. Among the pathogenic characteristics of ST131 are increased biofilm formation and adhesion to epithelial cells, both processes which require flagella motility, and which may be potentiated by LF82_p314. It is thought that many UTIs are seeded from a gut reservoir and colonisation of both epithelial surfaces occurs via similar mechanisms [78]. An ST131 *LF82_314* homologue may play a role both in establishment of a gut niche, as well as subsequent infection of the urinary epithelium. The presence of *LF82_314* in the genomes of numerous strains representing an emergent pathogen raises the possibility that acquisition of *LF82_314* may have been an important step in becoming such a successful pathogen.

Further work is required to understand the molecular function of LF82_p314; to assess its significance in higher-complexity infection systems; and to characterise fully its distribution, and whether this novel virulence factor is transmissible. What is clear is *LF82_314* is a novel player in flagellar-mediated motility with significance for host colonisation and biofilm formation, and is conserved in a range of important human pathogens. We therefore suggest a new designation for *LF82_314* and its homologues – *bcmA* (*b*iofilm *c*oupled to *m*otility *A*) – to facilitate future work without diverse nomenclature confusing the literature.

Materials and Methods

Strains and media

E. coli LF82, XL-1, and OP50, were grown in Lysogeny Broth (LB) or on LB agar with supplements, antibiotics, and agitation as appropriate, and incubated at 37°C unless otherwise stated. *E. coli* S17-1 carrying pMRE-Tn7-XXX plasmids were maintained at 25°C. *C.*

elegans SS104 [*glp-4(bn2)I.*] obtained from the *Caenorhabditis* Genetics Centre were cultured as in [79]. Strains used in this study are listed in Table 1.

Table 1

Strain	Genotype / Description	Source / Reference
Caenorhabditis elegans SS104	[*glp-4(bn2)I.*]	Caenorhabditis Genetics Centre
E. coli LF82	Prototypical AIEC type strain. Genome sequenced.	Arlette Darfeuille-Michaud, Université Clermont Auvergne
E. coli LF82 Δ*LF82_314*	LF82 with 428 bases of the *LF82_314* CDS deleted	This study
E. coli LF82 Δ*LF82_314* p*LF82_314*	*E. coli* LF82 Δ*LF82_314* with an *LF82_314*-3xFLAG contruct on pBAD18	This study
E. coli LF82-sGFP2	LF82 with chromosomally inserted sGFP2	This study
E. coli LF82 Δ*LF82_p314*-mScarlet-I	LF82 Δ*LF82_314* with chromosomally inserted mScarlet-I	This study
E. coli LF82-Kan^R	LF82 carrying pBAD18	This study
E. coli LF82-Cm^R	LF82 carrying pBAD33	This study
E. coli LF82 Δ*LF82_314*-Kan^R	LF82 ΔLF82_314 carrying pBAD18	This study
E. coli LF82 Δ*LF82_314*-Cm^R	LF82 ΔLF82_314 carrying pBAD33	This study
E. coli OP50	Uracil auxotrphic *C. elegans* food source. Genome sequenced. Tet^R	Steve Atkinson, University of Nottingham
E. coli S17-1 λ pir	Strain for conjugation of pMre-Tn7-XXX plasmids into LF82 strains	AddGene, (81)
E. coli XL-1 blue	Common cloning strain	

Strains used in this study. Cm^R = Chloramphenicol resistant; Kan^R = Kanamycin resistant; Tet^R = Tetracycline resistant.

Genetic manipulation

Genes of interest were amplified by PCR amplification using Phusion-HF DNA Polymerase (NEB). Sequences were inserted into pBAD18 or pBAD33 by restriction digest using EcoRI, KpnI, and XbaI restriction enzymes (NEB), and ligation using T4 DNA ligase (NEB). LF82

Δ*LF82_314* was generated from LF82 wild-type using the CRISPR-Cas9-based no-SCAR strategy [80]. Deletion was confirmed by Sanger sequencing, and strains were fully validated by whole genome Illumina sequencing (MicrobesNG, Birmingham, UK). LF82-sGFP2 and LF82 Δ*LF82_314*-mScarlet-I were constructed using pMRE-Tn7-132 and pMRE-Tn7-135 respectively, as in [81]. See Table 2 for a list of plasmids used in this study, and Table 3 for a list of primers.

Table 2

Plasmid	Description
pBAD18	Expression plasmid; Kan^R
pBAD33	Expression plasmid; Cm^R
pBAD -*LF82_314*-mEmerald	Expression plasmid for LF82_p314-mEmerald fusion protein; Cm^R
pCas9-cr4	Cas9- expressing plasmid for no-SCAR deletion strategy; Tet^R
pCMV-3xFLAG-*LF82_314*	Plasmid for expression of FLAG-tagged LF82_p314 in mammalian cells; Amp^R
pCMV-mEmerald-4GS	Plasmid for expression of mEmerald-tagged proteins in mammalian cells; Amp^R
pCMV-mEmerald-*LF82_314*	Plasmid for expression of mEmerald-tagged LF82_p314 in mammalian cells; Amp^R
pKD-sgRNA-p314	λ-Red recombinase- and *LF82_314*-targeting sgRNA-expressing plasmid for no-SCAR deletion strategy; Spc^R
p*LF82_314*	Complementation plasmid for LF82 Δ*LF82_314* encoding an LF82_p314-3xFLAG protein on pBAD18; Kan^R
pMRE-Tn7-132	Conjugative suicide plasmid for transposon cloning of sGFP2 construct into chromosomal sites; Amp^R, Cm^R
pMRE-Tn7-135	Conjugative suicide plasmid for transposon cloning of mScarlet-I construct into chromosomal sites; Amp^R, Cm^R

Plasmids used in this study. Amp^R = Ampicillin resistant; Cm^R = Chloramphenicol resistant; Kan^R = Kanamycin resistant; Spc^R = Spectinomycin resistant; Tet^R = Tetracycline resistant.

Table 3

Primer	Sequence	Description
3xFLAG F	GAggtaccGACTACAAGGACGACGATG	Amplifying 3xFLAG tag for insertion into pBAD plasmids
3xFLAG R	GAtctagactaGCCCCCTCCACCAATTCG	
LF82_314 F	GCgaatccAGGAGGAtaaataATGGCTACGATCCCCAC	Amplifying *LF82_314* from LF82 genome for insertion into pBAD plasmids
LF82_314 R	CCggtaccTGCTTTGGCCTCCACACC	
mEmerald F	TAggtaccATGGTGAGCAAGGGCGAG	Amplifying mEmerald for insertion into pBAD plasmids
mEmerald R	GCtctagaCTACTTGTACAGCTCGTCCATG	
p314_del F	GAGGAACATCAGCGATAGC	Confirming *LF82_314* deletion
p314_del R	GCGCGCTTAGCTACACC	
pTET_com	CCAATTGTCCATATTGCATCA	Amplifying pKD-sgRNA-p314 in two parts
pTET_p314	GTCTGACTCTGCTGCAACCCGTGCTCAGTATCTCTATCACTGA	
sgRNA_com	TTTATAACCTCCTTAGAGCTCGA	Amplifying pKD-sgRNA-p314 in two parts
sgRNA p314	GGGTTGCAGCAGAGTCAGACGTTTTAGAGCTAGAAATAGCAAG	

Primers used in this study. Upper-case letters denote complimentary sequences; lower-case bases represent restriction sites

Light microscopy

Light microscopy was conducted using an Olympus BX51 microscope at appropriate magnifications, using µManager software [82].

Electron microscopy

EM images were captured using a Tecnai T12 BioTwin Transmission Electron Microscope at an accelerating voltage of 100 kV. Images were captured using a Megaview III Soft Imaging System (SIS) camera.

