


Organization and architecture of AggR-dependent promoters from enteroaggregative *Escherichia coli*

Muhammad Yasir,^{1,2} Christopher Icke,¹
Radwa Abdelwahab,^{1,3} James R. Haycocks,¹
Rita E. Godfrey,¹ Pavelas Sazinas,⁴ Mark J. Pallen,²
Ian R. Henderson,¹ Stephen J. W. Busby^{1*} and
Douglas F. Browning ^{1*}

¹Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK.

²Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UA, UK.

³Faculty of Medicine, Assiut University, Assiut, Egypt.

⁴Department of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800, Kgs Lyngby, Denmark.

Summary

Enteroaggregative *Escherichia coli* (EAEC), is a diarrhoeagenic human pathogen commonly isolated from patients in both developing and industrialized countries. Pathogenic EAEC strains possess many virulence determinants, which are thought to be involved in causing disease, though, the exact mechanism by which EAEC causes diarrhoea is unclear. Typical EAEC strains possess the transcriptional regulator, AggR, which controls the expression of many virulence determinants, including the attachment adherence fimbriae (AAF) that are necessary for adherence to human gut epithelial cells. Here, using RNA-sequencing, we have investigated the AggR regulon from EAEC strain 042 and show that AggR regulates the transcription of genes on both the bacterial chromosome and the large virulence plasmid, pAA2. Due to the importance of fimbriae, we focused on the two AAF/II fimbrial gene clusters in EAEC 042 (*afaB-afaCB* and *aafDA*) and identified the promoter elements and AggR-binding sites required for fimbrial expression. In addition, we

examined the organization of the fimbrial operon promoters from other important EAEC strains to understand the rules of AggR-dependent activation. Finally, we generated a series of semi-synthetic promoters to define the minimal sequence required for AggR-mediated activation and show that the correct positioning of a single AggR-binding site is sufficient to confer AggR-dependence.

Introduction

Enteroaggregative *Escherichia coli* (EAEC) is an important human pathogen that is responsible for causing diarrhoea in both adults and children in industrialized and developing countries (Okeke *et al.*, 2000; Wilson *et al.*, 2001; Nataro *et al.*, 2006; Franca *et al.*, 2013). It has been shown to elicit travellers' diarrhoea, paediatric diarrhoea and persistent diarrhoea in HIV-infected patients, as well as extra-intestinal infections, such as urinary tract infections and septicaemia (Durrer *et al.*, 2000; Okeke *et al.*, 2000; Adachi *et al.*, 2001; Olesen *et al.*, 2012; Herzog *et al.*, 2014). EAEC strains have been linked to a number of serious diarrhoeal outbreaks, including the food-borne outbreak caused by a Shiga-toxin-producing EAEC O104:H7 in Germany in 2011, which infected over 4000 individuals and resulted in 54 deaths (Itoh *et al.*, 1997; Harada *et al.*, 2007; Frank *et al.*, 2011; Boisen *et al.*, 2015). In spite of its global importance as a human pathogen, the mechanisms by which EAEC causes disease are still poorly understood. In some instances, specific virulence determinants have been identified, but as EAEC strains are extremely heterogeneous in nature, many determinants are not present in all strains (Estrada-Garcia and Navarro-Garcia, 2012; Franca *et al.*, 2013).

EAEC pathogenesis is thought to proceed by the colonization of the human intestinal mucosa followed by the production of various toxins, such as plasmid-encoded toxin (Pet), the Pic mucinase, enteroaggregative heat-stable toxin (EAST-1) and *Shigella* enterotoxin 1 (ShET1), and the concurrent triggering of inflammation (Savarino *et al.*, 1991; Fasano *et al.*, 1997; Henderson *et al.*, 1999; Harrington *et al.*, 2009; Estrada-Garcia and Navarro-Garcia, 2012). Typical EAEC strains carry the

Accepted 22 November, 2018. *For correspondence. E-mail s.j.w.busby@bham.ac.uk; Tel. +44 (0)121-414-5439 (Stephen J. W. Busby); E-mail D.F.Browning@bham.ac.uk; Tel. +44 (0)121-414-5435 (Douglas F. Browning).

plasmid-encoded AggR transcription regulator protein, a member of the AraC-XylS family of transcription factors (Nataro *et al.*, 1994; Sarantuya *et al.*, 2004). AggR co-ordinately activates the expression of many genes thought to be required for pathogenesis, for example the attachment adherence fimbriae (AAF) required for colonization, the anti-aggregative protein dispersin (Aap) and its dedicated type I secretion system (T1SS) (Elias *et al.*, 1999; Sheikh *et al.*, 2002; Nishi *et al.*, 2003; Morin *et al.*, 2013). As AggR is central to activating the expression of essential virulence genes, it is key to understanding pathogenesis in this important human pathogen. Here, we use RNA-sequencing (RNA-seq) to examine the AggR regulon in the pathogenic EAEC strain 042 and show that AggR regulates genes on both the large virulence plasmid, pAA2, and the bacterial chromosome. As fimbrial biogenesis is central to EAEC pathogenesis, we examine the organization and architecture of AggR-dependent fimbrial promoters from EAEC strain 042 and from a number of other important EAEC strains, identifying the different promoter elements and functional AggR-binding sites required for expression.

Results

RNA-seq analysis of the AggR regulon in EAEC strain 042

AggR is the master regulator of EAEC virulence. Previously, Morin *et al.* (2013) examined the AggR regulon, using micro-arrays, for the archetypal pathogenic strain EAEC 042. As micro-array analysis can be influenced by probe design and genome annotation, and has issues with detecting low abundance transcripts (Zhao *et al.*, 2014), we repeated the analysis using high-throughput RNA-seq methodology. Briefly, wild-type EAEC 042 and an isogenic *aggR* mutant strain (EAEC 042 Δ *aggR*) (Table S1) were grown in triplicate until mid-logarithmic growth in high glucose Dulbecco's modified Eagle's medium (DMEM), which has been shown to induce biofilm formation and AggR-dependent gene expression in EAEC (Sheikh *et al.*, 2001; Morin *et al.*, 2013), and RNA was isolated and contaminating DNA removed. The isolated RNA was converted to cDNA and sequenced, generating over 7 million reads each with >90% of reads aligning to the EAEC 042 genome. Genes were considered to be differentially expressed if there was > 1 log₂-fold difference in expression accompanied by an adjusted *p*-value < 0.00001 between the mutant and the wild-type strains. In total, 112 genes were differentially expressed in EAEC 042 in comparison to the *aggR* mutant (Tables S2 and S3). These genes were located on both the chromosome and the large pAA2 plasmid (Fig. 1). Note that with the exception of EC042_pAA056,

all the AggR-regulated genes identified by Morin *et al.* (2013) were identified by our study (Tables S2 and S3).

Of the 112 genes that showed differential expression between the wild-type and the *aggR* mutant, 29 were located in clusters on the large virulence plasmid, pAA2 (Fig. 1B and Table S2). It is of note that these genes are all confined to one half of the plasmid, whilst the genes required for plasmid replication and conjugative transfer are located on the other half and are independent of AggR control (Fig. 1B) (Chaudhuri *et al.*, 2010). The expression of many of these genes has previously been shown to be dependent on AggR, for example *aggR* itself, *aar*, which encodes a repressor of AggR, the five genes encoding the Aat T1SS (*aatPABCD*) and its secreted substrate dispersin (*aap*), EC042_pAA003 and EC042_pAA004 that encode proteins that are involved in biofilm formation, the polysaccharide deacetylase encoded by *shf*, the *Shigella flexneri* virulence protein VirK and the AAF/II fimbriae (*aafDA* and *afaB-aafCB*) (Elias *et al.*, 1999; Nishi *et al.*, 2003; Fujiyama *et al.*, 2008; Chaudhuri *et al.*, 2010; Morin *et al.*, 2010; Morin *et al.*, 2013; Santiago *et al.*, 2014). Many of the genes, which are encoded on pAA2 and activated by AggR, have unknown function (*e.g.* EC042_pAA005, EC042_pAA005A, EC042_pAA019, EC042_pAA020 and EC042_pAA061) and, thus, their potential role in EAEC 042 pathogenicity is unclear.

From the genes differentially expressed in the *aggR* mutant, 83 were located on the chromosome and many of these genes are located in chromosomal islands (Fig. 1A and Table S3), for example the genes which encode the Aai type VI secretion system (T6SS) (EC042_4562 to EC042_4583 (*aaiA* to *aaiU*)). This cluster consists of the 16 genes encoding the T6SS machinery and 4 hypothetical proteins (EC042_4580, EC042_4581, EC042_4582 and EC042_4583) and has been shown to be activated by AggR (Dudley *et al.*, 2006; Morin *et al.*, 2013). A second AggR activated chromosomal island extends from EC042_3179A to EC042_3187 (Fig. 1A and Table S3). Previously, Morin *et al.* (2013) identified EC042_3182 and EC042_3184 as being AggR regulated. Whilst many genes within this region encode conserved proteins, with no homology to any known protein family (*e.g.* EC042_3179A, EC042_3180 and EC042_3184), EC042_3181 is of note as it is homologous to the transcription activator PerC from enteropathogenic *E. coli*, which regulates the *LEE1* pathogenicity island (Knutton *et al.*, 1997).

