Massively parallel transposon mutagenesis identifies temporally essential genes for biofilm formation in *Escherichia coli*

Authors: Emma R Holden[1], Muhammad Yasir[1], A Keith Turner[1], John Wain[1,2], Ian G. Charles[1,2], Mark A Webber[1,2] *

[1] Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, NR4 7UQ, U.K.

[2] Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, U.K.

* Corresponding author Mark.Webber@quadram.ac.uk

Keywords: Functional genomics, transposon mutagenesis, adhesion, TraDIS

Running title: *E. coli* TraDIS biofilm

Abstract

Biofilms complete a life cycle where cells aggregate, grow and produce a structured community before dispersing to seed biofilms in new environments. Progression through this life cycle requires controlled temporal gene expression to maximise fitness at each stage. Previous studies have focused on the essential genome for the formation of a mature biofilm, but here we present an insight into the genes involved at different stages of biofilm formation. We used TraDIS-Xpress; a massively parallel transposon mutagenesis approach using transposon-located promoters to assay expression of all genes in the genome. We determined how gene essentiality and expression affects the fitness of *E. coli* growing as a biofilm on glass beads after 12, 24 and 48 hours. A selection of genes identified as important were then validated independently by assaying biofilm biomass, aggregation, curli biosynthesis and adhesion ability of defined mutants. We identified 48 genes that affected biofilm fitness including genes with known roles and those not previously implicated in biofilm formation. Regulation of type 1 fimbriae and motility were important at all time points. Adhesion and motility were important for the early biofilm, whereas matrix production and purine biosynthesis were only important as the biofilm matured. We found strong temporal contributions to biofilm fitness for some genes including some which were both beneficial and detrimental depending on the stage at which they are expressed, including *dksA* and *dsbA*. Novel genes implicated in biofilm formation included *ychM* and *truA* involved in cell division, *crfC* and *maoP* in DNA housekeeping and *yigZ* and *ykgJ* of unknown function. This work provides new insights into the requirements for successful biofilm formation through the biofilm life cycle and demonstrates the importance of understanding expression and fitness through time.

Introduction

Bacteria rarely exist planktonically outside of the laboratory and are usually found as part of structured, aggregated communities called biofilms [1]. Clinically, approximately 80% of infections have been suggested to have a biofilm component [2]. Biofilm related infections are complicated by their intrinsic tolerance to antimicrobials, making infections difficult to treat and, often persistent [3-6]. Cells within a biofilm grow more slowly than those in planktonic culture and this reduced level of metabolic activity has been associated with tolerance to antimicrobials, allowing biofilms to be typically 10-1000-fold less sensitive to antibiotics than corresponding strains in planktonic conditions [7,8]. Aside from the clinical setting, there are many useful applications of biofilms, including wastewater treatment and bioprocessing [9]. Biofilms undergo a life cycle that commonly consists of initial attachment to a surface, growth and maturation of the biofilm over time with characteristic production of extracellular matrix components, followed by dispersal of planktonic cells to facilitate colonisation of new surfaces [10]. The switch between planktonic and biofilm lifestyles is driven by environmental stimuli promoting large scale changes in gene expression and regulation that are necessary to support the bacterial community through the life cycle, which is distinct to planktonic growth conditions. Expressing the right genes at the right time and place is critical for efficient production of a biofilm.

The main components of the biofilm extracellular matrix in *E. coli* are the amyloid protein curli, the polysaccharide cellulose and extracellular DNA [11]. Curli is transcribed by the divergent operons *csgBAC* and *csgDEFG*, with curli biosynthesis regulated by CsgD [12]. Cellulose biosynthetic machinery is encoded by *bcsRQABZC* and *bcsEFG*, and its production is regulated by c-di-GMP [13]. Several genes are known to be involved in the regulation of matrix production, including *ompR* [14,15], *cpxR* [14,16,17] and *rpoS* [18,19], amongst others [20-22]. Extracellular DNA is also an important component of the biofilm matrix, and the addition of DNase has negatively affected the biomass of biofilms formed by *Pseudomonas aeruginosa* [23], *Bacillus cereus* [24] and a range of Gram-negative pathogens, including *E. coli* [25].

Most previous studies focusing on identifying the genes and pathways required for biofilm formation in *E. coli* have concentrated on the mature biofilm rather than dissecting events across the life cycle. Various studies have taken a genome wide approach to identifying genes that impact biofilm formation. One assessed biofilm formation of all the mutants in the Keio collection [26], although analysis is limited to the effect of inactivation [27]. Another used a transcriptomic approach to identify genes with altered expression in biofilms compared to in suspension over time [28]. DNA microarrays have also been used to link the presence of different genes with biofilm capacity in panels of isolates [29].

Large scale transposon mutagenesis represent another high-throughput, sensitive whole genome approach to link phenotype to genotype methodologies [30-32]. These methods have the advantage of assessing the impact of a very large number of random insertions within and without genes on a phenotype. They have however been historically limited by an inability to assay essential genes where transposon insertions are not viable. We sought to investigate biofilm formation using TraDIS-Xpress, a method developed recently which allows essential genes to be assayed for their role in survival in test conditions. In this method, gene expression changes caused by an outward-facing inducible promoter incorporated into the transposon are captured as well as traditional essentiality measurements [33].

This study identified 48 genes that were found to be important at different stages of biofilm formation by *E. coli*. By investigating the genes important across the biofilm life cycle, we were able to get a dynamic view of the main pathways important at different stages of biofilm development. Genes identified by TraDIS-Xpress as important at each stage were then phenotypically validated using defined mutants. Our findings reinforced the importance of adhesion, motility and matrix production in the biofilm, and revealed the importance of genes not previously implicated in biofilm formation. This included genes with roles in cell division (*ychM* and *truA*), DNA housekeeping (*crfC* and *maoP*) and *yigZ* and *ykgJ*, the functions of which have not

been elucidated. We identified clear requirements for some pathways at specific points of the biofilm life cycle, furthering our understanding of how biofilms maintain fitness over time.