Biofilm assays

Crystal Violet (CV) biofilm assays were adapted from [83]. Overnight bacterial cultures were diluted 1:100 in LB, and 100 µl diluted culture was inoculated into each well of a 96-well microtitre plate before static incubation at 37°C. For *trans*-complementation assays, diluted

cultures were mixed in specified ratios before inoculation. At appropriate intervals, planktonic bacteria were removed from the plate, and biofilms were washed three times with phosphate buffered saline (PBS). Washed biofilms were stained with 0.1% CV dissolved in water. CV was removed, and stained biofilms were washed four times with PBS, before the plates were dried in a laminar flow cabinet. Dry stain was solubilised in 30% glacial acetic acid and moved to a clean 96-well plate. The OD_{550} of solubilised CV was read using an automated plate reader. Each experiment contained 3-4 technical replicates of 4 biological replicates.

Biofilm microscopy

Biofilms were grown for microscopy on acid-washed coverslips. Coverslips were placed in 12-well plates, which were then inoculated with 500 µl bacterial culture diluted as above. Inoculated plates were inclined at a 45° angle to ensure the air-liquid interface bisected the coverslip, and were incubated at 37°C for 16 h. At 16 h, culture media was aspirated, and biofilms were fixed in 4% formaldehyde in PBS for 1 h. Fixed biofilms were washed three times with PBS, and coverslips were mounted on slides in a 90% glycerol mounting medium with 0.1% DABCO (Sigma), before imaging at 40x and 100x magnification.

Motility assays

Motility was assessed using established soft agar protocols. 5µl of saturated overnight culture was inoculated into the centre of soft LB agar plates, solidified with either 0.15% (swimming) or 0.25% (swarming) agar (Sigma) supplemented with 0.4% glucose. Plates were incubated at 37°C. At appropriate intervals, the maximum diameter of the resulting bacterial cloud or swarm was determined, and plates were imaged using a handheld camera.

Chemotaxis assays

A chemotaxis assay was modified from [84]. 75mm Haemocrit capillary tubes (Hawksley & Sons Ltd, catalogue no. 01604-00) were sealed at one end in a Bunsen flame, before being passed quickly through the flame several times to heat the glass. Heated capillaries were immediately placed open-end down into LB with or without 0.4% (w/v) glucose, and left to draw in media for 15 minutes. Overnight cultures were diluted 1:100 in fresh LB, and inoculated into the wells of a 96-well plate. Media-loaded capillaries were placed into inoculated wells, and the plate was incubated in a laminar flow cabinet at room temperature

for 1 h. To recover bacteria, the outside of capillaries were washed with water, and the sealed ends were broken over tubes containing fresh LB to catch escaping culture. Remaining culture was removed by pipetting. Recovered bacteria were then plated at appropriate dilutions for colony forming unit (CFU) enumeration.

Flagella staining

To prepare bacteria for light microscopy, overnight cultures were diluted 1:33 in fresh LB and incubated at 37°C with agitation for 3 hours, before being spread onto glass slides and stained as in [65]. Stained bacteria were then mounted in immersion oil under a cover slip, sealed with nail varnish, and imaged at 100x magnification. TEM samples were prepared in a protocol modified from [85]. Bacterial cultures were prepared as above, before being absorbed onto carbon-coated copper grids (EM Resolutions) for 10 minutes. Excess fluid was blotted away, and bacteria were fixed in 3% glutaraldehyde in 0.1 M sodium cacodylate buffer for 2 minutes. Fixed samples were washed three times with 0.1 M sodium cacodylate buffer for 10 seconds. Samples were then stained using 2% phosphotungstic acid before imaging at 6000x magnification as above.

Surface staining

To stain the outer surface of LF82, we employed an amine-reactive dye conjugation, modified from [86]. An overnight culture of LF82 p*LF82_314-mEmerald* was sub-cultured as for flagella staining, and LF82_p314 expression was induced with 0.1% arabinose. At appropriate intervals, 1 ml culture was spun at 1200 x *g* for 10 minutes in a 15ml round-bottom tube, washed twice in 1 ml PBS, and resuspended in 100 µl PBS using minimal agitation. Cells were stained at 37°C using 300 µg/ml CF-633 succinimidyl ester (Sigma) for 30 minutes with agitation at 100 rpm, then washed in PBS. Finally, bacteria were fixed in 4% formaldehyde in PBS for 20 minutes, washed, and mounted on 2% agarose pads for microscopy. Images were captured using a 100x lens.

***C. elegans* kill assays**

A *C. elegans* infection model was modified from [66]. To preclude data being confounded by progeny, *C. elegans* SS104 cultures were synchronised as in [79], and maintained at 25°C to ensure development of sterile adult worms. To prepare kill plates, 100 µl 10x concentrated

449 overnight cultures was spread onto NGM agar plates with appropriate antibiotics, and
450 incubated at 37°C overnight. 30 synchronised L4 or young adult worms were transferred to
451 the prepared kill plates, and incubated at 25°C for up to 14 days. Plates were scored every 24
452 hours for death, and dead worms were removed. To prevent contamination, worms were
453 transferred to freshly prepared kill plates every 4 days.

454 **_C. elegans_ colonisation assays**

455 *C. elegans* colonisation assays were modified from [66]. Colonisation plates were prepared as
456 above, and 50 synchronised L4 or young adult worms were transferred to each plate, before
457 being incubated at 25°C. To assay stable gut colonisation, 10 worms were transferred to
458 freshly prepared OP50 lawns daily, and incubated at 25°C for 1 h to "wash" transient bacteria
459 from the gut and worm exterior. Worms were then picked and suspended in 1ml M9 buffer,
460 before being washed three times by pelleting at 1200 x *g* for 1 minute before removal of 750
461 µL buffer, which was replaced with fresh M9 buffer. To determine external bacterial numbers
462 following washing, a sample from the final wash was plated at appropriate dilutions on
463 selective agar. To release gut contents, the worms were homogenised by vortexing with
464 approximately 400 mg sterile 1mm diameter glass beads (BioSpec Products Inc., catalogue
465 no. 11079110) in 1% Triton-X (Sigma) in M9 buffer for 2 minutes, before plating the
466 homogenate at appropriate dilutions on selective agar. CFU per worm gut was defined as:

467 $$CFU\ per\ worm\ gut = Homogenate\ CFU\ per\ worm - External\ CFU\ per\ worm \quad (2)$$

468 Colonisation experiments were conducted at least twice, with three separate
469 biological replicates per experiment.

470 **_C. elegans_ competition assays**

471 Plates for competition assays were prepared as above, using 1:1 ratios of LF82-Kan^R:LF82
472 Δ*LF82_314*-Cm^R or LF82-Cm^R:LF82 Δ*LF82_314*-Kan^R. Plates with an input ratio substantially
473 different from 1 were discarded. CFU per worm gut was assessed as above, and a competitive
474 index (CI) was calculated. CI was defined as:

475 $$CI = \frac{(CFU\ per\ worm\ gut)_{\text{LF82 }\Delta LF82\_314}}{(CFU\ per\ worm\ gut)_{\text{LF82}}} \quad (2)$$

476 Competition assays were conducted three times, with three separate biological
477 replicates of each mix per experiment.

478 To image competition assays, infected worms were "washed" on OP50 lawns as
479 above, and immobilised in a 0.1% NaN_3 solution on a 2% agarose 0.05% NaN_3 pad on a glass
480 slide, which was sealed under a cover slip. Slides were imaged at 10x magnification.

481 **Data analysis**

482 Statistical analyses were conducted in GraphPad Prism 7. Error bars in graphs represent
483 standard deviation, unless stated otherwise. Light microscopy images were analysed and
484 processed using FIJI. Fluorescence co-localisation data analysis and Pearson correlation
485 analysis was conducted in CellProfiler. Raw *r* values were converted to z' values by Fisher's Z-
486 Transformation and used to calculate mean correlations and 95% CIs, before transformation
487 back to Pearson's *r* values for interpretation. Cross-sectional intensity measurements were
488 taken using ImageJ.