Interestingly, genes associated with flagellar motility were down regulated in the *aggR* mutant, whilst antigen 43 (Agn43) homologues (EC042_4803, *flu1*, and *flu2*) were up regulated, suggesting that AggR might regulate motility and cell aggregation in EAEC 042 (Table S3). To investigate this, the relative expression of three flagella genes (*fliA*, *flgB* and *fliC*) and EC042_4803 was

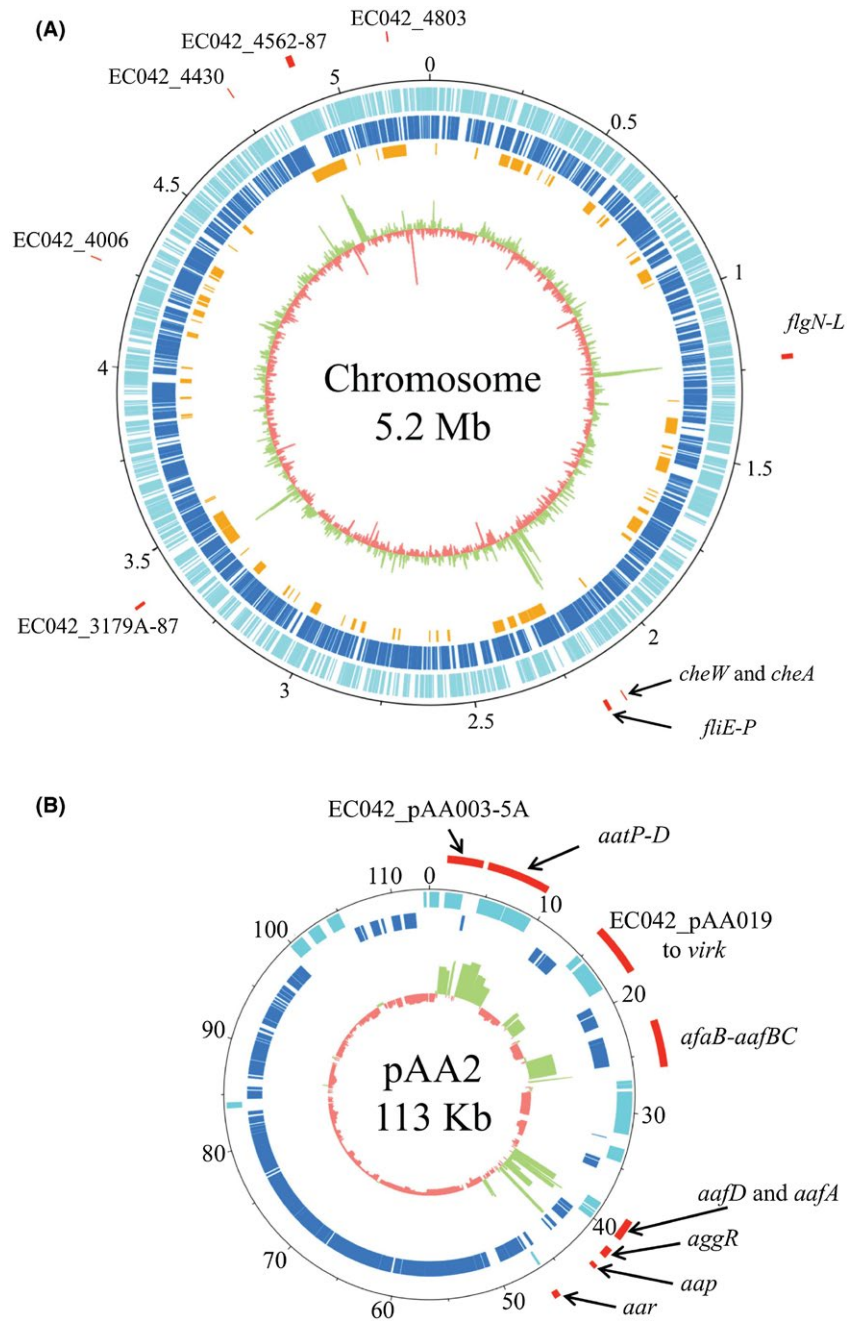


Fig. 1. AggR-regulated genes in EAEC strain 042. The figure shows the differential gene expression observed between wild-type EAEC 042 and its *aggR* mutant on A. the chromosome and B. plasmid pAA2, as determined by RNA-seq.

A. The data are displayed in rings from the outside inwards. The outermost red lines identify some of the differentially expressed genes (which are labelled with their gene name or number), followed by the base coordinates of the chromosome (labelled in Mb). The annotated genes of EAEC 042 are indicated in the forward and reverse orientation (light blue and dark blue respectively). The EAEC 042 chromosomal regions of difference (RODs) as identified by Chaudhuri *et al.* (2010) are presented in orange. The inner most circle shows the \log_2 fold difference for each gene compared between wild-type EAEC 042 and the *aggR* mutant. Positively differential expressed genes are presented in green and negatively differentially expressed genes are in red.

B. The rings depicting the data for plasmid pAA2 are the same as for the EAEC 042 chromosome in A. Note that base numbering for pAA2 is in Kb.

assessed by qRT-PCR. Results in Fig. S1 and Table S4 demonstrated that there was no statistically significant difference in the expression of *fliA*, *flgB*, *fliC* or

EC042_4803 between the wild-type EAEC 042 and the *aggR* mutant. Furthermore, there was no difference in their cell motility as observed on agar motility assay

plates (Fig. S2). As the expression of flagella genes is known to be stochastic (Spudich and Koshland, 1976; Korobkova *et al.*, 2004) and as *agn43* homologues are phase variable (Henderson *et al.*, 1997), we propose that the differential expression observed in our RNA-seq experiment for these genes is likely due to stochastic variation and phase variation, respectively, rather than direct regulation by AggR.

In order to confirm a direct role of AggR in the transcription of genes that showed some of the largest differential expression in our RNA-seq experiment (*i.e.* *aafD*, *afaB*, *aap*, *aatP* and *aaiA*), and to identify the AggR-dependent promoters that control their expression, ~400 bp of upstream DNA was amplified by PCR to generate the *aafD*100, *afaB*100, *aap*100, *aatP*100 and *aaiA*100 promoter fragments (Table S1). Each fragment was cloned into the low copy number *lac* expression vector, pRW50, to generate *lacZ* transcriptional fusions (Table S1) and pRW50 constructs were transferred into the Δ *lac* *E. coli* K-12 strain, BW25113. To investigate the role of AggR, cells also carried either plasmid pBAD/*aggR*, which encodes AggR expressed from an arabinose-inducible promoter, or empty pBAD24 vector as a control (Table S1) (Sheikh *et al.*, 2002). Transformants were grown with shaking in LB medium to mid-logarithmic phase, either with or without AggR induction by arabinose, and measured β -galactosidase activities were taken as a proxy for promoter activity. Results detailed in Fig. S3 show that for host cells carrying pRW50, containing each of the upstream regulatory region fragments, measured β -galactosidase levels are higher than levels with empty pRW50, showing that promoter activity is associated with each fragment. Furthermore, expression was markedly increased by arabinose in the presence of pBAD/*aggR*, but not increased with pBAD24. Thus, we conclude that each tested fragment carries an AggR-dependent promoter, corroborating the results of our RNA-seq analysis for these promoters.

Analysis of the AAF/II fimbrial operon promoters from EAEC 042

During infection, EAEC cells bind to human epithelial cells, using their AAF fimbriae (Harrington *et al.*, 2006). Due to the importance of fimbriae in EAEC pathogenesis and the role of AggR in their expression, we sought to characterize in detail the promoters that control the expression of the EAEC 042 AAF/II fimbrial genes. The AAF/II fimbrial genes are organized into two clusters (*aafDA* and *afaB-aafCB*) on the pAA2 virulence plasmid (Fig. 1B) (Elias *et al.*, 1999). Figs 2A and 3A detail the DNA sequence of the *aafD*100 and *afaB*100

promoter fragments, which carry DNA upstream of *aafD* and the *afaB* pseudogene respectively. Note that each fragment is flanked by EcoRI and HindIII sites, which were introduced to aid cloning, and sequences are numbered from the HindIII site. Inspection of both sequences identified several matches to the proposed AggR-binding site consensus (Morin *et al.*, 2010) (Figs 2A and 3A). Therefore, to identify the essential sequences for AggR-induced promoter activity, we initially focused on the *aafD*100 promoter fragment and constructed nested deletions from the EcoRI end of the fragment. Each shortened fragment (*i.e.* *aafD*99, *aafD*98, *aafD*97, *aafD*96, *aafD*95 and *aafD*94) (Fig. 2A, Table S1) was cloned into pRW50 and each plasmid construct was transferred into BW25113 cells, carrying pBAD/*aggR* or pBAD24. β -galactosidase activity was measured, as before, and results in Fig. 2B show that *aafD*96 is the shortest fragment where full AggR-dependent induction is retained, with induction being greatly reduced for *aafD*95 and absent for *aafD*94. To identify the transcript start, we extracted RNA from BW25113 cells carrying pRW50/*aafD*96 and containing either pBAD/*aggR* or pBAD24. Fig. 2C shows the result of the primer extension analysis, analysed by polyacrylamide gel electrophoresis, and identifies two clear bands, corresponding to transcripts starting at positions 56 and 54 of the cloned sequence (Fig. 2A). Note that these bands are seen in the sample from cells carrying pBAD/*aggR* but not with pBAD24. Examination of the *aafD*96 DNA sequence upstream of positions 56 and 54 revealed a potential -10 hexamer element (5'-TAGCAT-3') and a potential AggR-binding site (5'-GTTTATTTATC-3'), based on previously established consensus sequences (Morin *et al.*, 2010; Browning and Busby, 2016) (Fig. 2A). Therefore, to investigate the role of these sequences, site-directed mutagenesis was used to introduce the 65C and 92C/90C substitutions into the *aafD*96 fragment, to disrupt each element (Fig. 2A). Mutant derivatives were cloned into pRW50, transferred into BW25113 cells, carrying pBAD/*aggR* or pBAD24, and the promoter activity determined. Results in Fig. 2D show that AggR-dependent promoter activity from the *aafD*96 fragment was greatly decreased by these substitutions, consistent with our proposal of these elements as the -10 hexamer and AggR-binding site at the *aafD* promoter.