Results

Analysis of the TraDIS-Xpress data found 48 genes that considerably affected biofilm formation over time in *E. coli*: 42 were identified as being needed to maintain the fitness of a biofilm and expression of 6 genes was predicted to result in reduced biofilm fitness (Figure 1 and Supplementary table 1). The main pathways that were consistently important in the biofilm through all the time points included type 1 fimbriae, curli biosynthesis and regulation of flagella (Figure 1). All other loci identified affected biofilm formation at specific points in the life cycle.

<u>Fimbriae expression and motility are important at all stages of biofilm formation</u>

Only 4 genes were found to be important throughout 12, 24 and 48 hours (figure 2a). These included *fimB* and *fimE*. The recombinase gene *fimB* which helps mediate both 'ON-to-OFF' and 'OFF-to-ON' switching of fimbriae expression was beneficial for biofilm formation at all time points. Fewer *fimB* mutants were observed in biofilm conditions compared to planktonic, and this number decreased over time. In contrast, inactivation of *fimE*, responsible for only 'ON-to-OFF' fimbrial regulation [34], increased biofilm fitness at all time points. Initially, there were only slightly more *fimE* mutants in biofilm conditions compared to planktonic at 12 hours, but this increased over time with a stark contrast seen between biofilm and planktonic conditions at the 24 and 48 hour time points. The predicted impacts on biofilm formation were phenotypically validated by testing both gene knockout mutants from the Keio collection (which contains two, independent deletion mutants for most genes in *E. coli* BW25113) [35]. Analysis of these mutants showed for both that there was no deficit in biofilm biomass (Figure 3a), but both were deficient in cell aggregation (Figure 3b). Together, the TraDIS-*Xpress* and phenotypic data suggest that the ability to regulate fimbriae expression in a phase-dependent manner is important for fitness of a biofilm, rather than being constrained in a fimbriae 'ON' or 'OFF' state.

Disruption of *lrhA*, a regulator of motility and chemotaxis [36], was beneficial for biofilm formation at all time points (Figure 2b). LrhA also has a role in type 1 fimbriae expression through activating expression of *fimE* [37], but in addition represses flagella-mediated motility. Analysis of the Δ*lrhA*

biofilm showed initial formation of microcolonies occurred faster than the wild-type (Figure 4) but at later time points the biofilms formed by this mutant were less mature than seen with the wild-type. There was no significant change in biomass formed by this mutant (Figure 3a) and mutants appeared less aggregative than the wild-type (Figure 3b). These data suggest that inactivation of *lrhA* impacts both adhesion and aggregation differently at distinct stages of the biofilm life cycle and may result in a benefit to early surface colonisation but with a cost to later maturation.

Expression of the Hha toxin attenuator *tomB* was also found to be consistently important for biofilm formation at 12, 24 and 48 hours (Figure 2b). Consistent with this prediction, the Δ*tomB* mutant biofilm had reduced cell aggregation and curli biosynthesis, and reduced biofilm biomass (Figure 3 a,b,c).

<u>Regulatory genes are important in the early biofilm</u>

In the early biofilm, at 12 hours growth, only 13 genes were found to distinguish the planktonic and biofilm conditions. TraDIS-*Xpress* data indicated that inactivation of transcriptional factor *dksA* promoted biofilm formation at the 12- and 24-hour time points but not in the mature biofilm (Figure 2c). Supporting this, analysis of Δ*dksA* mutant biofilms under flow conditions showed an initial benefit with increased adhesion at both the 12 and 24 hour time points, but reduced microcolony formation at the 48 hour time point, suggesting *dksA* affects biofilm initiation (Figure 4). Inactivation of Δ*dksA* was also seen to reduce cell aggregation, curli biosynthesis and biofilm biomass (Figure 3 a,b,c). Expression of *hdfR*, a negative regulator of motility [38], was found to be detrimental to biofilm fitness in the early biofilm after 12 and 24 hours growth (Figure 2b), and Δ*hdfR* mutant biofilms had significantly reduced biomass (Figure 3a). In addition, the stress response regulator *marR* [39] and the 23S rRNA methyltransferase *rlmI* [40] were both found to be beneficial for biofilm fitness at the 12 hour time point only, and reduced biofilm biomass was found in deletion mutants (figure 3a). These genes have both previously been implicated in biofilm formation [40-42], but the effect on early biofilm formation has not been described previously.

Biofilms sampled after 24 hours demonstrate both adhesion and matrix production are important

More pathways were identified as being important to biofilm formation at 24 hours that at 12 hours. Two genes involved in DNA housekeeping were found to be involved in biofilm formation at this time point. These included *dam*, encoding DNA methyltransferase [43], insertional activation of which was not tolerated in the 24 hour biofilm, and Δ*dam* mutants were defective in aggregation compared to the wild type (figure 3b). Also, inactivation of *maoP,* involved in Ori macrodomain organisation [44], was predicted to confer a fitness advantage in the 24 hour biofilm compared to the planktonic condition. TraDIS-*Xpress* data showed more reads mapped to *maoP* in the biofilm conditions compared to the planktonic at 24 hours suggesting loss of this gene was beneficial. Phenotypic analysis of the Δ*maoP* mutant biofilm did demonstrate a phenotype although in opposition to the prediction, *maoP* mutants were significantly deficient in biofilm biomass production, curli biosynthesis and one mutant displayed reduced aggregation (Figure 3 a,c).

In the 12 and 24 hour biofilms, *dsbA* (encoding disulphide oxidoreductase) was essential with no insertions detected within this gene (Figure 2c). The role of *dsbA* in adhesion to abiotic surfaces and epithelial cells has previously been suggested (Lee et al., 2008, Bringer et al., 2007). Phenotypic validation of the Δ*dsbA* mutant showed a red, dry and rough (*rdar*) phenotype on Congo red plates (Figure 3c), indicative of increased curli biosynthesis. Cell aggregation in the Δ*dsbA* mutant was significantly improved compared to the wild type, implying a role of *dsbA* in inhibiting cell-cell aggregation. Our data showed that *dsbA* is important in the early biofilm, but its deletion appears to be beneficial to the formation of a mature biofilm, according to the Congo red and aggregation data. Supporting this hypothesis, *dsbA* was not essential at the 48 hour time point.