489 **Bioinformatics**

490 Protein and nucleotide sequences were retrieved from NCBI [87]. 16S rRNA sequences were
491 harvested from SILVA [88]. Protein homology searches were conducted using BLASTp and
492 JACKHMMER [89,90]. Nucleotide sequences for phylogenetic analysis were retrieved by
493 discontiguous megablast [91]. Multiple sequence alignments were generated using MAAFT
494 [92–94], before submission to PhyML [95] for automated tree generation. Trees were
495 visualised in PRESTO (Phylogenetic tReE viSualisaTiOn, available at http://www.atgc-
496 montpellier.fr/presto/).

497 ## Acknowledgements

498 The authors thank the Nanoscale and Microscale Research Centre (nmRC, University of
499 Nottingham, UK) for providing access to instrumentation. We would also like to thank Eric J G
500 Pollitt (University of Sheffield, UK) for help in preparing the manuscript.

501 ## References

502 1. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet. Elsevier Ltd; 2012;380: 1590–
503 1605. doi:10.1016/S0140-6736(12)60026-9

2. Hendy P, Hart A. A Review of Crohn's Disease. EMJ Gastroenterol. 2013;1: 116–123.

3. Levine JS, Burakoff R. Extraintestinal manifestations of inflammatory bowel disease. Gastroenterol Hepatol (N Y). 2011;7: 235–41. doi:10.1007/978-3-319-26811-8_18

4. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001;411: 603–6. doi:10.1038/35079114

5. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. Nature Publishing Group; 2010;42: 1118–25. doi:10.1038/ng.717

6. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. Nature Publishing Group; 2012;491: 119–24. doi:10.1038/nature11582

7. Kenny EE, Pe'er I, Karban A, Ozelius L, Mitchell AA, Ng SM, et al. A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. PLoS Genet. 2012;8: e1002559. doi:10.1371/journal.pgen.1002559

8. Yang SK, Hong M, Zhao W, Jung Y, Baek J, Tayebi N, et al. Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. Gut. 2014;63: 80–87. doi:10.1136/gutjnl-2013-305193

9. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet. 2007;39: 207–11. doi:10.1038/ng1954

10. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet. 2007;39: 830–2. doi:10.1038/ng2061

11. Tremelling M, Berzuini C, Massey D, Bredin F, Price C, Dawson C, et al. Contribution of

TNFSF15 gene variants to Crohn's disease susceptibility confirmed in UK population. Inflamm Bowel Dis. 2008;14: 733–737. doi:10.1002/ibd.20399

12. Csöngei V, Járomi L, Sáfrány E, Sipeky C, Magyari L, Faragó B, et al. Interaction of the major inflammatory bowel disease susceptibility alleles in Crohn's disease patients. World J Gastroenterol. 2010;16: 176–83. doi:10.3748/wjg.v16.i2.176

13. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47: 979–989. doi:10.1038/ng.3359

14. Somerville KW, Logan RF, Edmond M, Langman MJ. Smoking and Crohn's disease. Br Med J (Clin Res Ed). 1984;289: 954–6. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1443150&tool=pmcentrez&rendertype=abstract

15. Calkins BM. A meta-analysis of the role of smoking in inflammatory bowel disease. Dig Dis Sci. 1989;34: 1841–1854. doi:10.1007/BF01536701

16. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and Inflammatory Bowel Disease: A Meta-analysis. Mayo Clin Proc. Elsevier; 2016;81: 1462–1471. doi:10.4065/81.11.1462

17. Racine A, Carbonnel F, Chan SSM, Hart AR, Bueno-de-Mesquita HB, Oldenburg B, et al. Dietary Patterns and Risk of Inflammatory Bowel Disease in Europe: Results from the EPIC Study. Inflamm Bowel Dis. 2016;22: 345–54. doi:10.1097/MIB.0000000000000638

18. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 2016;167: 1339–1353.e21. doi:10.1016/j.cell.2016.10.043

19. Agus A, Denizot J, Thévenot J, Martinez-Medina M, Massier S, Sauvanet P, et al. Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. Sci Rep. Nature

561 Publishing Group; 2016;6: 19032. doi:10.1038/srep19032

562 20. Hou JK, Abraham B, El-Serag H. Dietary intake and risk of developing inflammatory
563 bowel disease: a systematic review of the literature. Am J Gastroenterol. Nature
564 Publishing Group; 2011;106: 563–573. doi:10.1038/ajg.2011.44

565 21. Wright EK, Kamm MA, Teo SM, Inouye M, Wagner J, Kirkwood CD. Recent advances in
566 characterizing the gastrointestinal microbiome in Crohn's disease: a systematic
567 review. Inflamm Bowel Dis. 2015;21: 1219–1228.
568 doi:10.1097/MIB.0000000000000382

569 22. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux J-J, et
570 al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium
571 identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S
572 A. 2008;105: 16731–6. doi:10.1073/pnas.0804812105

573 23. Nagalingam NA, Kao JY, Young VB. Microbial ecology of the murine gut associated
574 with the development of dextran sodium sulfate-induced colitis. Inflamm Bowel Dis.
575 2011;17: 917–926. doi:10.1002/ibd.21462

576 24. Davis WC, Madsen-Bouterse SA. Crohn's disease and Mycobacterium avium subsp.
577 paratuberculosis: The need for a study is long overdue. Vet Immunol Immunopathol.
578 Elsevier B.V.; 2012;145: 1–6. doi:10.1016/j.vetimm.2011.12.005

579 25. Feller M, Huwiler K, Stephan R, Altpeter E, Shang A, Furrer H, et al. Mycobacterium
580 avium subspecies paratuberculosis and Crohn's disease: a systematic review and
581 meta-analysis. Lancet Infect Dis. 2007;7: 607–613. doi:10.1016/S1473-
582 3099(07)70211-6

583 26. Darfeuille-Michaud A, Neut C, Barnich N, Lederman E, Martino PDI, Desreumaux P, et
584 al. Presence of adherent Escherichia coli strains in ileal mucosa of patients with
585 Crohn's disease. Gastroenterology. 1998;115: 1405–1413. doi:10.1016/S0016-
586 5085(98)70019-8

587 27. Kotlowski R, Bernstein CN, Sepehri S, Krause DO. High prevalence of Escherichia coli
588 belonging to the B2+D phylogenetic group in inflammatory bowel disease. Gut.
589 2007;56: 669–675. doi:10.1136/gut.2006.099796

590 28. Mylonaki M, Rayment NB, Rampton DS, Hudspith BN, Brostoff J. Molecular
591 characterization of rectal mucosa-associated bacterial flora in inflammatory bowel
592 disease. Inflamm Bowel Dis. 2005;11: 481–7.
593 doi:http://dx.doi.org/10.1097/01.MIB.0000159663.62651.4f

594 29. Swidsinski A, Ladhoff A, Pernthaler A, Swidsinski S, Loening-Baucke V, Ortner M, et al.
595 Mucosal flora in inflammatory bowel disease. Gastroenterology. 2002;122: 44–54.
596 doi:10.1053/gast.2002.30294

597 30. Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser A-L, Barnich N, et al. High
598 prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in
599 Crohn's disease. Gastroenterology. 2004;127: 412–21.
600 doi:10.1053/j.gastro.2004.04.061

601 31. Baumgart M, Dogan B, Rishniw M, Weitzman G, Bosworth B, Yantiss R, et al. Culture
602 independent analysis of ileal mucosa reveals a selective increase in invasive
603 Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's
604 disease involving the ileum. ISME J. 2007;1: 403–18. doi:10.1038/ismej.2007.52

605 32. Conte MP, Longhi C, Marazzato M, Conte AL, Aleandri M, Lepanto MS, et al.
606 Adherent-invasive Escherichia coli (AIEC) in pediatric Crohn's disease patients:
607 phenotypic and genetic pathogenic features. BMC Res Notes. 2014;7: 748.
608 doi:10.1186/1756-0500-7-748