To locate the essential promoter sequences required for *afaB-aafCB* expression, we also constructed nested deletions of the *afaB*100 promoter fragment (Fig. 3A). Again, each of the shorter fragments (*afaB*99, *afaB*98 and *afaB*97) (Fig. 3A, Table S1) was cloned into pRW50 and promoter activity determined. Results in Fig. 3B show that *afaB*100 is the only fragment where AggR-dependent induction is observed. Thus, to identify the

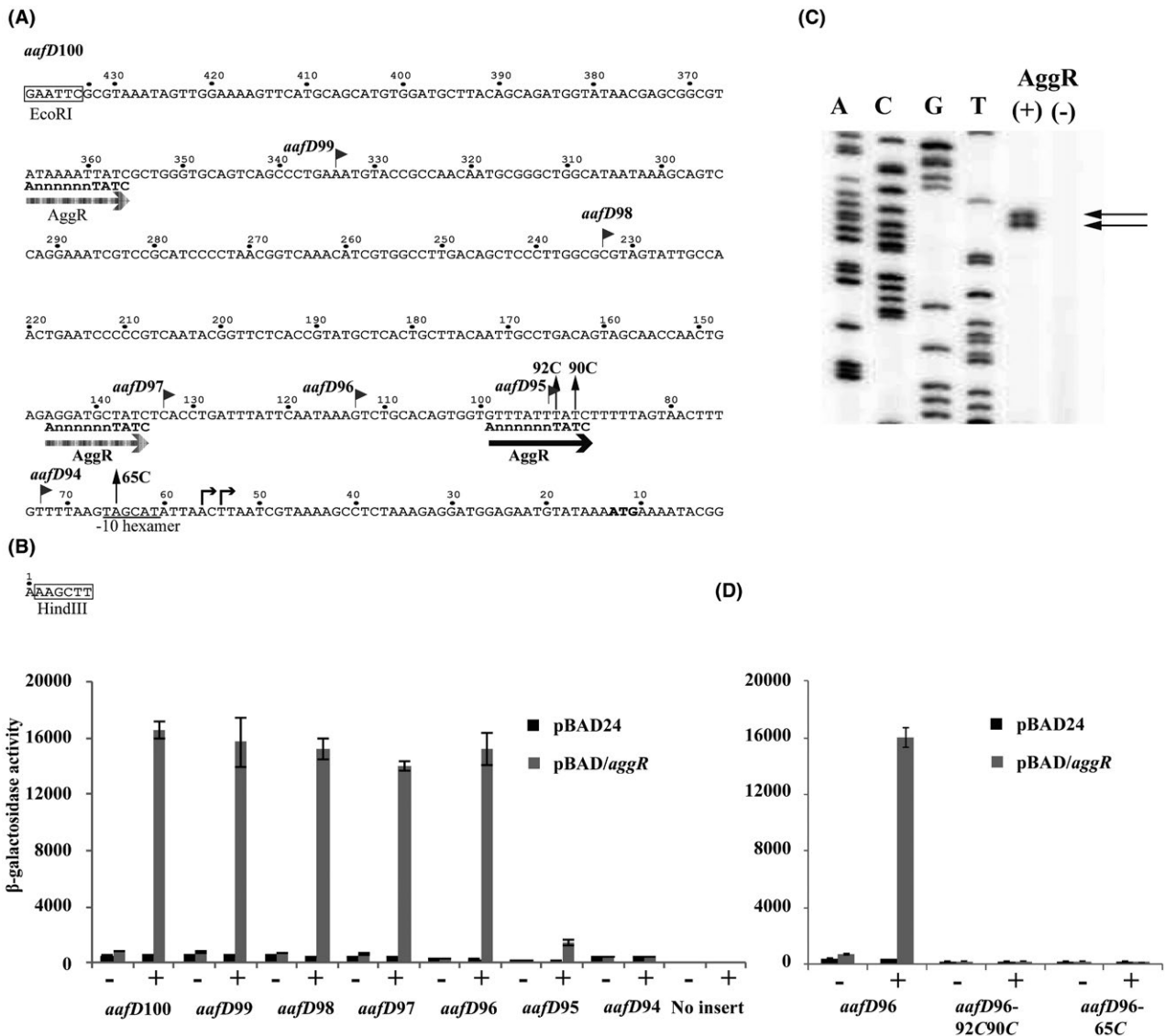


Fig. 2. Analysis of the *aafD100* promoter fragment from EAEC strain 042.

A. The panel shows the base sequence of the EAEC 042 *aafD100* regulatory region fragment, which includes the start of the *aafD* coding sequence. The sequence is flanked by upstream EcoRI and downstream HindIII sites and is numbered from the base immediately upstream of the HindIII site. The limits of the *aafD99*, *aafD98*, *aafD97*, *aafD96*, *aafD95* and *aafD94* nested deletions are indicated by flags. The proposed promoter -10 hexamer element is underlined, the experimentally determined transcript start sites are indicated by bent horizontal arrows and the initiating ATG codon is in bold. Potential AggR-binding sites are indicated by horizontal arrows, with functional and non-functional sites denoted by dark and light shading respectively. Each site is aligned with the AggR-binding consensus (Morin *et al.*, 2010). The locations of the 65C and 92C/90C substitutions, which disrupt the -10 element and the functional AggR-binding site, respectively, are shown.

B. The panel illustrates measured β -galactosidase activities in *E. coli* K-12 BW25113 Δ *lac* cells, containing pRW50 carrying the *aafD100* fragment, shortened derivatives or no insert. Cells also carried either pBAD/*aggR* (grey bars) or pBAD24 (black bars).

C. The panel shows an autoradiogram of a denaturing polyacrylamide gel run to determine the primer extension products from RNA synthesis initiating at the *aafD* promoter in BW25113 cells carrying pRW50/*aafD96*. AggR (+) and AggR (-) indicates cells carried pBAD/*aggR* or pBAD24. Reactions are calibrated with the M13mp18 phage reference sequence (A, C, G and T), which serves as sequence ladder. Primer extension products, produced in the presence of AggR, are indicated by arrows.

D. The panel shows the β -galactosidase activities of BW25113 cells, containing pRW50 carrying either the *aafD96* fragment or mutant derivatives. Cells also carried either pBAD/*aggR* (grey bars) or pBAD24 (black bars). In panels B. and D. cells were grown in LB medium in presence (+) or absence (-) of 0.2% arabinose. β -galactosidase activities are expressed as nmol of ONPG hydrolysed min⁻¹ mg⁻¹ dry cell mass. Each activity is the average of three independent determinations and standard deviations are shown for all data points.

start of transcription, we again extracted RNA from BW25113 cells carrying pRW50/*afaB100* with either pBAD/*aggR* or pBAD24. Fig. 3C shows the result

of the primer extension analysis and identifies two bands, corresponding to transcripts starting at positions 280 and 266 of the cloned sequence (Fig. 3A),

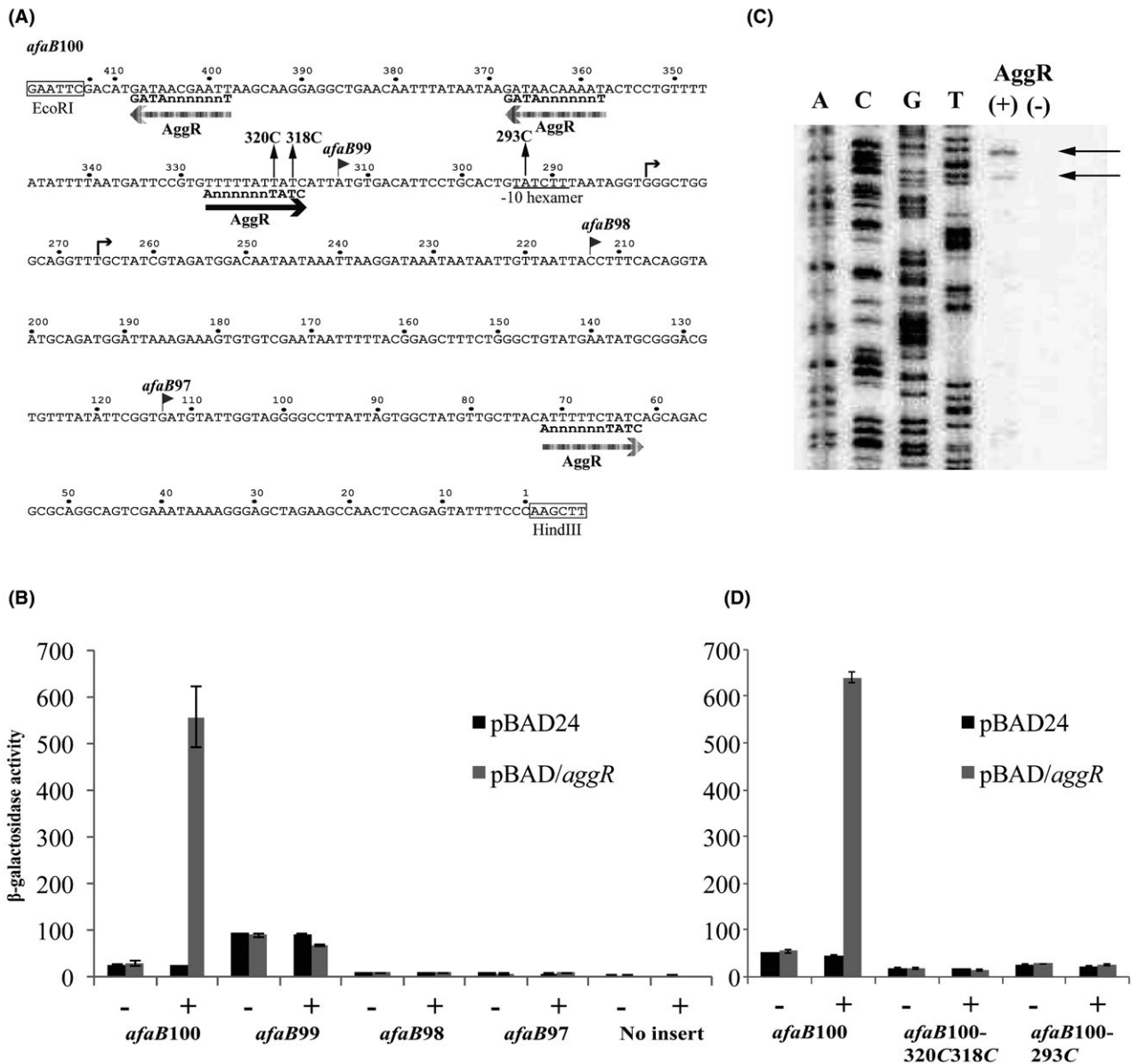


Fig. 3. Analysis of *afaB100* promoter fragment from EAEC strain 042.

A. The panel shows the base sequence of the EAEC 042 *afaB100* regulatory region fragment flanked by upstream EcoRI and downstream HindIII sites. The sequence is numbered from the base immediately upstream of the HindIII site. The limits of the *afaB99*, *afaB98* and *afaB97* nested deletions are indicated by flags. The proposed -10 hexamer element is underlined and the experimentally determined transcript start sites are indicated by bent horizontal arrows. Potential AggR-binding sites are indicated by horizontal arrows, with functional and non-functional sites denoted by dark and light shading respectively. Each site is aligned with the AggR-binding consensus (Morin *et al.*, 2010). The location of the 293C and 320C/318C substitutions, which disrupt the -10 element and the functional AggR-binding site, respectively, is shown.

B. The panel illustrates measured β -galactosidase activities in *E. coli* K-12 BW25113 cells containing pRW50, carrying the *afaB100* fragment, shortened derivatives, or no insert. Cells also carried either pBAD/*aggR* (grey bars) or pBAD24 (black bars).

C. The panel shows an autoradiogram of a denaturing polyacrylamide gel run to determine the primer extension products from RNA initiating from the *afaB* promoter in BW25113 cells, carrying pRW50/*afaB100*. AggR (+) and AggR (-) indicates cells carried pBAD/*aggR* or pBAD24. Reactions are calibrated with the M13mp18 phage reference sequence (A, C, G and T), which serves as sequence ladder. Primer extension products, produced in the presence of AggR, are indicated by arrows.