Inactivation of the RNase III modulator *ymdB* [45] was found to reduce fitness in the biofilm, with fewer reads mapping here in biofilm conditions compared to planktonic at both the 24 and 48 hour time points. This follows previous findings that both inactivation and overexpression of *ymdB*

negatively affects biofilm biomass [45]. In concordance with these findings, a Δ*ymdB* mutant had significantly improved cell aggregation and reduced biofilm biomass (figure 3 a,b).

Curli biosynthesis became important by the 24 hour time point as no insertions mapped to *csgC*, encoding a curli subunit chaperone [46] and more transposon insertions mapped upstream of the curli biosynthesis regulator *csgD* [12], likely indicating its increased expression benefitted biofilm formation. At the 48 hour time point, both genes were essential for biofilm formation, which was also the case for the known *csgD* regulator, *ompR* [14], supported by significantly reduced biofilm biomass and reduced aggregation in knockout mutants (figure 3a,b).

<u>The mature biofilm grown for 48 hours requires purine biosynthesis, matrix production, motility and solute transport</u>

There were 38 genes found to be important for fitness of the mature biofilm after 48 hours growth, and 25 of these genes were identified as essential at this time point only. The major pathway implicated in biofilm formation at 48 hours was purine ribonucleotide biosynthesis, with four genes, *purD*, *purH*, *purL* and *purE* [47], found to be essential at this time point only. TraDIS-Xpress did not identify mutants in any of these genes in biofilms sampled at 48 hours, whereas several reads mapped to these loci under planktonic conditions, as well as under both biofilm and planktonic conditions earlier at 12 and 24 hours. Visualisation of a Δ*purD* mutant biofilm under flow conditions saw poor biofilm formation and no microcolony formation at any time compared to the wild type (Figure 4). Additionally, Δ*purD* and Δ*purE* mutants were deficient in biofilm biomass production, curli biosynthesis, and Δ*purE* also showed reduced cell aggregation (Figure 3 a,b,c), confirming an important role for purine biosynthesis in matrix production and curli biosynthesis in the mature biofilm.

The flagella master regulatory system *flhDC* was identified as important in the mature biofilm. Biofilms sampled after 48 hours saw fewer *flhC* mutants, while insertions interpreted as over-expressing *flhD* increased in numbers both at the 24 and 48 hour time points, compared to

planktonic conditions. No mutants in *flgD* and *fliE*, encoding flagellar filament proteins, were identified at 24 and 48 hours, respectively. It has previously been shown that motility is important for initial biofilm formation [48,49], but this may not relate to biomass formation where no differences were seen for Δ*flhD*, Δ*flhC*, Δ*fliE* and Δ*flgD* mutants.

Various pleiotropic transcriptional regulators were also important in the mature biofilm. This included the H-NS antagonist *leuO* [50]. Increased insertions upstream of *leuO* under biofilm conditions after 12 hours growth, as well as no *leuO* mutants in 48 hour biofilms, indicated it was beneficial to biofilm formation. A Δ*leuO* mutant did not aggregate as well as the wild type, and one Δ*leuO* mutant had reduced biofilm biomass (figure 3 a,b). The Δ*leuO* mutant biofilm under flow conditions demonstrated an inability to form microcolonies after 48 hours growth (figure 4). The leucine-responsive global regulator *lrp* [51] and a transcriptional regulator responsible for survival under acid stress, *gadW* [52] were also found to have fewer mutants in the 48 hour biofilm compared to the planktonic condition, indicating their importance in the mature biofilm. Reduced biofilm biomass, aggregation and curli biosynthesis were observed for one copy of Δ*lrp*, but no differences in biofilm formation or aggregation were seen for Δ*gadW* mutant biofilms (figure 3 a,b,c).

Inactivation of the outer membrane channels *mscL* [53], *tolC* [54] and *ompF* [55] was not tolerated in the mature biofilm grown for 48 hours. This would indicate the importance of transport in the mature biofilm, however inactivation of *tolC* and *ompF* did not result in a change in biofilm biomass (figure 3a).

Three genes involved in cell division or DNA replication, *crfC*, *ychM* and *truA*, were identified as important in the 48 hour biofilm. Expression of *yhcM*, involved in cell division [56] was predicted to be essential in the 48 hour biofilm, but this was not reflected in the phenotype of deletion mutants (Figure 3 a). The pseudouridine synthase *truA* [57], also reported to be involved in cell division [58] was found to be essential in the mature biofilm grown for 24 and 48 hours. Although a Δ*truA* mutant showed no change in biofilm biomass or curli biosynthesis, aggregation was significantly improved

compared to the wild type albeit to a small degree (figure 3b), contrary to the prediction generated by TraDIS-*Xpress*.

Discussion

We have characterised the essential genome of *E. coli* biofilms across the biofilm lifecycle by using the high throughput transposon mutagenesis screen TraDIS-*Xpress*. The identification of genes and pathways already described to be involved in biofilm formation validates the efficacy of this experimental model. The early biofilm established 12 hours after inoculation was characterised by genes involved in adhesion. The 24-hour biofilm required both adhesion and matrix production, transitioning into matrix production being of the upmost importance in the mature biofilm after 48 hours. In concordance with previous work identifying genes whose importance varies with time in the *E. coli* biofilm, we also reported that control of fimbriae expression and motility remained important at each stage of the biofilm life cycle rather than just being involved in initial attachment [28].