609 33. Martinez-Medina M, Aldeguer X, Lopez-Siles M, Gonz??lez-Huix F, L??pez-Oliu C,
610 Dahbi G, et al. Molecular diversity of Escherichia coli in the human gut: New
611 ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's
612 disease. Inflamm Bowel Dis. 2009;15: 872–882. doi:10.1002/ibd.20860

613 34. Darfeuille-Michaud A. Adherent-invasive Escherichia coli: a putative new E. coli
614 pathotype associated with Crohn's disease. Int J Med Microbiol. 2002;292: 185–93.
615 doi:10.1078/1438-4221-00201

616 35. Carvalho FA, Koren O, Goodrich JK, Johansson ME V, Nalbantoglu I, Aitken JD, et al.
617 Transient inability to manage proteobacteria promotes chronic gut inflammation in
618 TLR5-deficient mice. Cell Host Microbe. 2012;12: 139–152.

doi:10.1016/j.chom.2012.07.004

36. Chassaing B, Koren O, Carvalho F a, Ley RE, Gewirtz AT. AIEC pathobiont instigates chronic colitis in susceptible hosts by altering microbiota composition. Gut. 2014;63: 1069–1080. doi:10.1136/gutjnl-2013-304909

37. Chassaing B, Gewirtz AT. Mice harboring pathobiont-free microbiota do not develop intestinal inflammation that normally results from an innate immune deficiency. Chamaillard M, editor. PLoS One. 2018;13: e0195310. doi:10.1371/journal.pone.0195310

38. Small CL, Xing L, McPhee JB, Law HT, Coombes BK. Acute Infectious Gastroenteritis Potentiates a Crohn's Disease Pathobiont to Fuel Ongoing Inflammation in the Post-Infectious Period. Baumler AJ, editor. PLoS Pathog. 2016;12: e1005907. doi:10.1371/journal.ppat.1005907

39. Craven M, Egan CE, Dowd SE, McDonough SP, Dogan B, Denkers EY, et al. Inflammation drives dysbiosis and bacterial invasion in murine models of ileal Crohn's Disease. PLoS One. 2012;7: 1–10. doi:10.1371/journal.pone.0041594

40. Boudeau J, Barnich N, Darfeuille-Michaud A. Type 1 pili-mediated adherence of Escherichia coli strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. Mol Microbiol. 2001;39: 1272–84. doi:10.1046/j.1365-2958.2001.02315.x

41. Boudeau J, Glasser AL, Masseret E, Joly B, Darfeuille-Michaud A. Invasive ability of an Escherichia coli strain isolated from the ileal mucosa of a patient with Crohn's disease. Infect Immun. 1999;67: 4499–509. Available: http://www.ncbi.nlm.nih.gov/pubmed/10456892

42. Glasser AL, Boudeau J, Barnich N, Perruchot MH, Colombel JF, Darfeuille-Michaud A. Adherent invasive Escherichia coli strains from patients with Crohn's disease survive and replicate within macrophages without inducing host cell death. Infect Immun. 2001;69: 5529–37. doi:10.1128/IAI.69.9.5529

43. Bringer M-A, Glasser A-L, Tung C-H, Méresse S, Darfeuille-Michaud A. The Crohn's disease-associated adherent-invasive Escherichia coli strain LF82 replicates in mature

phagolysosomes within J774 macrophages. Cell Microbiol. 2006;8: 471–84. doi:10.1111/j.1462-5822.2005.00639.x

44. Barnich N, Boudeau J, Claret L, Darfeuille-Michaud A. Regulatory and functional co-operation of flagella and type 1 pili in adhesive and invasive abilities of AIEC strain LF82 isolated from a patient with Crohn's disease. Mol Microbiol. 2003;48: 781–94. doi:10.1046/j.1365-2958.2003.03468.x

45. Dreux N, Cendra M del M, Massier S, Darfeuille-Michaud A, Barnich N, Torrents E. Ribonucleotide reductase NrdR as a novel regulator for motility and chemotaxis during adherent-invasive Escherichia coli infection. Infect Immun. 2015;83: 1305–1317. doi:10.1128/IAI.02772-14

46. Sevrin G, Massier S, Chassaing B, Agus A, Delmas J, Denizot J, et al. Adaptation of adherent-invasive E. coli to gut environment: Impact on flagellum expression and bacterial colonization ability. Gut Microbes. Taylor & Francis; 2018;0: 1–17. doi:10.1080/19490976.2017.1421886

47. Gibold L, Garenaux E, Dalmasso G, Gallucci C, Cia D, Mottet-Auselo B, et al. The Vat-AIEC protease promotes crossing of the intestinal mucus layer by Crohn's disease-associated Escherichia coli. Cell Microbiol. 2015;18: 617–631. doi:10.1111/cmi.12539

48. Chassaing B, Rolhion N, de Vallée A, Salim SY, Prorok-Hamon M, Neut C, et al. Crohn disease--associated adherent-invasive E. coli bacteria target mouse and human Peyer's patches via long polar fimbriae. J Clin Invest. 2011;121: 966–75. doi:10.1172/JCI44632

49. Barnich N, Carvalho FA, Glasser A-L, Darcha C, Jantscheff P, Allez M, et al. CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease. J Clin Invest. 2007;117: 1566–74. doi:10.1172/JCI30504

50. Carvalho FA, Barnich N, Sivignon A, Darcha C, Chan CHF, Stanners CP, et al. Crohn's disease adherent-invasive Escherichia coli colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. J Exp Med. 2009;206: 2179–89. doi:10.1084/jem.20090741

51. Dreux N, Denizot J, Martinez-Medina M, Mellmann A, Billig M, Kisiela D, et al. Point

677 mutations in FimH adhesin of Crohn's disease-associated adherent-invasive
678 Escherichia coli enhance intestinal inflammatory response. PLoS Pathog. 2013;9:
679 e1003141. doi:10.1371/journal.ppat.1003141

680 52. Rolhion N, Barnich N, Claret L, Darfeuille-Michaud A. Strong Decrease in Invasive
681 Ability and Outer Membrane Vesicle Release in Crohn's Disease-Associated Adherent-
682 Invasive Escherichia coli Strain LF82 with the yfgL Gene Deleted. J Bacteriol. 2005;187:
683 2286–2296. doi:10.1128/JB.187.7.2286-2296.2005

684 53. Rolhion N, Barnich N, Bringer M-A, Glasser A-L, Ranc J, Hébuterne X, et al. Abnormally
685 expressed ER stress response chaperone Gp96 in CD favours adherent-invasive
686 Escherichia coli invasion. Gut. 2010;59: 1355–1362. doi:10.1136/gut.2010.207456

687 54. Cieza RJ, Hu J, Ross BN, Sbrana E, Torres AG. The IbeA invasin of adherent-invasive
688 Escherichia coli mediates nteraction with intestinal epithelia and macrophages. Infect
689 Immun. 2015;83: 1904–1918. doi:10.1128/IAI.03003-14

690 55. Bringer MA, Rolhion N, Glasser AL, Darfeuille-Michaud A. The oxidoreductase DsbA
691 plays a key role in the ability of the Crohn's disease-associated adherent-invasive
692 Escherichia coli strain LF82 to resist macrophage killing. J Bacteriol. 2007;189: 4860–
693 4871. doi:10.1128/JB.00233-07

694 56. Chassaing B, Darfeuille-Michaud A. The σE pathway is involved in biofilm formation
695 by Crohn's disease-associated adherent-invasive Escherichia coli. J Bacteriol.
696 2013;195: 76–84. doi:10.1128/JB.01079-12

697 57. Chassaing B, Garénaux E, Carriere J, Rolhion N, Guérardel Y, Barnich N, et al. Analysis
698 of the σ^E regulon in Crohn's disease-associated Escherichia coli revealed involvement
699 of the waaWVL operon in biofilm formation. J Bacteriol. 2015;197: 1451–1465.
700 doi:10.1128/JB.02499-14