D. The panel shows the β -galactosidase activities in BW25113 cells containing pRW50 carrying either the *afaB100* fragment or mutant derivatives. Cells also carried either pBAD/*aggR* (grey bars) or pBAD24 (black bars). In panels B. and D. cells were grown in LB medium in presence (+) or absence (-) of 0.2% arabinose. β -galactosidase activities are expressed as nmol of ONPG hydrolysed $\text{min}^{-1} \text{mg}^{-1}$ dry cell mass. Each activity is the average of three independent determinations and standard deviations are shown for all data points.

which were only present in the sample from cells carrying pBAD/*aggR*. Examination of the *afaB*100 DNA sequence upstream of these positions revealed a potential –10 element (5'-TATCTT-3') and AggR-binding site (5'-TTTTTATTATC-3') (Fig. 3A). These elements were, therefore, disrupted by introducing the 293C and 320C/318C substitutions into the promoter region (Fig. 3A) and mutant *afaB*100 fragments were cloned into pRW50. As expected, AggR-dependent promoter activity was substantially decreased by these substitutions (Fig. 3D), consistent with our hypothesis that these sequences constitute a functional –10 element and AggR-binding site at the *afaB* promoter. Previously, Elias *et al.* (1999) predicted that the promoter controlling *aafCB* expression was immediately upstream of *aafC*. Therefore, using our dual reporter system, we checked for promoter activity in different *afaB*-*aafCB* fragments. However, we could not find any evidence of a second promoter (Fig. S4). Thus, our results with both *aafD* and *afaB* promoter fragments indicate that each EAEC 042 AAF/II fimbrial gene cluster is expressed from a single upstream AggR-dependent promoter.

Other EAEC fimbrial operon promoters possess similar promoter organization

To date, five AAF systems (AAF/I to AAF/V) have been identified in EAEC strains and the genes that encode these fimbrial components, together with the corresponding chaperones and ushers, are all found on large virulence plasmids (Savarino *et al.*, 1994; Elias *et al.*, 1999; Bernier *et al.*, 2002; Boisen *et al.*, 2008; Jonsson *et al.*, 2015). As the promoters that control the expression of different AAF variants have not been characterized, we investigated some of these promoters in more detail to uncover their promoter organization and determine whether AggR regulates them similarly. It has been shown that EAEC strain 17-2 and the highly virulent Shiga-toxin-producing EAEC O104:H4 strain C227-11, produce AAF/I fimbriae, and the fimbrial genes are organized in a single operon (*aggDCBA*) (Savarino *et al.*, 1994; Rasko *et al.*, 2011; Rohde *et al.*, 2011). Therefore, to identify the fimbrial operon promoter from EAEC 17-2, PCR was used to amplify the DNA upstream of *aggD* to generate the *aggD*100 promoter fragment (Fig. 4A). This was cloned into pRW50 and assayed for promoter activity in BW25113 cells, carrying either pBAD24 or pBAD/*aggR*, as before. Results detailed in Fig. 4B show that expression from *aggD*100 fragment was greatly increased by AggR induction, confirming that the EAEC 17-2 *aggD* promoter is AggR regulated. To pinpoint the location of important regulatory sequences, nested deletions were

constructed and the shortened fragments (*i.e.* *aggD*99, *aggD*98 and *aggD*97) were cloned into pRW50 and assayed (Fig. 4). Results in Fig. 4B show that AggR-mediated induction is absent with the *aggD*97 fragment, and that *aggD*98 is the shortest of the fragments where AggR-dependent promoter activity is observed. Examination of the *aggD*98 DNA sequence revealed a potential promoter –10 element (5'-TATAAT-3') and an AggR-binding site (5'-ATTTTTTTAGC-3') (Fig. 4A). Disruption of these elements in the *aggD*98 fragment, by introducing the 60C and 86C substitutions, respectively, greatly decreased promoter expression (Fig. 4C), supporting our proposal that these are the functional –10 element and AggR-binding site at this promoter.

Savarino *et al.* (1994) noted that the EAEC 17-2 AAF/I *aggD* promoter carried six direct repeats of the hexamer 5'-TCAAGT-3', which are positioned between the –10 element and the *aggD* translation initiation codon (Fig. S5). Interestingly, these repeats are more extensive in the *aggD* promoters from other pathogenic EAEC strains, *e.g.* the EAEC O104:H4 strain C227-11 possesses 15 repeats (Table S5 and Fig. S5). As such tandem repeats are unusual in bacteria and can play a role in gene expression (Browning and Busby, 2016), we examined if the different number of repeats carried by the EAEC 17-2 and C227-11 *aggD* promoters affected promoter activity. Results detailed in Fig. S5 show that the two promoters had similar promoter activity and, thus, although these repeat tracts are substantial, they do not appear to influence *aggD* promoter activity.

Alignment of the nucleotide sequence of the AAF/I and AAF/II fimbrial operon promoters (Fig. 5A) indicated that in each case, the DNA-binding site for AggR is located 21 or 22 bp upstream from the –10 element. This suggests that all AggR-dependent fimbrial promoters may have similar promoter organization. Using this information, we examined the DNA upstream of the *agg3D* and *agg4D* genes, which are the first genes in the AAF/III and AAF/IV fimbrial operons from the pathogenic EAEC strains 55989 and C1010-00, respectively, and identified suitably positioned AggR-binding sites and –10 elements (Bernier *et al.*, 2002; Boisen *et al.*, 2008) (Figs 5A and S6). To investigate the regulation of these AAF variants, the DNA upstream of *agg3D* and *agg4D* was cloned into plasmid pRW50, to generate the *agg3D*100 and *agg4D*100 promoter fragments, and point mutations were introduced to disrupt the potential AggR-binding sites and –10 elements identified (Figs 5A and S6). The β -galactosidase activity of BW25113 cells, carrying these constructs, was then measured, as before. Results in Fig. 5B and 5C indicated that expression from both the wild-type *agg3D*100 and *agg4D*100

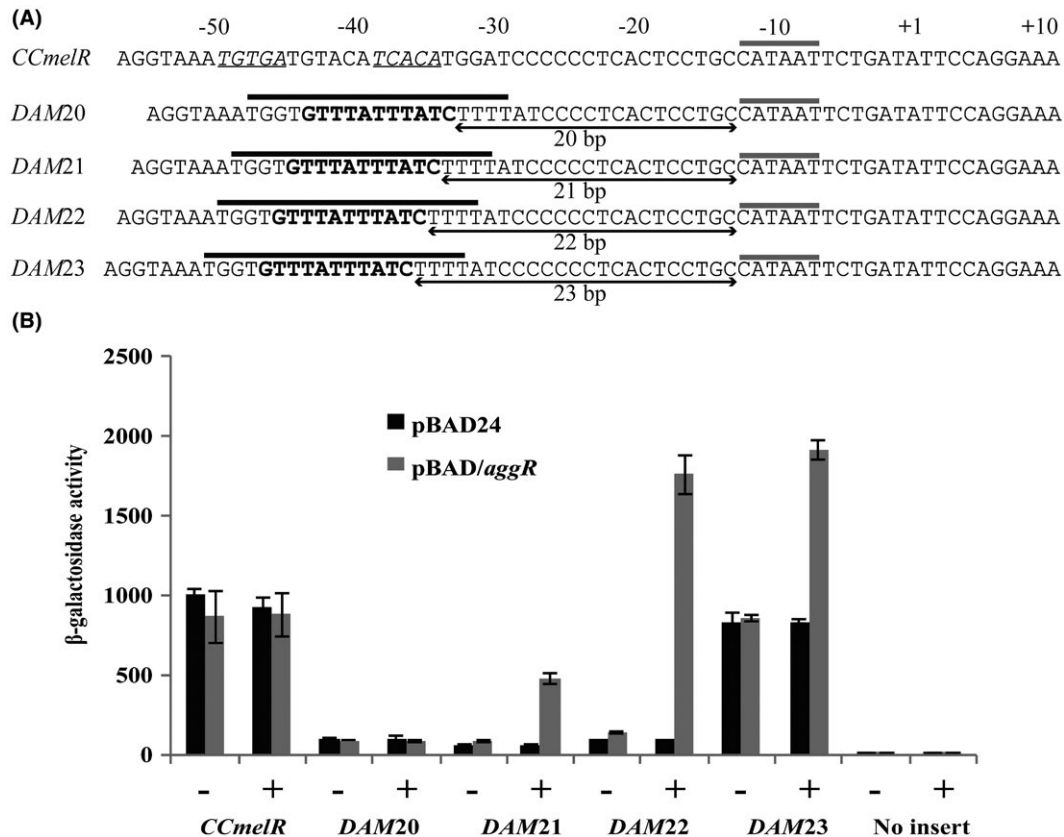


Fig. 6. Construction and analysis of semi-synthetic AggR-dependent promoters.

A. The panel illustrates the DNA sequence of the *CCmelR* promoter region and the *DAM20*, *DAM21*, *DAM22* and *DAM23* promoter constructs. In these promoters, the AggR-binding site from the *aafD* promoter has been transplanted at different distances from the -10 elements (20 bp to 23 bp). The CRP-binding half-sites in the *CCmelR* promoter are italicized and underlined. Thick black lines indicate the *aafD* promoter sequence transplanted, with the AggR-binding site in bold, and the -10 elements are indicated by grey lines. Sequence is numbered from the *CCmelR* promoter transcript start site (+1).

B. The panel illustrates the measured β -galactosidase activities in BW25113 cells, containing pRW50 carrying the *CCmelR* and various *DAM* promoter derivatives. Cells also carried either pBAD/*aggR* (grey bars) or pBAD24 (black bars). Cells were grown in LB medium in presence (+) or absence (-) of 0.2% arabinose. β -galactosidase activities are expressed as nmol of ONPG hydrolysed $\text{min}^{-1} \text{mg}^{-1}$ dry cell mass. Each activity is the average of three independent determinations and standard deviations are shown for all data points.

Discussion

Using an RNA-seq approach, we identified the genes regulated by AggR in the archetypal pathogenic EAEC strain 042. Some members of the AggR regulon are conserved hypothetical genes with unknown function and their regulation by AggR suggests a possible role in EAEC 042 intestinal colonization. As EAEC strains are heterogeneous in nature, further investigation of these candidate genes has the potential to enhance our knowledge of EAEC pathogenicity.