As well as identifying known pathways, TraDIS-*Xpress* was also able to identify genes not previously known to be involved in biofilm formation, including *yigZ*, *ykgJ*, *ychM*, *maoP*, *truA*, *crfC*. We found that expression of *maoP* was detrimental to the fitness of biofilms grown for 24 hours, but a Δ*maoP* mutant biofilm had reduced biofilm biomass and reduced curli biosynthesis compared to the wild type. Chromosomal organisation of the Ori macrodomain requires both *maoP* and *maoS* [44], however no signal is seen in our data for *maoS*. This warrants further investigation into how chromosomal macrodomain organisation affects biofilm formation. The importance of cell division in the mature biofilm was further supported by our findings of fewer *ychM* and *truA* mutants compared to planktonic conditions, both of which have not before been implicated in biofilm formation. However, deletion of either gene had no effect on biofilm biomass or curli production, and the Δ*truA* deletion mutant had improved aggregation compared to the wild type. Genes involved in cell division are clearly important for the fitness of the mature biofilm, but the essential roles of *ychM* and *truA* in this process are currently unclear but unlikely to relate to crude biomass production.

Expression of *dsbA* and repression of *dksA* was found in this study to benefit early biofilm fitness. Based on previous studies and phenotypic analysis of knockout mutants in this study, we believe the increase in biofilm fitness seen is due to increased adhesion in these mutants [59,60]. This study has highlighted the benefit of close temporal gene regulation in the biofilm, where the expression of certain genes can have a different effect on biofilm fitness at different stages of the biofilm life cycle. For example, we found that *dsbA* was important for the early biofilm, but a *dsbA* mutant biofilm had increased curli expression and increased aggregation. Expression of *dsbA* has been previously found to result in repression of the curli regulator *csgD* and curli subunit *csgA*, essential for optimal fitness of the mature biofilm [61]. Conversely, we found that expression of the transcription factor *dksA* was detrimental in the early biofilm, whilst a *dksA* knockout biofilm had reduced biofilm biomass, reduced curli biosynthesis and reduced aggregation. The relationship between *dksA* expression and curli biosynthesis has been previously characterised in similar studies [32,62]. Again, these data show differential expression of important genes at different stages of the biofilm life cycle is essential for optimising biofilm fitness.

Purine biosynthesis was found to be important in the mature biofilm, through the essentiality of *purD*, *purE*, *purL* and *purH* in biofilms grown for 48 hours. Similar findings have previously been described in another transposon mutagenesis experiment in uropathogenic *E. coli* [32]. Inactivation of purine biosynthetic genes was also found to impair biofilm formation in *Bacillus cerus*, but this was thought to be due to reduced extracellular DNA in the biofilm matrix [24]. Extracellular DNA is thought to aid adhesion and has been found to be important in the biofilms of a wide range of bacterial species [23,25]. Our data suggest the importance is in the mature biofilm rather than initial adhesion. A relationship between both purine and pyrimidine biosynthesis and curli production in the biofilm has been reported [32,62,63]. More recently, curli biosynthesis in a *purL* mutant was reported to be abrogated through addition of inosine, which is involved in the *de novo* purine biosynthetic pathway for production of adenosine monophosphate (AMP) and guanine monophosphate (GMP) [64]. This suggests that nucleotide production itself, rather than the regulatory effects of the genes involved, affects curli biosynthesis, supporting one hypothesis that

disruption of the purine biosynthetic pathway may directly result in a reduction of cyclic-di-GMP, known to regulate biofilm formation and affect curli biosynthesis [32]. Quantification of intracellular c-di-GMP or further investigation of other c-di-GMP-dependent pathways in these mutants would uncover the relationship between these pathways and biofilm formation.

TraDIS-*Xpress* data suggested that expression of *fimB* and deletion of *fimE* was necessary for optimal biofilm fitness at all points in the biofilm, rather than just for initial attachment. Analysis of Δ*fimB* and Δ*fimE* deletion mutants found no significant change in biofilm biomass and reduced cell-cell aggregation. This finding supports previous work that reported increased fimbriae expression across all stages of biofilm development [28]. Previous work has observed a positive correlation between type 1 fimbriae and exopolysaccharide production in the mature biofilm [65], but the increase in biofilm biomass to support this was not seen in our study. These data suggest that for a population the ability to present cells both with and without fimbriae is beneficial for fitness throughout the life cycle.

Analysis of biofilms under flow conditions found that Δ*lrhA* and Δ*tomB* mutant biofilms had a similar appearance after 12 hours growth, could potentially indicate a similar role in the biofilm. The role of *lrhA* in motility regulation has been well documented [36,37,66], and expression of *tomB* has been seen to reduce motility through repression of *fliA* [67]. We found deletion of eiher *lrhA* or *tomB* resulted in reduced aggregation. Although Δ*lrhA* and Δ*tomB* deletion mutants shared many similar phenotypes, TraDIS-*Xpress* data predicted that *tomB* was beneficial and *lrhA* was detrimental to biofilm formation at 12, 24 and 48 hours. Previous studies on Δ*lrhA* mutant biofilms have reported increased adhesion, aggregation and biomass compared to the wild type [37]. This was not seen in our study, but we did find aggregation of the Δ*lrhA* mutant changed over time, with increased aggregates seen in biofilms grown in flow cells after 12 hours (figure 4). However, decreased aggregation was observed in planktonic cultures after 24 hours (figure 3). The initial benefit to biofilm formation resulting from the inactivation of *lrhA* predicted by the TraDIS-*Xpress* data agrees with the initial phenotype of the Δ*lrhA* biofilm in the flow cells. This may be due to reduced

induction of *fimE* by LrhA [37], thereby allowing expression of type 1 fimbriae to facilitate adhesion. We have already described how expression of both *fimB* and *fimE* is necessary for optimal fitness of the mature biofilm, and the effect of *lrhA* on biofilm formation correlates with these findings, with reduced aggregation in Δ*lrhA* biofilms after 24 hours (also seen in *fimB* and *fimE* mutants) and no microcolony formation under flow conditions at 24 and 48 hours. Therefore, the importance of *lrhA* to biofilm formation clearly appears to be time dependent, with the most important role in early events.