701 58. Martinez-Medina M, Naves P, Blanco J, Aldeguer X, Blanco JE, Blanco M, et al. Biofilm
702 formation as a novel phenotypic feature of adherent-invasive Escherichia coli (AIEC).
703 BMC Microbiol. 2009;9: 202. doi:10.1186/1471-2180-9-202

704 59. Elhenawy W, Tsai CN, Coombes BK. Host-Specific Adaptive Diversification of Crohn's
705 Disease-Associated Adherent-Invasive Escherichia coli. Cell Host Microbe. Elsevier

706 Inc.; 2019; 1–12. doi:10.1016/j.chom.2018.12.010

707 60. Ormsby MJ, Johnson SA, Meikle LM, Goldstone RJ, McIntosh A, Wessel HM, et al.
708 Propionic acid enhances the virulence of Crohn's disease-associated adherent-
709 invasive Escherichia coli. bioRxiv. 2018; 387647. doi:10.1101/387647

710 61. Delmas J, Gibold L, Faïs T, Batista S, Leremboure M, Sinel C, et al. Metabolic
711 adaptation of adherent-invasive Escherichia coli to exposure to bile salts. Sci Rep.
712 2019;9: 2175. doi:10.1038/s41598-019-38628-1

713 62. Collins A, Huett A. A multi-phenotypic imaging screen to identify bacterial effectors by
714 exogenous expression in a HeLa cell line. Sci Data. 2018;5: 180081.
715 doi:10.1038/sdata.2018.81

716 63. Pugsley AP. The complete general secretory pathway in gram-negative bacteria.
717 Microbiol Rev. 1993;57: 50–108. doi:10.1016/0966-842X(93)90043-Q

718 64. Wood TK, González Barrios AF, Herzberg M, Lee J. Motility influences biofilm
719 architecture in Escherichia coli. Appl Microbiol Biotechnol. 2006;72: 361–367.
720 doi:10.1007/s00253-005-0263-8

721 65. Kodaka H, Armfield AY, Lombard GL, Dowell VR. Practical procedure for
722 demonstrating bacterial flagella. J Clin Microbiol. 1982;16: 948–52. Available:
723 http://www.ncbi.nlm.nih.gov/pubmed/6185531

724 66. Simonsen KT, Nielsen G, Bjerrum JV, Kruse T, Kallipolitis BH, Møller-Jensen J. A role
725 for the RNA chaperone Hfq in controlling Adherent-invasive Escherichia coli
726 colonization and virulence. PLoS One. 2011;6. doi:10.1371/journal.pone.0016387

727 67. Miquel S, Peyretaillade E, Claret L, de Vallée A, Dossat C, Vacherie B, et al. Complete
728 genome sequence of Crohn's disease-associated adherent-invasive E. coli strain LF82.
729 PLoS One. 2010;5: 481–7. doi:10.1371/journal.pone.0012714

730 68. Krüger A, Lucchesi PMA. Shiga toxins and stx phages: Highly diverse entities.
731 Microbiol (United Kingdom). 2015;161: 1–12. doi:10.1099/mic.0.000003

732 69. Pickard D, Wain J, Baker S, Line A, Chohan S, Fookes M, et al. Composition,
733 Acquisition, and Distribution of the Vi Exopolysaccharide-Encoding Salmonella

734 enterica Pathogenicity Island SPI-7. J Bacteriol. 2003;185: 5055–5065.
735 doi:10.1128/JB.185.17.5055-5065.2003

736 70. Seth-Smith HMB, Fookes MC, Okoro CK, Baker S, Harris SR, Scott P, et al. Structure,
737 diversity, and mobility of the salmonella pathogenicity island 7 family of integrative
738 and conjugative elements within enterobacteriaceae. J Bacteriol. 2012;194: 1494–
739 1504. doi:10.1128/JB.06403-11

740 71. Nicolas-Chanoine MH, Bertrand X, Madec JY. Escherichia coli st131, an intriguing
741 clonal group. Clin Microbiol Rev. 2014;27: 543–574. doi:10.1128/CMR.00125-13

742 72. Claret L, Miquel S, Vieille N, Ryjenkov DA, Gomelsky M, Darfeuille-Michaud A. The
743 flagellar sigma factor FliA regulates adhesion and invasion of Crohn disease-
744 associated Escherichia coli via a cyclic dimeric GMP-dependent pathway. J Biol Chem.
745 2007;282: 33275–33283. doi:10.1074/jbc.M702800200

746 73. Rooks MG, Veiga P, Reeves AZ, Lavoie S, Yasuda K, Asano Y, et al. QseC inhibition as
747 an antivirulence approach for colitis-associated bacteria. Proc Natl Acad Sci U S A.
748 2017;114: 142–147. doi:10.1073/pnas.1612836114

749 74. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: a common cause of
750 persistent. Infect Sci. 1999;284: 1318–1322.

751 75. Stewart PS, William Costerton J. Antibiotic resistance of bacteria in biofilms. Lancet.
752 2001;358: 135–138. doi:10.1016/S0140-6736(01)05321-1

753 76. Townsend CM, Parker CE, MacDonald JK, Nguyen TM, Jairath V, Feagan BG, et al.
754 Antibiotics for induction and maintenance of remission in Crohn's disease. Cochrane
755 Database Syst Rev. 2019; doi:10.1002/14651858.CD012730.pub2

756 77. Baran B, Karaca C. Practical Medical Management of Crohn's Disease. ISRN
757 Gastroenterol. 2013;2013: 1–12. doi:10.1155/2013/208073

758 78. Spaulding CN, Klein RD, Ruer S, Kau AL, Schreiber HL, Cusumano ZT, et al. Selective
759 depletion of uropathogenic E. coli from the gut by a FimH antagonist. Nature. Nature
760 Publishing Group; 2017;546: 528–532. doi:10.1038/nature22972

761 79. Stiernagle T. Maintenance of C. elegans. WormBook. 2006. pp. 1–11.

762 doi:10.1895/wormbook.1.101.1

763 80. Reisch CR, Prather KLJ. The no-SCAR (Scarless Cas9 Assisted Recombineering) system
764 for genome editing in Escherichia coli. Sci Rep. Nature Publishing Group; 2015;5:
765 15096. doi:10.1038/srep15096

766 81. Schlechter R, Jun H, Bernach M, Oso S, Boyd E, Munoz-Lintz D, et al. Chromatic
767 bacteria – A broad host-range plasmid and chromosomal insertion toolbox for
768 fluorescent protein expression in bacteria. bioRxiv. 2018; 1–29. doi:10.1101/402172

769 82. Schneider C a, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image
770 analysis. Nat Methods. Nature Publishing Group; 2012;9: 671–675.
771 doi:10.1038/nmeth.2089

772 83. O'Toole GA. Microtiter dish biofilm formation assay. J Vis Exp. 2011; 10–11.
773 doi:10.3791/2437

774 84. Adler J. A Method for Measuring Chemotaxis and Use of the Method to Determine
775 Optimum Conditions for Chemotaxis by Escherichia coli. J Gen Microbiol. 1973;74:
776 77–91. doi:10.1099/00221287-74-1-77

777 85. Kirov SM, Tassell BC, Semmler ABT, O'Donovan LA, Rabaan AA, Shaw JG. Lateral
778 Flagella and Swarming Motility in Aeromonas Species. J Bacteriol. 2002;184: 547–555.
779 doi:10.1128/JB.184.2.547-555.2002

780 86. Turner L, Berg HC. Labeling bacterial flagella with fluorescent dyes. Methods Mol Biol.
781 2018;1729: 71–76. doi:10.1007/978-1-4939-7577-8_7

782 87. NCBI Resource Coordinators. Database Resources of the National Center for
783 Biotechnology Information. Nucleic Acids Res. 2017;45: D12–D17.
784 doi:10.1093/nar/gkw1071