The main aim of this study was to determine the organization and architecture of AggR-dependent promoters. Focusing on the fimbrial operon promoters in EAEC strain 042, we found single AggR-dependent promoters upstream of the *aafDA* and *afaB-aafCB* regions on the pAA2 virulence plasmid that encode genes for fimbrial

assembly. The *aafD* promoter is located immediately upstream of the *aafD* gene, which encodes the AAF/II chaperone protein, whilst the *afaB* promoter is upstream of the *afaB* pseudogene, which is followed by functional *aafC* and *aafB* genes, encoding the fimbrial usher protein and the fimbrial adhesin respectively (Figs 1B, 2A, 3A and S4). It is of note that the level of expression from the *aafD* promoter is considerably higher than that of *afaB* (Figs 2 and 3) with the fold increase in RNA sequence reads for *aafDA* genes being higher than that of the *aafCB* genes (Table S2). As the *aafD* promoter controls the expression of the AafD chaperone and AafA fimbrial subunit, both of which are required in large amounts, in comparison to the AafC usher protein and AafB adhesin, this regulation is likely to help ensure that each component of the AAF/II fimbriae are made to the appropriate level. For the other AAF systems examined

(e.g. AAF/I, AAF/III and AFF/IV) the fimbrial genes exist in a single operon and were expressed from a strong upstream AggR-regulated promoter that had similar organization to the EAEC 042 fimbrial promoters (Fig. 5A).

Previous studies have shown that DNA sites for AggR-binding resemble sites for the Rns ‘master’ regulator from enterotoxigenic *E. coli* (ETEC) (Munson, 2013) and, following studies of the *aggR* promoter, a consensus sequence was also suggested for AggR, in which the importance of a TATC motif and an A base, seven nucleotides upstream of this motif, was highlighted (Morin *et al.*, 2010). Thus, we first identified putative DNA sites for AggR using this consensus. Based on the promoters characterized here, we now propose a revised consensus logo for the AggR-binding site and AggR-dependent promoters (Fig. 7). Note that, in our AggR-binding site consensus the upstream A base, noted by Morin *et al.* (2010), is not always conserved (Figs 5A and 7A). Indeed, using the *aafD* promoter from EAEC 042, which has a G at this position, we observed that any base can be tolerated at this position, with only a small effect on AggR-dependent

activation (Fig. S7). For each fimbrial promoter, we found a single functional DNA site for AggR, located 21 to 22 base pairs upstream from the promoter –10 element and this juxtaposition suggests that bound AggR must overlap the –35 element and is able to interact directly with Domain 4 of the RNA polymerase σ subunit. This is consistent with AggR being a member of the AraC-XylS family of bacterial transcription factors, many of which activate transcription initiation by making such a direct contact that serves to assist the recruitment of RNA polymerase to the target promoter (Martin and Rosner, 2001; Egan, 2002; Browning and Busby, 2004). It is also evident from our promoter logo that the DNA between AggR-binding site and the –10 element contains phased A/T tracts (Fig. 7B), which is indicative of bent DNA. Indeed, modelling of the *aggD*, *aafD* and *afaB* promoters suggests that AggR-dependent promoters possess a bent promoter architecture (Fig. S8). Our experiment, where a single DNA site for AggR was ‘transplanted’ into the context of the *E. coli melR* promoter, indicates that it is easy for AggR-dependence to be conferred onto a target promoter (Fig. 6). Since AggR-binding sites are relatively

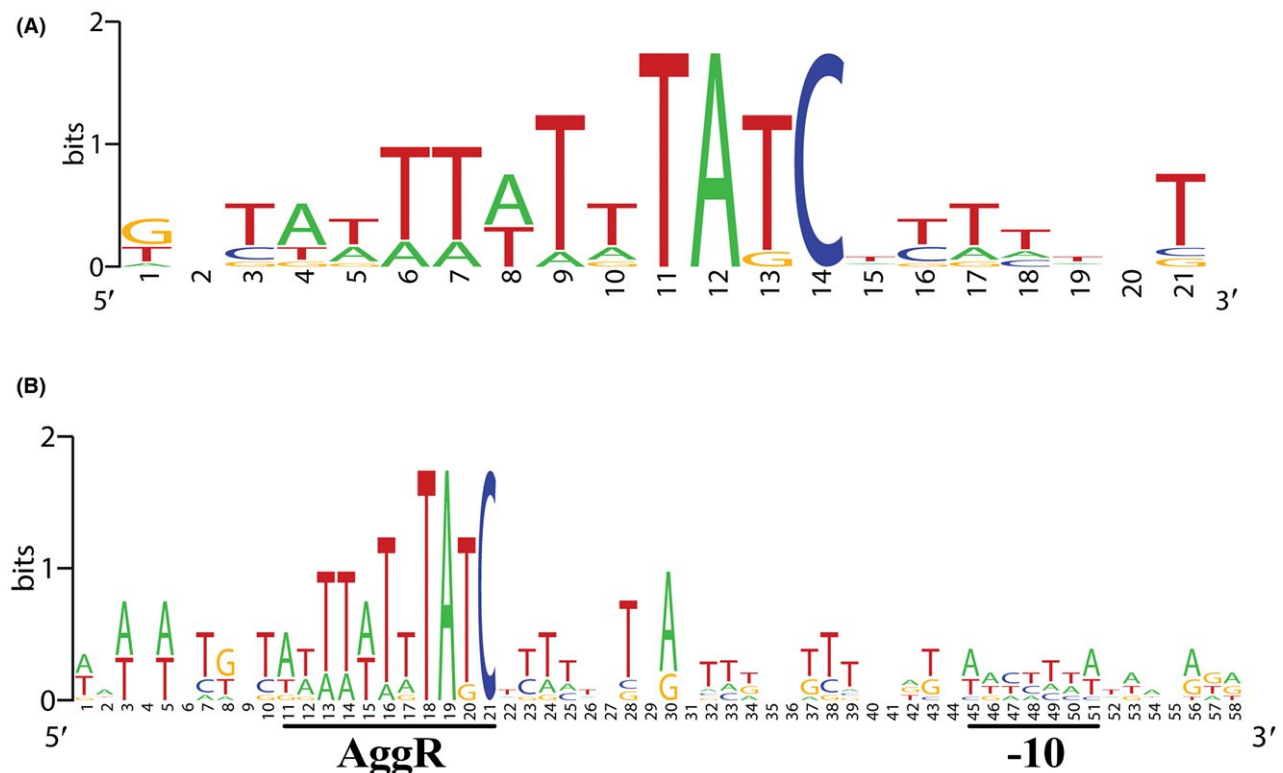


Fig. 7. The AggR-binding site consensus. The figure shows motifs for:

A. the AggR-binding site consensus sequence and

B. AggR-dependent promoter organization. Motifs were generated using the WebLogo server (Crooks *et al.*, 2004) with sequences from the EAEC 042 *aafD* and *afaB* promoters, the EAEC 17-2 *aggD* promoter, the EAEC 55989 *agg3D* promoter and the EAEC C1010-00 *agg4D* promoter identified by experiments in Figs. 2–5, the *aap*, *aatP* and *aaiA* promoters identified by similar experiments by Yasir (2017) and the AggR-binding site at the *aggR* promoter (Morin *et al.*, 2010).

degenerate, essentially consisting of a conserved TATC motif with an upstream A/T tract, it may be simple for the promoters expressing A/T-rich horizontally acquired genes, to become AggR-dependent and assimilated into the AggR regulon.

As our transcriptomics data identified a number of genes that had not previously been included in the AggR regulon (Fig. 1, Tables S2 and S3) (Morin *et al.*, 2013) we used the information from our AggR-dependent promoter logo (Fig. 7B) to interrogate the genes identified in our RNA-seq data set. Thus, we were able to find AggR-binding sites and suitably positioned -10 promoter elements (with a spacing of 21 to 23 bp) upstream of many leading genes in the transcription units that we found to be AggR regulated on both pAA2 and the chromosome (Tables 1 and 2, respectively). This analysis confirmed the organization of AggR-dependent promoters characterized by this study (*i.e.* *aggR*, *aatP*, *aap* and *aaiA*) and it is of note that these promoter sequences are conserved in other pathogenic EAEC strains, *e.g.* C227-11 and 55989, suggesting that these genes are similarly regulated in these strains (Fig. S9).

AggR-dependent biofilm formation is a hallmark of EAEC infection and, in addition to the expression of AAF fimbriae, other plasmid-encoded genes are required (*e.g.* EC042_pAA003, EC042_pAA004 and *shf*) (Czeczulin *et al.*, 1997; Fujiyama *et al.*, 2008; Morin *et al.*, 2013). Our analysis indicates chromosomally encoded genes, EC042_4006 (*yicS*) and *bssS*, are also AggR regulated (Tables 2 and S3). Both genes have been implicated in biofilm formation in *E. coli*, whilst YicS plays a role in pathogenicity in avian pathogenic *E. coli* (Domka *et al.*, 2006; Verma *et al.*, 2018). Thus, it is likely that AggR deploys both specialized plasmid- and chromosomally-encoded factors to ensure formation of its trademark biofilm.

Strikingly, AggR appears to control the expression of a number of transposases and transposon remnants (Tables S2 and S3) and putative AggR-dependent promoters are located upstream of these transcription units (Tables 1 and 2). This suggests that genomic rearrangements, on both the chromosome and the pAA2 plasmid, may occur more frequently in EAEC 042 during the initiation of the AggR virulence programme and lead to genome evolution, something which has been observed in other bacterial species (Lindsay, 2014; Singh *et al.*, 2014; Wan *et al.*, 2017).

AggR belongs to a subgroup of AraC-XylS family members, which control virulence gene regulation, and includes Rns/CfaD/CfaR from ETEC and VirF from *Shigella flexneri*. These family members are highly similar and often interchangeable, for example Rns can replace

VirF in *S. flexneri* and CfaR can complement for the loss of AggR in EAEC (Caron and Scott, 1990; Nataro *et al.*, 1994; Porter *et al.*, 1998). However, it is worth noting that this arrangement is not always reciprocal, as VirF is unable to replace Rns in ETEC, and this might reflect subtle differences in the mechanisms by which each regulator activates transcription (Porter *et al.*, 1998). As well as directly activating transcription, both Rns and VirF have been shown to activate at promoters by counteracting the repressive effects of the heat-stable nucleoid structuring protein, H-NS, which silences many horizontally acquired genes (Jordi *et al.*, 1992; Tobe *et al.*, 1993; Murphree *et al.*, 1997; Singh *et al.*, 2016). Experiments, which examined AggR-dependent activation at the *afaB* and *aafD* promoters in an *hns* null strain (Fig. S10), indicated that, although H-NS marginally represses both promoters, AggR still substantially activates transcription in the absence of H-NS. Consistent with this, a recent transcriptomic analysis in EAEC 042 indicated that neither *afaB* nor *aafD* were derepressed by the absence of H-NS or its homologue H-NS2 (EC042_2824) (Prieto *et al.*, 2018). Therefore, we propose that, at the *afaB* and *aafD* promoters, AggR primarily activates transcription by directly interacting with RNA polymerase rather than alleviating H-NS repression.