Studies on the effect of *tomB* on biofilm formation have focused on its toxin-antitoxin relationship with *hha*, which has been found to reduce expression of fimbrial subunit *fimA* and activate prophage lytic genes causing cell death [68]. Deletion of *hha* was found to reduce motility through *flhDC* and increase curli production through *csgD* [69]. We found no obvious benefit to biofilm fitness with insertional inactivation of *hha*, but this may not be visible in our data due to these mutants having a functional copy of *tomB*. The role of *tomB* in the biofilm has not previously been suggested, but we predict *tomB* is involved in regulation of adhesion in the early biofilm and matrix biosynthesis in the late biofilm.

The relationship between motility and biofilm formation is complex. Although it is widely understood that motility is crucial for biofilm formation [48,49], it is also true that motility and curli production have an inverse relationship, where *csgD* directly represses *fliE* [70] and induces c-di-GMP synthesis through *adrA*, which reduces motility through *ycgR* [71-73]. This is often referred to as a lifestyle 'switch', where biofilm matrix production represses motility for a motile-to-sessile lifestyle transition [74]. We found that although insertional inactivation of negative motility regulators *lrhA* and *hdfR* improved biofilm fitness according to the TraDIS-*Xpress* data, a Δ*hdfR* deletion mutant actually had reduced biofilm biomass, and deletion of either *lrhA* or *hdfR* impaired cell aggregation. Our data found an important role for structural flagella components in the mature biofilm. Previous work has suggested that flagella filaments are important for initial attachment and adhesion [75] , however we did not find this to be the case, with the expression of flagella filaments only appearing to

increase biofilm fitness in the mature biofilm grown for 48 hours. This is supported by previous work that found expression of flagella is important at all stages of the developing biofilm [28]. It appears that regulation of flagella and motility, rather than their fixed expression or absence, is important for optimal biofilm fitness.

Previous genome-wide screens on *E. coli* biofilms have identified some of the same genes as this study [26,30,32]. The TraDIS-*Xpress* technology used here makes for a more a powerful analysis of biofilm formation by predicting roles of changes in gene expression as well as essentiality and by analysing important genes over time. Differences between this work and previous studies may reflect biofilms being grown under different conditions on different surfaces, as these environmental factors greatly affect the pattern of gene expression and gene essentiality in the biofilm [76]. Analysis of more strains and species, grown on different abiotic and biotic surfaces under a range of environmental conditions may provide a wider list of essential genes for biofilm formation shared amongst a majority of human pathogens, as well as substrate-, condition- and species-specific genes and pathways for specific industrial, clinical and drug-development applications. As well as temporal changes in gene expression, spatial changes have been shown to affect biofilm development [77]. Integration of the spatial component into this model, to assay how gene expression throughout the biofilm over time affects biofilm fitness, would be the next logical step in furthering our understanding of biofilm development.

This study had revealed important time-specific roles for known and novel genes in biofilm formation. We reveal some pathways have a more important role in the mature biofilm than previously appreciated and identify genes with time dependent conditional essentiality within the biofilm. We also identify potential new candidate genes essential for biofilm formation, which could be targeted for novel anti-biofilm therapies. Further work using high-density transposon mutant libraries across time and in different conditions is likely to further our understanding of biofilm biology.

Methods

Transposon mutant library

The *E. coli* BW25113 transposon mutant library containing over 800,000 distinct mutants that was used in this study has recently described [33]. The transposon used to construct this library incorporates an outward-transcribing IPTG-inducible promoter.

Biofilm model conditions

The library was used to inoculate parallel cultures of 5 mL LB broth (without salt) with approximately 10^7 cells. Cultures were grown in 6-well plates containing 40 sterile 5 mm glass beads per well (Sigma, 18406). Two replicates were set up, with or without 1 mM IPTG. Plates were incubated at 30 ˚C with light shaking for 48 hours. After 12, 24, and 48 hours of incubation, 2 mL of planktonic sample was collected from each culture and 70 beads were taken to constitute the biofilm sample. Beads were washed twice in sterile PBS and vortexed in tubes containing PBS to resuspend cells from the biofilm. Both planktonic and biofilm samples were centrifuged at 2100 x g to form pellets for DNA extraction.

TraDIS-Xpress *Sequencing*

DNA was extracted from pellets following the protocol described in Trampari, et al. [78]. A customised sequencing library was prepared to identify transposon insertions. DNA was tagmented using a MuSeek DNA fragment library preparation kit (ThermoFisher) and customised Tn5-i5 and i7 primers were used in PCR for 28 cycles [33]. DNA fragments of 300-500 bp were size selected using AMPure beads (Beckman Coulter) and nucleotide sequences were generated using a NextSeq 500 and a NextSeq 500/550 High Output v2 kit (75 cycles) (Illumina).

Informatics

Fastq files were aligned to the *E. coli* BW25113 reference genome (CP009273) using the BioTraDIS (version 1.4.3) software suite [79] using SMALT (version 0.7.6). Insertion frequencies were determined per gene using tradis_gene_insert_sites within the BioTraDIS toolkit. Insertion

frequencies per gene for each replicate were plotted against each other to determine the experimental error between replicates as well as differences in insertion frequency between control and test conditions (supplementary figure 1). The tradis_comparison.R command (also part of the BioTraDIS toolkit) was also used to determine significant differences ($p < 0.05$, after correction for false discovery) in insertion frequencies per gene between control and test conditions. For all candidate loci, plot files generated by BioTraDIS were also examined manually in Artemis (version 17.0.1) [80] to confirm the results from these two approaches, as well as to identify regions where inserts were under differential selection but did not fall within coding regions of the genome.