785 88. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal
786 RNA gene database project: Improved data processing and web-based tools. Nucleic
787 Acids Res. 2013;41: 590–596. doi:10.1093/nar/gks1219

788 89. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database
789 search programs. Nucleic Acids Res. 1997;25: 3389–3402.

doi:10.1093/nar/25.17.3389

90. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. Nucleic Acids Res. 2015;43: W30–W38. doi:10.1093/nar/gkv397

91. Zhang Z, Schwartz S, Wagner L, Miller W. A Greedy Algorithm for Aligning DNA Sequences. J Comput Biol. 2000;7: 203–214. doi:10.1089/10665270050081478

92. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

93. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 2015;43: W580–W584. doi:10.1093/nar/gkv279

94. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 2013;41: 597–600. doi:10.1093/nar/gkt376

95. Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017;34: 2422–2424. doi:10.1093/molbev/msx149

Fig 1. *LF82_314* is required for optimal biofilm formation

(A) LF82 Δ*LF82_314* do not form biofilms as strongly as wild-type LF82, and LF82 Δ*LF82_314* biofilms do not appear to mature and disperse as readily as wild-type biofilms. Wild-type biofilm formation and dispersal behaviour is restored by episomal expression of *LF82_314*. Coloured asterisks represent significant difference between mutant (magenta) and complemented (black) groups, and LF82 (two-way ANOVA with multiple comparisons to LF82; see Table S2 for significance levels). (B) Microscopic analysis of biofilms at 16 h shows LF82 Δ*LF82_314* form patchy, less dense biofilms than wild-type LF82, and that this phenotype can be complemented by *LF82_314* expression. (C) To assess whether LF82_p314 functions in *trans*, we assessed the biofilm formation of wild-type LF82, LF82 Δ*LF82_314*, and a 1:1 mix of LF82:LF82 Δ*LF82_314*. An LF82 Δ*LF82_314* biofilm defect was observed as in (A); however, a 1:1 mix of LF82:LF82 Δ*LF82_314* displayed an intermediate phenotype. Coloured asterisks represent significant difference between mutant (magenta) and mixed (black) cultures, and LF82 (two-way ANOVA with multiple comparisons to LF82; see Table S1 for significance levels). (D) LF82-sGFP2 (green) and LF82 Δ*LF82_314*_mScarlet-I (magenta) form biofilms comparable to non-fluorescent LF82 and LF82 Δ*LF82_314*. A 1:1 co-culture of LF82-sGFP2 and LF82 Δ*LF82_314*_mScarlet-I show that mixed biofilm are composed of strain-exclusive islands. (E) Biofilms in which LF82 express an LF82_p314-mEmerald fusion protein (green) shows LF82_p314 forms cell-associated filaments, as observed in HeLa cells [62]. (F) Fluorescence intensity cross-section analysis of LF82 expressing *LF82_314-mEmerald* with CF633 succinimidyl ester-stained outer membrane proteins shows the extracellular stain and mEmerald fluorescence intensity peaks do not overlap, demonstrating an intracellular localisation for *LF82_314* filaments. White lines represent the 30 pixels of the cross-section. Cyan = CF633; green = LF82_p314-mEmerald.

Fig. 2. *LF82_314* promotes flagella-mediated motility *via* an uncharacterised mechanism.

LF82 Δ*LF82_314* has notable defects in (A) swimming and (B) swarming motility (One-way ANOVA, *** = $p \leq 0.001$), which are complemented by *LF82_314* expression. Each dot represents one technical replicate, or plate; separate colours represent biological replicates. Swim plates were measured at 10 h post-inoculation, and swarm plates at 24h. Motility plates imaged at 24 h post-inoculation show LF82 Δ*LF82_314* have atypical swimming motility lacking chemotactic rings observed on wild-type plates, and the swarming defect. (C) A

847 capillary-based chemotaxis assay demonstrated increased recovery of both LF82 and LF82
848 Δ*LF82_314* CFU from media supplemented with glucose compared to LB alone, with no
849 significant difference between strains in CFU recovered in either condition. Each dot
850 represents one biological replicate. Bright-field microscopy of Kodaka stained (D) and
851 negative stain TEM (E) of LF82 and LF82 Δ*LF82_314* demonstrates no differences in flagella
852 morphology, and flagella counts from 30 TEM micrographs (F) show no difference in flagella
853 numbers per flagellated cell, suggesting flagella biosynthesis is intact in both strains. TEM
854 scale bar represents 5 µm.

855 **Fig. 3. *LF82_314* promotes gut colonisation in *Caenorhabditis elegans***

856 (A) Survival of *C. elegans* SS104 is significantly decreased when cultivated on LF82 or LF82
857 Δ*LF82_314*, compared to OP50, however no significant difference was noted between
858 survival on LF82 or LF82 Δ*LF82_314*. (B) No significant difference in stable colonisation of the
859 *C. elegans* gut by LF82 or LF82 Δ*LF82_314* was found when worm were fed on each strain
860 exclusively. (C) A competition assay demonstrated LF82 Δ*LF82_314* has a competitive
861 colonisation disadvantage compared to LF82 throughout the course of infection (one-tailed
862 Wilcoxon match-pairs signed rank test, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$. CI's were
863 compared to an ideal "no-disadvantage" CI = 1). CI ratios below 1 represent a competitive
864 disadvantage. Data was pooled from 3 independent experiments. Error bars show 95%
865 confidence intervals. (D) Fluorescence microscopy of *C. elegans* fed LF82-sGFP2 and LF82
866 Δ*LF82_314*-mScarlet-I alone demonstrates both LF82 and LF82 Δ*LF82_314* are capable of
867 establishing gut and pseudocoelomic infections; however, when fed to worms in a 1:1 ratio,
868 LF82-sGFP2 appears to outcompete LF82 Δ*LF82_314*-mScarlet-I, mirroring Fig. 3C. These data
869 suggest that *LF82_314* is required for efficient host colonisation. Scale bars represent 50 µm.

870 **Fig. 4. *LF82_314* is conserved among enterobacterial pathogens**

871 *LF82_314* is encoded in a region of the LF82 genome (A) containing several ORFs with
872 predicted transposase and integrase functions, and a toxin-antitoxin addiction module (B),
873 suggesting the region may represent an MGE. (C) A ML tree of 77 *LF82_314* homologues from
874 68 strains reveals that *LF82_314* is conserved in a wide variety of pathogenic enterobacteria,
875 and that *LF82_314* homologue-derived phylogenies do not recapitulate expected
876 phylogenetic relationships (see Fig. S2). Of note is the large clade *E. coli* strains, which

represents members of the clonal, UTI-associated ExPEC, ST131. Red = human pathogen; gold = human and animal or zoonotic pathogen; green = plant pathogen; brown = commensal; blue = environmental; black = insufficient data. Scale bar represents number of substitutions per site.

Supporting information

S1 Fig. *LF82_314* self-assembles into large filaments

When expressed heterologously in HeLa cells, *LF82_314* forms large filaments. Filament formation is independent of the tag used for visualisation, suggesting this phenotype is not a tag-dependent artefact. See S3 File Supplementary methods for protocol.

S2 Fig. *LF82_314* homologue distribution does not mirror true phylogenies

The distribution of *LF82_314* homologues represented in the *LF82_314* ML phylogenetic tree (A) differs significantly from the 16S rRNA phylogenies (B). For example, in (A), *E. coli* LF82 and NRG 857c cluster with *Shigella boydii* Sb227 and ATCC 9210 in a separate clade to other *E. coli* isolates; however, in (B), LF82, NRG 857c, Sb227, and ATCC 9210 cluster as expected with *E. coli* in a distinct clade. Similarly, (A) suggests a loose relationship exists between *Salmonella* strains, whereas in (B), a distinct phylogenetic group is generated. These data suggest *LF82_314* may be horizontally inherited. Strains are coloured by genus. Scale bars = substitutions per site.