To characterize AggR-dependent promoters, we, as have others, used a simple two-plasmid system with a laboratory strain of *E. coli* K-12 as host (Dudley *et al.*, 2006; Morin *et al.*, 2010). As expected for promoters that control the expression of virulence determinants, coupling of expression to AggR is tight, with high induction ratios. For some AraC-XylS family members that control bacterial virulence, specific host-derived signals are often sensed by the protein, which modulates the transcription factors activity (Yang *et al.*, 2009; Childers *et al.*, 2011). However, neither temperature nor specific molecules, such as bicarbonate ions or bile salts, seem to play a major role in AggR-dependent activation (Morin *et al.*, 2013) (Table S6). Thus, it is unclear what signal, if any, is sensed by AggR, especially as we were able to observe AggR-dependent activation in laboratory *E. coli* K-12, without any special induction conditions. It is of note that in EAEC, AggR activity is controlled by the Aar repressor protein (Santiago *et al.*, 2014), which could explain why we were able to detect AggR-dependent activity in its absence. Thus, it is clear that understanding the signal and mechanism by which the AggR-mediated regulation is initiated in EAEC strains will be key to understanding and designing small molecule inhibitors which can short circuit virulence in this important *E. coli* pathotype.

Table 2. Potential AggR-dependent promoters located on the chromosome of EAEC strain 042.

Gene name	Promoter sequence ^a	Spacing (bp) ^b	Distance to start (bp) ^c
<i>bssS</i>	GTACGTTATGCCGAGCGGAATAATTTATTCGGTTATGGCGCAACAAAAGAAGATAAACAGCGCATTAGCGAAATT	22	383
<i>figB</i>	TGTGGCCTGCACATTAACCTGTAATAATCGTTTGTCTGCCGAGCGGTGCCAACACATTCACATTGCCCCACAG	22	408
<i>cspl</i>	ATGGTGTCTGGTTTGTACAAATTTATCTGAAGCAGTCATTTGTTAATTTTATTTATTTGACCTCTTGAGATT	23	124
<i>fljE</i>	ACTGGTAGGCTTGGTACCAGAAATATCCGGGAGATGAAAATGTCAGCGATACAGGGGATGAAGGGGTTATCA	22	12
EC042_2219	TCTCCTTATACTAAAGAAATAATCAATATCAATAAATAATTAACACAGTCACACATTAATAAATAACACCAACA	23	491
EC042_2219	CGAGATAATATATGCAAAAATAATACGCCATTTGCTATACTATATGTCACACATGGGTGCATATAAC	5	5
EC042_2249	TTAGCCGATTTCTGTAAGGATTTATCGTTCAGACACACTCCCGGGCAACACTCCCGAGCACAATCACG	23	457
EC042_2249	TACCTCAGCAGTTGGTTACGTTTTATCTGGCTGTTATCCGAGCCCGAAATGAAAATTAACCTCCAGAA	22	123
EC042_2249	TACCCAGCCACTTTCCATTTTATATCATATACAGAAAATCTTCAGTATGAAAACATACCTGTTATACC	22	23
EC042_3188	TCACTGACACTGATGCTGTTTTATATCGCCTTTATCTCTCAGCCCGGGATACCCAGATCACCGAGATGA	22	24
EC042_4006	TGATGTGTGTCAGTTTCATCTTTTATCTCTTTTAACTGTTAATGACTATATGCGCAAGTACATACGATT	22	200
EC042_4430	TTTTGTTCAATTAATTTGAACTATATCAATGATGTTGTTAGCCAGATTAGCCGTGACTTTTATGGCC	22	286
EC042_4430	ATTCCTGTAGGAAATAGTTTTGAATATCAATGAAATTTTATTCAGGTGACGATTAATAAGGTATCAATTC	22	154
EC042_4430	TTTTATCAGGTGACGATTAATAAGGTATCAATTTCAATCAGGCAAAAAGTCTATTTATACCGTAAGATTATCT	23	114
<i>aatA</i> ^d	GTTATATCATTAATCAGCAAAAATGATACATGCTCCTCTTTTATGGTATCACATATAGAAATCCATGAA	23	237

^aPromoters were aligned to the AggR binding site consensus (WWWVWWWTATC), only allowing a maximum of 2 mismatches within the A/T tract region, and the extended -10 region consensus (TGnTATAAT). Each element is underlined and matches to each consensus are bold.

^bDistance between the AggR binding site and the -10 hexamer element.

^cDistance between the AggR binding site and the predicted translational start site of each gene. Sequence corresponding to the open reading frame of each gene is in red, where applicable.

^dCharacterized AggR-dependent promoter.

Experimental procedures

Bacterial strains, plasmids, primers and growth conditions

The bacterial strains, plasmids and promoter fragments used in this study are listed in Table S1. The oligonucleotide primers used for primer extension analysis and to amplify and mutate the various DNA fragments are listed in Table S7. Standard procedures for PCR, cloning and DNA manipulation were used throughout (Sambrook and Russell, 2001). All DNA fragments used in this study are flanked by EcoRI and HindIII sites and the DNA sequence of each fragment is numbered from the base adjacent to the HindIII site. Base substitutions are defined by the position of the nucleotide base altered and the substituted base introduced. Cells were routinely grown in Lysogeny Broth (LB medium) at 37°C with shaking. To measure promoter activities, fragments were cloned into the *lac* expression vector pRW50 (Lodge *et al.*, 1992) and maintained with 15 µg ml⁻¹ tetracycline. To examine the effect of *aggR* expression, cells were transformed with either pBAD/*aggR* or pBAD24, which were maintained in cells with 100 µg ml⁻¹ ampicillin or carbenicillin. AggR expression, using pBAD/*aggR*, was induced by the inclusion of 0.2% w/v arabinose in the medium, where appropriate (Sheikh *et al.*, 2002).

RNA isolation, rRNA depletion and cDNA synthesis for RNA-seq

Triplicate overnight cultures of EAEC 042 and EAEC 042 Δ *aggR* were used to inoculate 50 ml of Dulbecco's modified Eagle's medium with 0.45% glucose (DMEM high glucose) (Sigma) to an OD₆₀₀ of 0.05. Cultures were grown at 37°C with shaking to an OD₆₀₀ of 0.6. RNA was isolated using an RNeasy Mini Kit (Qiagen) and contaminating DNA was removed using an RNase-free DNase kit (Qiagen). The quality of the RNA was checked using an Agilent RNA 6000 Nano Chip (Agilent Technologies). RNA samples with a RIN (RNA integrity number) above 8 were then used for RNA-seq. A total of 3.5 µg of isolated RNA was used for each sample for rRNA depletion using a Ribo-Zero™ rRNA Removal Kit for bacteria (Illumina). Successful rRNA depletion was confirmed using an Agilent RNA 6000 Pico Chip (Agilent Technologies). The TruSeq® Stranded mRNA LT Sample Prep Kit (Illumina) was used to produce cDNA libraries, which were sequenced using a MiSeq Desktop Sequencer (Illumina). Raw sequence data were deposited under accession number PRJEB27566.

Differential gene expression analysis

Sequencing reads were filtered using Trimmomatic-0.36 and reads that did not pass the filter were discarded (Bolger *et al.*, 2014). Filtered reads were aligned using Burrows-Wheeler aligner to the EAEC 042 chromosome (FN554766.1) and the EAEC 042 pAA2 plasmid (FN554767.1) (Chaudhuri *et al.*, 2010; Li and Durbin, 2010). The alignment of reads to genes was counted using featureCounts (Liao *et al.*, 2014).

DESeq2 was used to determine differentially expressed genes (Love *et al.*, 2014). Genes were termed differentially expressed if there was a $>1 \log_2$ -fold difference and an adjusted p -value $< 1E-5$ between wild-type EAEC 042 and the $\Delta aggR$ mutant.

qRT-PCR

For qRT-PCR analysis, overnight cultures of EAEC 042 pBAD24, EAEC 042 $\Delta aggR$ pBAD24 and EAEC 042 $\Delta aggR$ pBAD/aggR, in triplicate, were used to inoculate 4 ml DMEM high glucose supplemented with $100 \mu\text{g ml}^{-1}$ carbenicillin to a final OD₆₀₀ of 0.05. Cultures were grown at 37°C with shaking as described above. At an OD₆₀₀ of 0.4, L-arabinose was added to a final concentration of 2%. Cultures were grown for 1 hour and RNA was extracted as described above. DNA was removed using TURBO DNA-free™ (Ambion). RNA was reverse transcribed to cDNA using the Tetro cDNA Synthesis Kit (Bioline). Reactions for qRT-PCR were prepared using the manufacturer's instructions for the Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies) and primers are detailed in Table S7. Relative gene expression was calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001), with the *polA* gene used as a reference.

Motility assays

Triplicate cultures of EAEC 042 $\Delta aggR$ pBAD24 and EAEC 042 $\Delta aggR$ pBAD/aggR were grown from overnight cultures to an OD₆₀₀ of 1, each culture was inoculated into the centre of LB 0.25% agar plate supplemented with 0.2% L-arabinose and incubated for 16 hours at 37°C. Plates were assessed for a difference in motility.

Promoter fragment and plasmid construction

The promoter fragments *aafD*100, *afaB*100, *aggD*100, *aap*100, *aatP*100 and *aaiA*100 were amplified by PCR using the primer pairs listed in Table S7 with EAEC 042 or EAEC 17-2 genomic DNA as template. The *aggD*101, *agg3D*100 and *agg4D*100 promoter fragments from EAEC strains C227-11, 55989 and C1010-00, respectively, were synthesized by Invitrogen Life Technologies. All DNA fragments are flanked by EcoRI and HindIII sites to facilitate cloning into pRW50 to generate *lacZ* transcriptional fusions. For shorter fragments, amplification was carried out using pRW50/*aafD*100, pRW50/*afaB*100 and pRW50/*aggD*100 as a template with the respective primers detailed in Table S7. Point mutations were introduced into fragments using megaprimer PCR, when necessary (Sarkar and Sommer, 1990). All constructs were verified by Sanger DNA sequencing.