Validation experiments

Knockout mutants for genes identified by TraDIS-Xpress data were sourced from the Keio collection [35] and tested for their biofilm-forming abilities. Crystal violet assays, used to assess biofilm biomass production, were undertaken by inoculating 10^4 of each mutant strain into 200 µL LB broth without salt in a 96-well plate. After 48 hours incubation at 30 ˚C, the culture was removed, wells were rinsed with water, and the residual biofilms were stained for 10 minutes with 200 µL 0.01% crystal violet. The plate was then rinsed with water to remove the stain and 200 µL 70% ethanol was added to the wells to solubilise the stained biofilm. The optical density (OD) was measured using a FLUOstar Omega plate reader (BMG Labtech) at 590 nM. Cell aggregation was measured by leaving overnight cultures (normalised to an $OD_{600\ nm}$ of 0.3) on an unagitated surface at room temperature. After 24 hours, the supernatant of each culture was removed by pipetting, diluted in PBS and measured in a plate reader at 600 nm. Biofilm matrix composition was investigated through spotting 10 µL of each mutant (representing 10^5 CFU) on agar supplemented with 40 µg/mL Congo red (Sigma, C6277) to examine curli production. Plates were incubated at 30˚C for 48 hours and photographed to compare mutant biofilm composition to the wild type. Adhesion and biofilm architecture were investigated under flow conditions for selected mutants using the Bioflux system. Flow cells were primed with LB broth without salt at 5 dyne/cm^2 and seeded with approximately 10^7 cells. The plate was left at room temperature for 2.5 hours to allow attachment, and subsequently incubated at 30 °C at a flow rate of 0.3 dyne/cm^2. After 12, 24 and

48 hours, biofilms were visualised with an inverted light microscope and representative images at 10x, 20x and 40x magnification were taken at three locations of the flow cell. Experiments were performed in duplicate.

Data availability

Sequence data supporting the analysis in this study has been deposited in ArrayExpress. Temporary accession number for review E-MTAB-9873.

Code availability statement

All software packages used are described in the methods. No bespoke code was used in this study.

Acknowledgements

ERH was supported by a studentship funded by the Quadram Institute Bioscience. The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); AKT, MY, JW, IGC and MAW were supported by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10349.

Author contributions

ERH designed and performed experiments, analysed the data and wrote the paper. AKT and IGC helped design experiments and wrote the paper. MY helped design experiments. JW heled analyse data and wrote the paper. MAW designed the experiments, analysed the data and wrote the paper.

Competing interests

The authors have no competing interests.

References

1 Berlanga, M. & Guerrero, R. Living together in biofilms: the microbial cell factory and its biotechnological implications. *Microbial Cell Factories* **15**, 165, doi:10.1186/s12934-016-0569-5 (2016).

2 Bjarnsholt, T. *et al.* Biofilm formation – What we can learn from recent developments. *J. Intern. Med.* **0**, doi:10.1111/joim.12782 (2018).

3 Gbejuade, H. O., Lovering, A. M. & Webb, J. C. The role of microbial biofilms in prosthetic joint infections. *Acta Orthop.* **86**, 147-158, doi:10.3109/17453674.2014.966290 (2015).

4 Davis, S. C., Martinez, L. & Kirsner, R. The diabetic foot: the importance of biofilms and wound bed preparation. *Curr. Diab. Rep.* **6**, 439-445 (2006).

5 Vestby, L. K., Gronseth, T., Simm, R. & Nesse, L. L. Bacterial Biofilm and its Role in the Pathogenesis of Disease. *Antibiotics (Basel, Switzerland)* **9**, doi:10.3390/antibiotics9020059 (2020).

6 Wang, H., Tay, M., Palmer, J. & Flint, S. Biofilm formation of *Yersinia enterocolitica* and its persistence following treatment with different sanitation agents. *Food Control* **73**, 433-437, doi:https://doi.org/10.1016/j.foodcont.2016.08.033 (2017).

7 Mah, T.-F. *et al.* A genetic basis for *Pseudomonas aeruginosa* biofilm antibiotic resistance. *Nature* **426**, 306, doi:10.1038/nature02122 https://www.nature.com/articles/nature02122#supplementary-information (2003).

8 Hoyle, B. D. & Costerton, J. W. Bacterial resistance to antibiotics: the role of biofilms. *Prog. Drug Res.* **37**, 91-105 (1991).

9 Flemming, H.-C. *et al.* Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* **14**, 563, doi:10.1038/nrmicro.2016.94 (2016).

10 Kostakioti, M., Hadjifrangiskou, M. & Hultgren, S. J. Bacterial biofilms: development, dispersal, and therapeutic strategies in the dawn of the postantibiotic era. *Cold Spring Harb. Perspect. Med.* **3**, a010306-a010306, doi:10.1101/cshperspect.a010306 (2013).

11 Flemming, H.-C. & Wingender, J. The biofilm matrix. *Nature Reviews Microbiology* **8**, 623-633, doi:10.1038/nrmicro2415 (2010).

12 Barnhart, M. M. & Chapman, M. R. Curli Biogenesis and Function. *Annu. Rev. Microbiol.* **60**, 131-147, doi:10.1146/annurev.micro.60.080805.142106 (2006).

13 Serra, D. O. & Hengge, R. in *Extracellular Sugar-Based Biopolymers Matrices* (eds Ephraim Cohen & Hans Merzendorfer) 355-392 (Springer International Publishing, 2019).

14 Jubelin, G. *et al.* CpxR/OmpR Interplay Regulates Curli Gene Expression in Response to Osmolarity in *Escherichia coli*. *J. Bacteriol.* **187**, 2038-2049, doi:10.1128/jb.187.6.2038-2049.2005 (2005).

15 Vidal, O. *et al.* Isolation of an *Escherichia coli* K-12 Mutant Strain Able To Form Biofilms on Inert Surfaces: Involvement of a New *ompR* Allele That Increases Curli Expression. *J. Bacteriol.* **180**, 2442-2449 (1998).

16 Dorel, C., Vidal, O., Prigent-Combaret, C., Vallet, I. & Lejeune, P. Involvement of the Cpx signal transduction pathway of *E. coli* in biofilm formation. *FEMS Microbiol. Lett.* **178**, 169-175, doi:10.1111/j.1574-6968.1999.tb13774.x (1999).

17 Otto, K. & Silhavy, T. J. Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proceedings of the National Academy of Sciences* **99**, 2287-2292, doi:10.1073/pnas.042521699 (2002).