S1 File *LF82_314* is co-inherited with the general secretion pathway

S2 File *LF82_314* homologues from blastn discontiguous megablast

S3 File Supplementary methods

S1 Table p-values for Fig. 1

Fig. 1

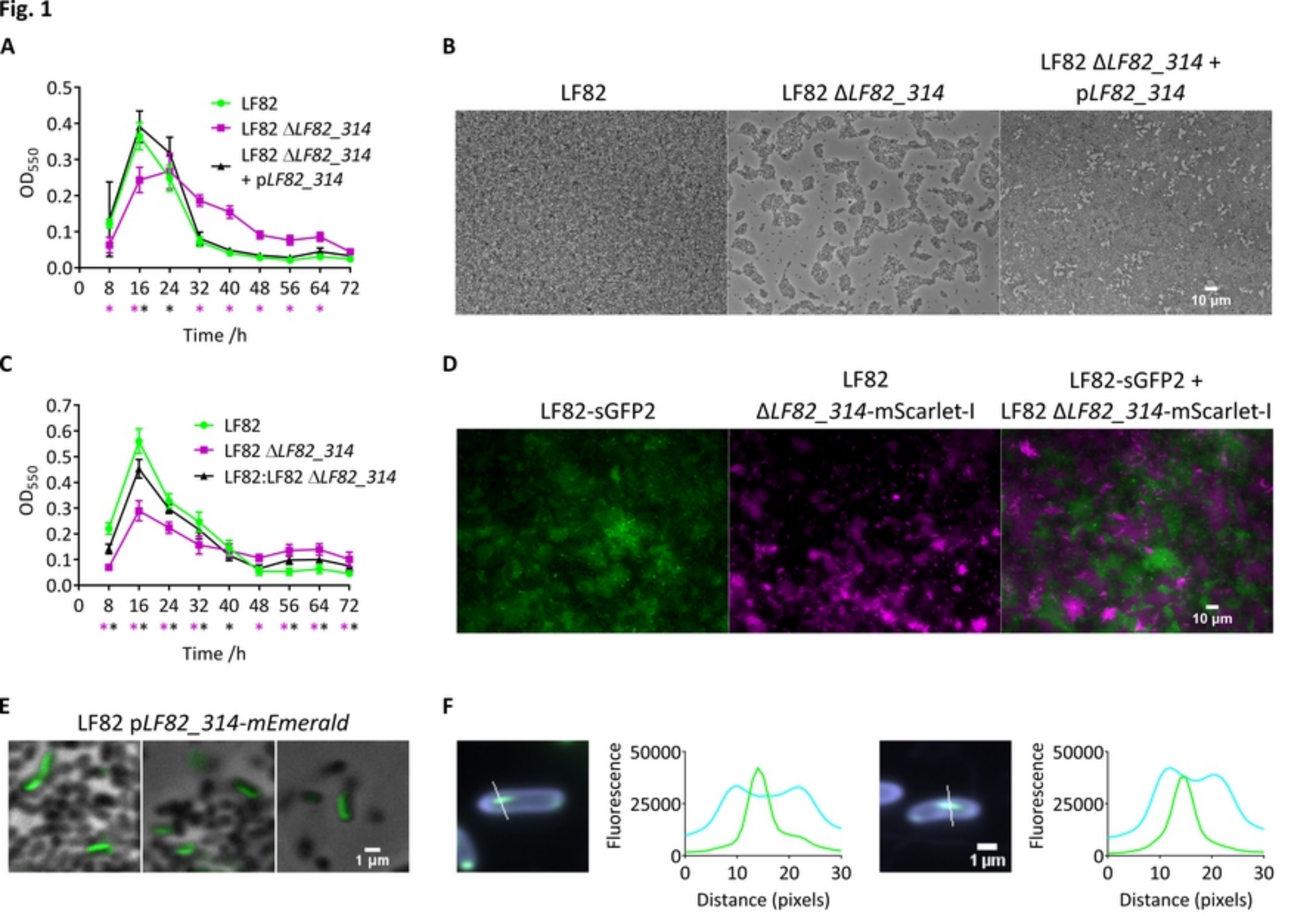

Figure 1

Fig. 2

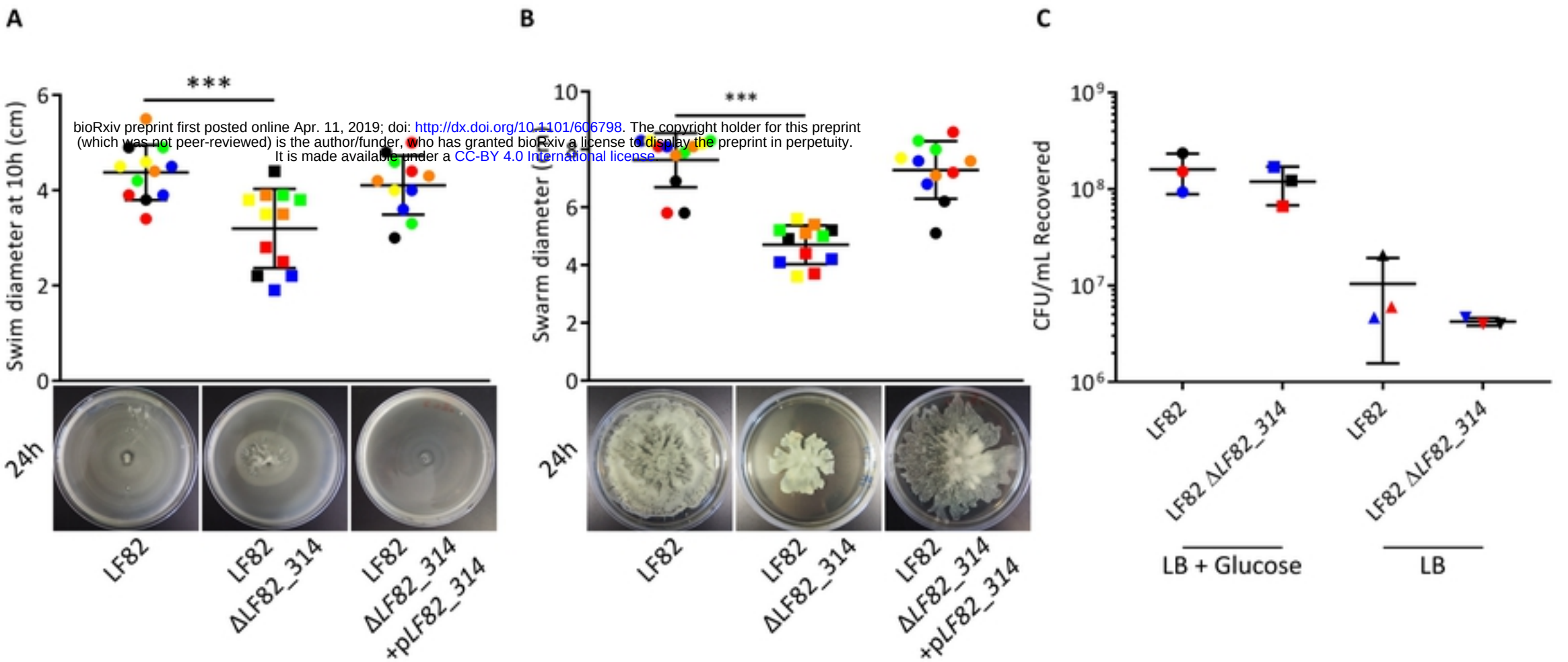

Figure 2

Fig. 3

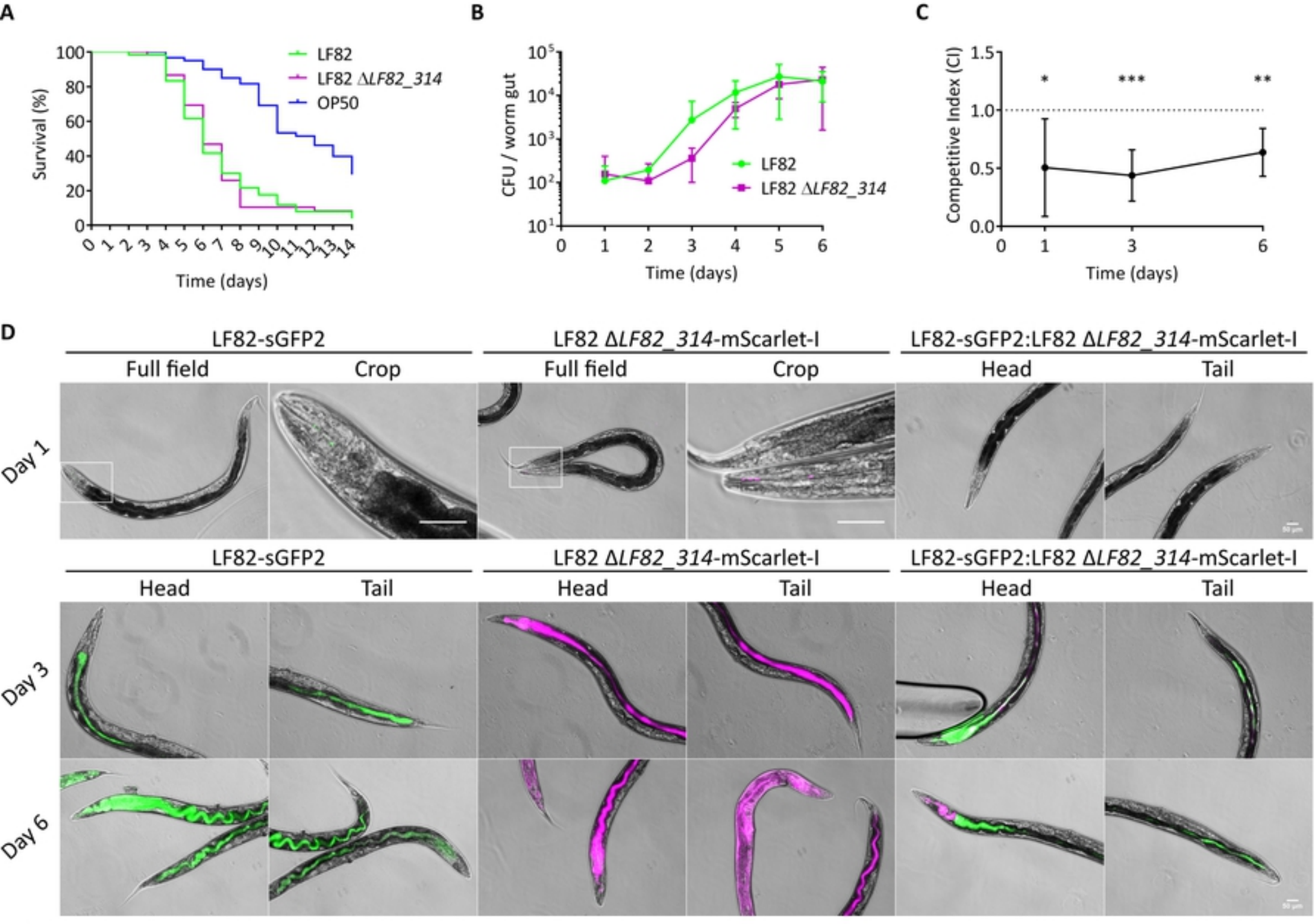

Figure 3

Fig. 4

A

B

Gene	Annotation	Gene	Annotation
asnV	Asparagine tRNA	*LF82_314*	Conserved hypothetical protein