Bioinformatic analysis of DNA sequences

DNA target sites for the binding of AggR and the closely related Rns protein have been previously investigated, using *in vivo* and *in vitro* approaches (Morin *et al.*, 2010; Munson, 2013). From these studies, it has been proposed

that the potential AggR-binding site consensus sequence is 5'-AnnnnnnTATC-3'. Thus, based on this consensus, promoter fragments were screened for potential AggR-binding sites on both strands, allowing for one mismatch to this consensus sequence. When predicted sites were found not to be necessary for AggR-mediated regulation, as judged by deletion analysis, they were discounted. Potential AggR-binding sequences, present in the smallest AggR-regulated fragment, were then investigated using mutational analysis to identify the functional site.

The WebLogo motifs for the AggR-binding site consensus sequence and AggR-dependent promoter organization were generated by the WebLogo server (<http://weblogo.berkeley.edu/logo.cgi>) (Crooks *et al.*, 2004) using sequences from the EAEC 042 *aafD*, *afaB*, *aap*, *aatP*, *aaiA* and *aggR* promoters, the EAEC 17-2 *aggD* promoter, the EAEC 55989 *agg3D* promoter and the EAEC C1010-00 *agg4D* promoter (Fig. 5) (Morin *et al.*, 2010; Yasir, 2017). The 3D models of DNA promoter architecture, for the EAEC 17-2 *aggD* and EAEC 042 *aafD* and *afaB* promoters, were produced by the model. it server using standard parameters (http://pongor.itk.ppke.hu/dna/model_it.html#modelit_intro) (Munteanu *et al.*, 1998) and PyMOL (Schrodinger, 2010).

To identify AggR-dependent promoters from our RNA-seq data (Tables S2 and S3) 600 bp of DNA upstream of the first gene in each operon was searched for AggR-binding sites, using the consensus sequence WWWWWWTATC (Fig. 7A), only allowing two mismatches in the A/T rich tract and no mismatches in the conserved TATC motif. The presence of a -10 element was then examined by determining if there was a suitable match to the -10 region consensus sequence (TGnTATAAT) at a spacing of 21 to 23 bp, ensuring that first A in the -10 hexamer was present, as this is an important determinant of promoter strength (Browning and Busby, 2004).

Assays of promoter activity

To assay the expression from promoter derivatives cloned into the *lac* expression vector pRW50, *E. coli* K-12 BW25113 Δlac strain was transformed with each construct and β -galactosidase activity was measured as described in our previous work (Jayaraman *et al.*, 1987). AggR was expressed from pBAD/aggR, which carries *aggR* cloned downstream of the arabinose inducible promoter, *paraBAD* (Sheikh *et al.*, 2002). Cells were grown in LB medium at 37°C with shaking to mid-logarithmic phase (OD₆₅₀ = 0.4–0.6) and 0.2% w/v arabinose was included in the medium to induce AggR expression, where appropriate. β -galactosidase activities are expressed as nmol of ONPG hydrolysed min⁻¹ mg⁻¹ dry cell mass and each activity is the average of three independent determinations.

Primer extension assay

Primer extension analysis was carried out as described in our previous work (Lloyd *et al.*, 2008). *E. coli* K-12 BW25113 cells, carrying various pRW50 derivatives and either pBAD/aggR or pBAD24, were grown in LB medium, containing 0.2% w/v arabinose, until mid-logarithmic

phase. RNA was extracted using an RNeasy Kit (Qiagen) and hybridized to ^{32}P end-labelled D49724 primer, which corresponds to sequence downstream of the HindIII site in pRW50 (Table S7). Primer extension products were run on a 6% denaturing polyacrylamide gel, containing $1 \times \text{TBE}$, and were analysed using a Bio-Rad Molecular Imager FX and Quantity One software (Bio-Rad). Gels were calibrated using an M13 sequence ladder, which was generated using a T7 sequencing kit (USB) with single-stranded M13mp18 phage DNA and the M13 Universal Primer (Table S7).

Acknowledgements

This work was generously supported by a Darwin Trust of Edinburgh PhD studentship to MY and by the Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) to support the translation, development and commercialization of innovative Industrial Biotechnology processes.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

DFB, SJWB, MP, IRH and MY conceived and designed the research project. MY, CI, RA and REG performed the experiments, aided by DFB, JRH and PS. MY, CI, SJWB, IRH and DFB wrote the manuscript with input from all authors.

References

- Adachi, J.A., Jiang, Z.D., Mathewson, J.J., Verenkar, M.P., Thompson, S., Martinez-Sandoval, F., *et al.* (2001) Enteroaggregative *Escherichia coli* as a major etiologic agent in traveler's diarrhea in 3 regions of the world. *Clinical Infectious Diseases*, **32**, 1706–1709.
- Bernier, C., Gounon, P. and Le Bouguenec, C. (2002) Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia coli* as a sensitive probe for detecting the AAF-encoding operon family. *Infection and Immunity*, **70**, 4302–4311.
- Boisen, N., Melton-Celsa, A.R., Scheutz, F., O'Brien, A.D. and Nataro, J.P. (2015) Shiga toxin 2a and enteroaggregative *Escherichia coli* – a deadly combination. *Gut Microbes*, **6**, 272–278.
- Boisen, N., Struve, C., Scheutz, F., Krogfelt, K.A. and Nataro, J.P. (2008) New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. *Infection and Immunity*, **76**, 3281–3292.
- Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Browning, D.F. and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, **2**, 57–65.
- Browning, D.F. and Busby, S.J. (2016) Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, **14**, 638–650.
- Caron, J. and Scott, J.R. (1990) A *rms*-like regulatory gene for colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. *Infection and Immunity*, **58**, 874–878.
- Chaudhuri, R.R., Sebaihia, M., Hobman, J.L., Webber, M.A., Leyton, D.L., Goldberg, M.D., *et al.* (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS ONE*, **5**, e8801.
- Childers, B.M., Cao, X., Weber, G.G., Demeler, B., Hart, P.J. and Klose, K.E. (2011) N-terminal residues of the *Vibrio cholerae* virulence regulatory protein ToxT involved in dimerization and modulation by fatty acids. *Journal of Biological Chemistry*, **286**, 28644–28655.
- Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Research*, **14**, 1188–1190.
- Czczulin, J.R., Balepur, S., Hicks, S., Phillips, A., Hall, R., Kothary, M.H., *et al.* (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. *Infection and Immunity*, **65**, 4135–4145.
- Domka, J., Lee, J. and Wood, T.K. (2006) YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signaling. *Applied and Environmental Microbiology*, **72**, 2449–2459.
- Dudley, E.G., Thomson, N.R., Parkhill, J., Morin, N.P. and Nataro, J.P. (2006) Proteomic and microarray characterization of the AggR regulon identifies a *pheU* pathogenicity island in enteroaggregative *Escherichia coli*. *Molecular Microbiology*, **61**, 1267–1282.
- Durrer, P., Zbinden, R., Fleisch, F., Altwegg, M., Ledergerber, B., Karch, H., and *et al.* (2000) Intestinal infection due to enteroaggregative *Escherichia coli* among human immunodeficiency virus-infected persons. *Journal of Infectious Diseases*, **182**, 1540–1544.
- Egan, S.M. (2002) Growing repertoire of AraC/XylS activators. *Journal of Bacteriology*, **184**, 5529–5532.
- Elias, W.P. Jr, Czczulin, J.R., Henderson, I.R., Trabulsi, L.R. and Nataro, J.P. (1999) Organization of biogenesis genes for aggregative adherence fimbria II defines a virulence gene cluster in enteroaggregative *Escherichia coli*. *Journal of Bacteriology*, **181**, 1779–1785.
- Estrada-Garcia, T. and Navarro-Garcia, F. (2012) Enteroaggregative *Escherichia coli* pathotype: a genetically heterogeneous emerging foodborne enteropathogen. *FEMS Immunology and Medical Microbiology*, **66**, 281–298.
- Fasano, A., Noriega, F.R., Liao, F.M., Wang, W. and Levine, M.M. (1997) Effect of shigella enterotoxin 1 (ShET1) on rabbit intestine *in vitro* and *in vivo*. *Gut*, **40**, 505–511.
- Franca, F.L., Wells, T.J., Browning, D.F., Nogueira, R.T., Sarges, F.S., Pereira, A.C., *et al.* (2013) Genotypic and phenotypic characterisation of enteroaggregative *Escherichia coli* from children in Rio de Janeiro. *Brazil. PLoS One*, **8**, e69971.