18 Adams, J. L. & McLean, R. J. Impact of *rpoS* deletion on *Escherichia coli* biofilms. *Appl. Environ. Microbiol.* **65**, 4285-4287 (1999).

501 19 Corona-Izquierdo, F. P. & Membrillo-Hernandez, J. A mutation in *rpoS* enhances biofilm
502 formation in *Escherichia coli* during exponential phase of growth. *FEMS Microbiol. Lett.*
503 **211**, 105-110 (2002).
504 20 Gerstel, U., Park, C. & Römling, U. Complex regulation of *csgD* promoter activity by global
505 regulatory proteins. *Mol. Microbiol.* **49**, 639-654, doi:doi:10.1046/j.1365-
506 2958.2003.03594.x (2003).
507 21 Gerstel, U. & Romling, U. The *csgD* promoter, a control unit for biofilm formation in
508 *Salmonella* Typhimurium. *Res. Microbiol.* **154**, 659-667, doi:10.1016/j.resmic.2003.08.005
509 (2003).
510 22 Amores, G. R., de las Heras, A., Sanches-Medeiros, A., Elfick, A. & Silva-Rocha, R. Systematic
511 identification of novel regulatory interactions controlling biofilm formation in the
512 bacterium *Escherichia coli*. *Sci. Rep.* **7**, 16768, doi:10.1038/s41598-017-17114-6 (2017).
513 23 Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C. & Mattick, J. S. Extracellular DNA required
514 for bacterial biofilm formation. *Science* **295**, 1487, doi:10.1126/science.295.5559.1487
515 (2002).
516 24 Vilain, S., Pretorius, J. M., Theron, J. & Brözel, V. S. DNA as an Adhesin: *Bacillus cereus*
517 Requires Extracellular DNA To Form Biofilms. *Appl. Environ. Microbiol.* **75**, 2861,
518 doi:10.1128/AEM.01317-08 (2009).
519 25 Tetz, G. V., Artemenko, N. K. & Tetz, V. V. Effect of DNase and Antibiotics on Biofilm
520 Characteristics. *Antimicrob. Agents Chemother.* **53**, 1204, doi:10.1128/AAC.00471-08
521 (2009).
522 26 Niba, E. T. E., Naka, Y., Nagase, M., Mori, H. & Kitakawa, M. A Genome-wide Approach to
523 Identify the Genes Involved in Biofilm Formation in *E. coli*. *DNA Res.* **14**, 237-246,
524 doi:10.1093/dnares/dsm024 (2008).
525 27 Aedo, S. J., Ma, H. R. & Brynildsen, M. P. Checks and Balances with Use of the Keio
526 Collection for Phenotype Testing. *Methods Mol. Biol.* **1927**, 125-138, doi:10.1007/978-1-
527 4939-9142-6_9 (2019).
528 28 Domka, J., Lee, J., Bansal, T. & Wood, T. K. Temporal gene-expression in *Escherichia coli* K-
529 12 biofilms. *Environ. Microbiol.* **9**, 332-346, doi:10.1111/j.1462-2920.2006.01143.x (2007).
530 29 Schembri, M. A., Kjærgaard, K. & Klemm, P. Global gene expression in *Escherichia coli*
531 biofilms. *Mol. Microbiol.* **48**, 253-267, doi:doi:10.1046/j.1365-2958.2003.03432.x (2003).
532 30 Puttamreddy, S., Cornick, N. A. & Minion, F. C. Genome-Wide Transposon Mutagenesis
533 Reveals a Role for pO157 Genes in Biofilm Development in *Escherichia coli* O157:H7
534 EDL933. *Infect. Immun.* **78**, 2377, doi:10.1128/IAI.00156-10 (2010).
535 31 Goh, K. G. K. *et al.* Genome-Wide Discovery of Genes Required for Capsule Production by
536 Uropathogenic *Escherichia coli*. *mBio* **8**, e01558-01517, doi:10.1128/mBio.01558-17 (2017).
537 32 Nhu, N. T. K. *et al.* Discovery of New Genes Involved in Curli Production by a Uropathogenic
538 *Escherichia coli* Strain from the Highly Virulent O45:K1:H7 Lineage. *mBio* **9**, e01462-01418,
539 doi:10.1128/mBio.01462-18 (2018).
540 33 Yasir, M. *et al.* TraDIS-Xpress: a high-resolution whole-genome assay identifies novel
541 mechanisms of triclosan action and resistance. *Genome Res*, doi:10.1101/gr.254391.119
542 (2020).
543 34 Klemm, P. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1
544 fimbriae in *Escherichia coli*. *The EMBO Journal* **5**, 1389-1393, doi:10.1002/j.1460-
545 2075.1986.tb04372.x (1986).
546 35 Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout
547 mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008-2006.0008,
548 doi:10.1038/msb4100050 (2006).