LF82_309	Phage integrase domain-containing	*yhhI*	Transposase
LF82_310	Transposase	*LF82_315*	Hypothetical protein
LF82_311	Phage integrase domain-containing	*LF82_316*	Hypothetical protein
LF82_312	MazE family anti-toxin protein	*LF82_317*	ATP-dependent endonuclease
LF82_313	PemK-like growth inhibitor	*LF82_318*	Superfamily I DNA and RNA helicases
LF82_774	XRE family transcription factor		

C

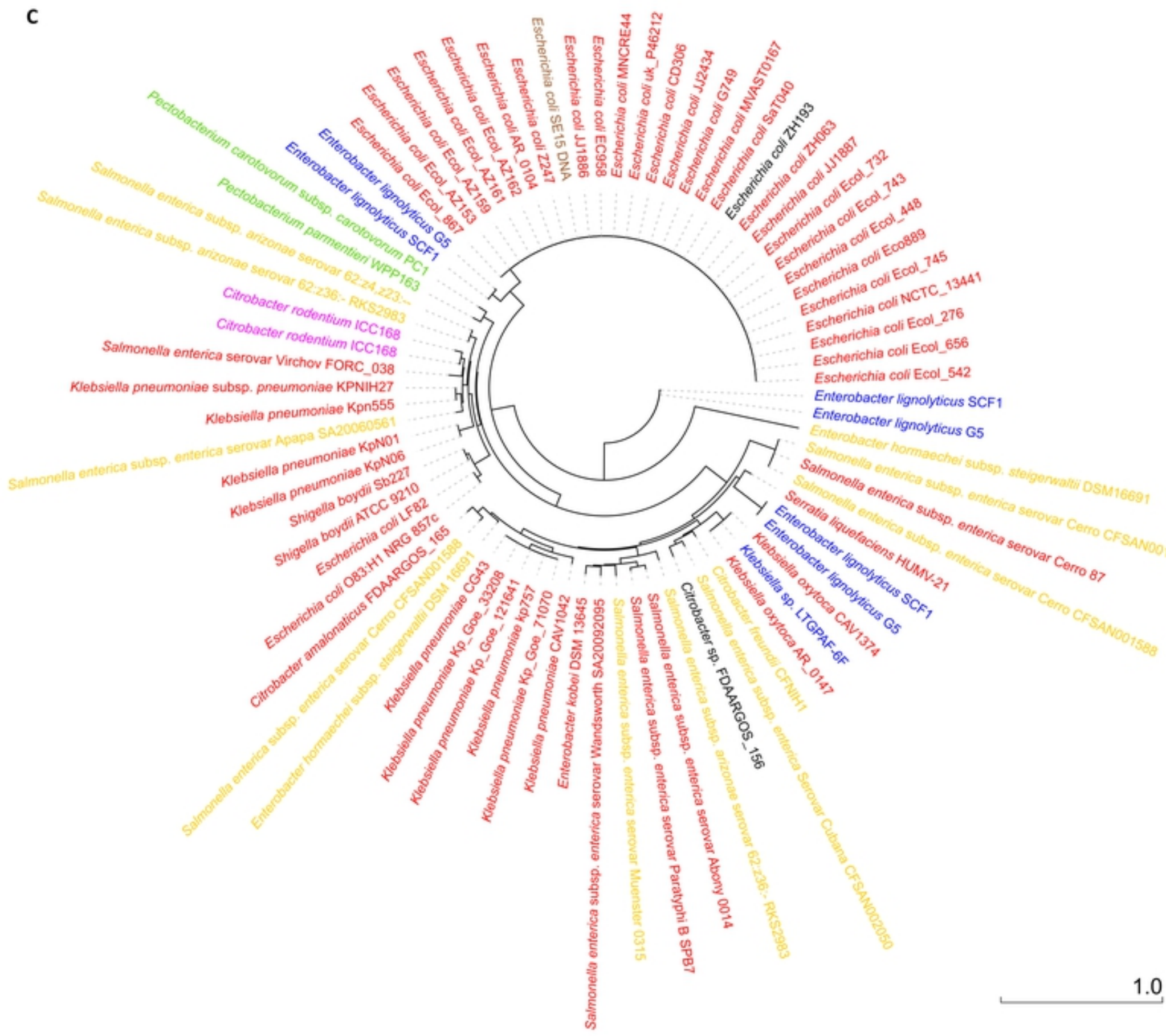

Figure 4