- Frank, C., Werber, D., Cramer, J.P., Askar, M., Faber, M., an der Heiden, M., *et al.* (2011) Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *New England Journal of Medicine*, **365**, 1771–1780.
- Fujiyama, R., Nishi, J., Imuta, N., Tokuda, K., Manago, K. and Kawano, Y. (2008) The *shf* gene of a *Shigella flexneri* homologue on the virulent plasmid pAA2 of enteroaggregative *Escherichia coli* O42 is required for firm biofilm formation. *Current Microbiology*, **56**, 474–480.
- Harada, T., Hiroi, M., Kawamori, F., Furusawa, A., Ohata, K., Sugiyama, K., *et al.* (2007) A food poisoning diarrhea outbreak caused by enteroaggregative *Escherichia coli* serogroup O126:H27 in Shizuoka, Japan. *Japanese Journal of Infectious Diseases*, **60**, 154–155.
- Harrington, S.M., Dudley, E.G. and Nataro, J.P. (2006) Pathogenesis of enteroaggregative *Escherichia coli* infection. *FEMS Microbiology Letters*, **254**, 12–18.
- Harrington, S.M., Sheikh, J., Henderson, I.R., Ruiz-Perez, F., Cohen, P.S. and Nataro, J.P. (2009) The Pic protease of enteroaggregative *Escherichia coli* promotes intestinal colonization and growth in the presence of mucin. *Infection and Immunity*, **77**, 2465–2473.
- Henderson, I.R., Hicks, S., Navarro-Garcia, F., Elias, W.P., Phillips, A.D. and Nataro, J.P. (1999) Involvement of the enteroaggregative *Escherichia coli* plasmid-encoded toxin in causing human intestinal damage. *Infection and Immunity*, **67**, 5338–5344.
- Henderson, I.R., Meehan, M. and Owen, P. (1997) A novel regulatory mechanism for a novel phase-variable outer membrane protein of *Escherichia coli*. *Advances in Experimental Medicine and Biology*, **412**, 349–355.
- Herzog, K., Engeler Dusel, J., Hugentobler, M., Beutin, L., Sagesser, G., Stephan, R., *et al.* (2014) Diarrheagenic enteroaggregative *Escherichia coli* causing urinary tract infection and bacteremia leading to sepsis. *Infection*, **42**, 441–444.
- Itoh, Y., Nagano, I., Kunishima, M. and Ezaki, T. (1997) Laboratory investigation of enteroaggregative *Escherichia coli* O untypeable:H10 associated with a massive outbreak of gastrointestinal illness. *Journal of Clinical Microbiology*, **35**, 2546–2550.
- Jayaraman, P.S., Peakman, T.C., Busby, S.J., Quincey, R.V. and Cole, J.A. (1987) Location and sequence of the promoter of the gene for the NADH-dependent nitrite reductase of *Escherichia coli* and its regulation by oxygen, the Fnr protein and nitrite. *Journal of Molecular Biology*, **196**, 781–788.
- Jonsson, R., Struve, C., Boisen, N., Mateiu, R.V., Santiago, A.E., Jenssen, H., *et al.* (2015) Novel aggregative adherence fimbria variant of enteroaggregative *Escherichia coli*. *Infection and Immunity*, **83**, 1396–1405.
- Jordi, B.J., Dagberg, B., de Haan, L.A., Hamers, A.M., van der Zeijst, B.A., Gaastra, W., *et al.* (1992) The positive regulator CfaD overcomes the repression mediated by histone-like protein H-NS (H1) in the CFA/I fimbrial operon of *Escherichia coli*. *EMBO Journal*, **11**, 2627–2632.
- Knutton, S., Adu-Bobie, J., Bain, C., Phillips, A.D., Dougan, G. and Frankel, G. (1997) Down regulation of intimin expression during attaching and effacing enteropathogenic *Escherichia coli* adhesion. *Infection and Immunity*, **65**, 1644–1652.
- Korobkova, E., Emonet, T., Vilar, J.M., Shimizu, T.S. and Cluzel, P. (2004) From molecular noise to behavioural variability in a single bacterium. *Nature*, **428**, 574–578.
- Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, **26**, 589–595.
- Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
- Lindsay, J.A. (2014) Evolution of *Staphylococcus aureus* and MRSA during outbreaks. *Infection, Genetics and Evolution*, **21**, 548–553.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
- Lloyd, G.S., Hollands, K., Godfrey, R.E. and Busby, S.J. (2008) Transcription initiation in the *Escherichia coli* K-12 *mal*-*malX* intergenic region and the role of the cyclic AMP receptor protein. *FEMS Microbiology Letters*, **288**, 250–257.
- Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N.R. (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiology Letters*, **74**, 271–276.
- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.
- Martin, R.G. and Rosner, J.L. (2001) The AraC transcriptional activators. *Current Opinion in Microbiology*, **4**, 132–137.
- Morin, N., Santiago, A.E., Ernst, R.K., Guillot, S.J. and Nataro, J.P. (2013) Characterization of the AggR regulon in enteroaggregative *Escherichia coli*. *Infection and Immunity*, **81**, 122–132.
- Morin, N., Tirling, C., Ivison, S.M., Kaur, A.P., Nataro, J.P. and Steiner, T.S. (2010) Autoactivation of the AggR regulator of enteroaggregative *Escherichia coli* *in vitro* and *in vivo*. *FEMS Immunology and Medical Microbiology*, **58**, 344–355.
- Munson, G.P. (2013) Virulence regulons of enterotoxigenic *Escherichia coli*. *Immunologic Research*, **57**, 229–236.
- Munteanu, M.G., Vlahovicek, K., Parthasarathy, S., Simon, I. and Pongor, S. (1998) Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends in Biochemical Sciences*, **23**, 341–347.
- Murphree, D., Froehlich, B. and Scott, J.R. (1997) Transcriptional control of genes encoding CS1 pili: negative regulation by a silencer and positive regulation by Rns. *Journal of Bacteriology*, **179**, 5736–5743.
- Nataro, J.P., Mai, V., Johnson, J., Blackwelder, W.C., Heimer, R., Tirrell, S., *et al.* (2006) Diarrheagenic *Escherichia coli* infection in Baltimore, Maryland, and New Haven, Connecticut. *Clinical Infectious Diseases*, **43**, 402–407.
- Nataro, J.P., Yikang, D., Yingkang, D. and Walker, K. (1994) AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. *Journal of Bacteriology*, **176**, 4691–4699.

- Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., *et al.* (2003) The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *Journal of Biological Chemistry*, **278**, 45680–45689.
- Okeke, I.N., Lamikanra, A., Czczulin, J., Dubovsky, F., Kaper, J.B. and Nataro, J.P. (2000) Heterogeneous virulence of enteroaggregative *Escherichia coli* strains isolated from children in Southwest Nigeria. *Journal of Infectious Diseases*, **181**, 252–260.
- Olesen, B., Scheutz, F., Andersen, R.L., Menard, M., Boisen, N., Johnston, B., *et al.* (2012) Enteroaggregative *Escherichia coli* O78:H10, the cause of an outbreak of urinary tract infection. *Journal of Clinical Microbiology*, **50**, 3703–3711.
- Porter, M.E., Smith, S.G. and Dorman, C.J. (1998) Two highly related regulatory proteins, *Shigella flexneri* VirF and enterotoxigenic *Escherichia coli* Rns, have common and distinct regulatory properties. *FEMS Microbiology Letters*, **162**, 303–309.
- Prieto, A., Bernabeu, M., Aznar, S., Ruiz-Cruz, S., Bravo, A., Queiroz, M.H., *et al.* (2018) Evolution of bacterial global modulators: role of a novel H-NS paralogue in the enteroaggregative *Escherichia coli* strain 042. *mSystems*, **3**, e00220–17.
- Rasko, D.A., Webster, D.R., Sahl, J.W., Bashir, A., Boisen, N., Scheutz, F., *et al.* (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *New England Journal of Medicine*, **365**, 709–717.
- Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N.J., Hentschke, M., *et al.* (2011) Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *New England Journal of Medicine*, **365**, 718–724.
- Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Santiago, A.E., Ruiz-Perez, F., Jo, N.Y., Vijayakumar, V., Gong, M.Q. and Nataro, J.P. (2014) A large family of antivirulence regulators modulates the effects of transcriptional activators in Gram-negative pathogenic bacteria. *PLoS Pathogens*, **10**, e1004153.
- Sarantuya, J., Nishi, J., Wakimoto, N., Erdene, S., Nataro, J.P., Sheikh, J., *et al.* (2004) Typical enteroaggregative *Escherichia coli* is the most prevalent pathotype among *E. coli* strains causing diarrhea in Mongolian children. *Journal of Clinical Microbiology*, **42**, 133–139.
- Sarkar, G. and Sommer, S.S. (1990) The “megaprimer” method of site-directed mutagenesis. *BioTechniques*, **8**, 404–407.
- Savarino, S.J., Fasano, A., Robertson, D.C. and Levine, M.M. (1991) Enteroaggregative *Escherichia coli* elaborate a heat-stable enterotoxin demonstrable in an *in vitro* rabbit intestinal model. *Journal of Clinical Investigation*, **87**, 1450–1455.
- Savarino, S.J., Fox, P., Deng, Y. and Nataro, J.P. (1994) Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. *Journal of Bacteriology*, **176**, 4949–4957.
- Schrodinger, L.L.C. (2010) *The PyMOL Molecular Graphics System*. Version 1.3r1.
- Sheikh, J., Czczulin, J.R., Harrington, S., Hicks, S., Henderson, I.R., Le Bouguenec, C., *et al.* (2002) A novel dispersin protein in enteroaggregative *Escherichia coli*. *Journal of Clinical Investigation*, **110**, 1329–1337.
- Sheikh, J., Hicks, S., Dall’Agnol, M., Phillips, A.D. and Nataro, J.P. (2001) Roles for Fis and YafK in biofilm formation by enteroaggregative *Escherichia coli*. *Molecular Microbiology*, **41**, 983–997.
- Singh, G., Srinivasan, R., Cheng, J., Peng, Z., Fujimura, K., Baek, M.S., *et al.* (2014) Rearrangement of a large novel *Pseudomonas aeruginosa* gene island in strains isolated from a patient developing ventilator-associated pneumonia. *Journal of Clinical Microbiology*, **52**, 2430–2438.
- Singh, K., Milstein, J.N. and Navarre, W.W. (2016) Xenogeneic silencing and its impact on bacterial genomes. *Annual Review of Microbiology*, **70**, 199–213.
- Spudich, J.L. and Koshland, D.E. Jr (1976) Non-genetic individuality: chance in the single cell. *Nature*, **262**, 467–471.
- Tobe, T., Yoshikawa, M., Mizuno, T. and Sasakawa, C. (1993) Transcriptional control of the invasion regulatory gene *virB* of *Shigella flexneri*: activation by VirF and repression by H-NS. *Journal of Bacteriology*, **175**, 6142–6149.
- Verma, R., Rojas, T.C.G., Maluta, R.P., Leite, J.L., Nakazato, G. and de Silveira, W.D. (2018) Role of hypothetical protein YicS in the pathogenicity of Avian Pathogenic *Escherichia coli* in vivo and in vitro. *Microbiological Research*, **214**, 28–36.
- Wan, T.W., Higuchi, W., Khokhlova, O.E., Hung, W.C., Iwao, Y., Wakayama, M., *et al.* (2017) Genomic comparison between *Staphylococcus aureus* GN strains clinically isolated from a familial infection case: IS1272 transposition through a novel inverted repeat-replacing mechanism. *PLoS ONE*, **12**, e0187288.
- Webster, C., Gaston, K. and Busby, S. (1988) Transcription from the *Escherichia coli* *melR* promoter is dependent on the cyclic AMP receptor protein. *Gene*, **68**, 297–305.
- Wilson, A., Evans, J., Chart, H., Cheasty, T., Wheeler, J.G., Tompkins, D., *et al.* (2001) Characterisation of strains of enteroaggregative *Escherichia coli* isolated during the infectious intestinal disease study in England. *European Journal of Epidemiology*, **17**, 1125–1130.
- Yang, J., Dogovski, C., Hocking, D., Tauschek, M., Perugini, M. and Robins-Browne, R.M. (2009) Bicarbonate-mediated stimulation of RegA, the global virulence regulator from *Citrobacter rodentium*. *Journal of Molecular Biology*, **394**, 591–599.
- Yasir, M. (2017) *Regulation of virulence determinants in enteroaggregative Escherichia coli*. Ph.D. Thesis. University of Birmingham, Birmingham, UK.
- Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K. and Liu, X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, **9**, e78644.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.