549 36 Lehnen, D. *et al.* LrhA as a new transcriptional key regulator of flagella, motility and
550 chemotaxis genes in *Escherichia coli*. *Mol. Microbiol.* **45**, 521-532, doi:10.1046/j.1365-
551 2958.2002.03032.x (2002).
552 37 Blumer, C. *et al.* Regulation of type 1 fimbriae synthesis and biofilm formation by the
553 transcriptional regulator LrhA of *Escherichia coli*. *Microbiology* **151**, 3287-3298,
554 doi:10.1099/mic.0.28098-0 (2005).
555 38 Ko, M. & Park, C. H-NS-Dependent regulation of flagellar synthesis is mediated by a LysR
556 family protein. *J. Bacteriol.* **182**, 4670-4672, doi:10.1128/jb.182.16.4670-4672.2000 (2000).
557 39 Alekshun, M. N. & Levy, S. B. Alteration of the Repressor Activity of MarR, the Negative
558 Regulator of the *Escherichia coli marRAB* Locus, by Multiple Chemicals In Vitro. *J. Bacteriol.*
559 **181**, 4669-4672 (1999).
560 40 Herzberg, M., Kaye, I. K., Peti, W. & Wood, T. K. YdgG (TqsA) controls biofilm formation in
561 *Escherichia coli* K-12 through autoinducer 2 transport. *J. Bacteriol.* **188**, 587-598,
562 doi:10.1128/jb.188.2.587-598.2006 (2006).
563 41 Holden, E. R. & Webber, M. A. MarA, RamA and SoxS as mediators of the Stress Response:
564 Survival at a Cost. *Front. Microbiol.* **11**, 828 (2020).
565 42 Kettles, R. A. *et al.* The *Escherichia coli* MarA protein regulates the *ycgZ-ymgABC* operon to
566 inhibit biofilm formation. *Mol. Microbiol.* **0**, doi:10.1111/mmi.14386 (2019).
567 43 Szyf, M. *et al.* DNA methylation pattern is determined by the intracellular level of the
568 methylase. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 3278-3282, doi:10.1073/pnas.81.11.3278
569 (1984).
570 44 Valens, M., Thiel, A. & Boccard, F. The MaoP/*maoS* Site-Specific System Organizes the Ori
571 Region of the *E. coli* Chromosome into a Macrodomain. *PLoS Genet.* **12**, e1006309-
572 e1006309, doi:10.1371/journal.pgen.1006309 (2016).
573 45 Kim, T., Lee, J. & Kim, K.-s. *Escherichia coli* YmdB regulates biofilm formation independently
574 of its role as an RNase III modulator. *BMC Microbiol.* **13**, 266-266, doi:10.1186/1471-2180-
575 13-266 (2013).
576 46 Evans, Margery L. *et al.* The Bacterial Curli System Possesses a Potent and Selective
577 Inhibitor of Amyloid Formation. *Mol. Cell* **57**, 445-455,
578 doi:https://doi.org/10.1016/j.molcel.2014.12.025 (2015).
579 47 Zhang, Y., Morar, M. & Ealick, S. E. Structural biology of the purine biosynthetic pathway.
580 *Cell. Mol. Life Sci.* **65**, 3699-3724, doi:10.1007/s00018-008-8295-8 (2008).
581 48 Pratt, L. A. & Kolter, R. Genetic analysis of *Escherichia coli* biofilm formation: roles of
582 flagella, motility, chemotaxis and type I pili. *Mol. Microbiol.* **30**, 285-293,
583 doi:10.1046/j.1365-2958.1998.01061.x (1998).
584 49 Wang, F. *et al.* Flagellar Motility Is Critical for *Salmonella enterica* Serovar Typhimurium
585 Biofilm Development. *Front. Microbiol.* **11**, doi:10.3389/fmicb.2020.01695 (2020).
586 50 Shimada, T., Bridier, A., Briandet, R. & Ishihama, A. Novel roles of LeuO in transcription
587 regulation of *E. coli* genome: antagonistic interplay with the universal silencer H-NS. *Mol.*
588 *Microbiol.* **82**, 378-397, doi:10.1111/j.1365-2958.2011.07818.x (2011).
589 51 Kroner, G. M., Wolfe, M. B. & Freddolino, P. L. *Escherichia coli* Lrp Regulates One-Third of
590 the Genome via Direct, Cooperative, and Indirect Routes. *J. Bacteriol.* **201**, e00411-00418,
591 doi:10.1128/jb.00411-18 (2019).
592 52 Tramonti, A., De Canio, M. & De Biase, D. GadX/GadW-dependent regulation of the
593 *Escherichia coli* acid fitness island: transcriptional control at the *gadY–gadW* divergent
594 promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol.*
595 *Microbiol.* **70**, 965-982, doi:10.1111/j.1365-2958.2008.06458.x (2008).

596 53 Sukharev, S. I., Blount, P., Martinac, B., Blattner, F. R. & Kung, C. A large-conductance
597 mechanosensitive channel in *E. coli* encoded by *mscL* alone. *Nature* **368**, 265-268,
598 doi:10.1038/368265a0 (1994).
599 54 Morona, R., Manning, P. A. & Reeves, P. Identification and characterization of the TolC
600 protein, an outer membrane protein from *Escherichia coli*. *J. Bacteriol.* **153**, 693-699
601 (1983).
602 55 Cai, S. J. & Inouye, M. EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *J.*
603 *Biol. Chem.* **277**, 24155-24161, doi:10.1074/jbc.M110715200 (2002).
604 56 Marteyn, B. S. *et al.* ZapE is a novel cell division protein interacting with FtsZ and
605 modulating the Z-ring dynamics. *mBio* **5**, e00022-00014, doi:10.1128/mBio.00022-14
606 (2014).
607 57 Hamma, T. & Ferré-D'Amaré, A. R. Pseudouridine Synthases. *Chem. Biol.* **13**, 1125-1135,
608 doi:https://doi.org/10.1016/j.chembiol.2006.09.009 (2006).
609 58 Tsui, H. C., Arps, P. J., Connolly, D. M. & Winkler, M. E. Absence of *hisT*-mediated tRNA
610 pseudouridylation results in a uracil requirement that interferes with *Escherichia coli* K-12
611 cell division. *J. Bacteriol.* **173**, 7395-7400, doi:10.1128/jb.173.22.7395-7400.1991 (1991).
612 59 Bringer, M. A., Rolhion, N., Glasser, A. L. & Darfeuille-Michaud, A. The oxidoreductase DsbA
613 plays a key role in the ability of the Crohn's disease-associated adherent-invasive
614 *Escherichia coli* strain LF82 to resist macrophage killing. *J. Bacteriol.* **189**, 4860-4871,
615 doi:10.1128/jb.00233-07 (2007).
616 60 Magnusson, L. U., Gummesson, B., Joksimović, P., Farewell, A. & Nyström, T. Identical,
617 Independent, and Opposing Roles of ppGpp and DksA in *Escherichia coli*. *J. Bacteriol.* **189**,
618 5193-5202, doi:10.1128/jb.00330-07 (2007).
619 61 Anwar, N., Rouf, S. F., Römling, U. & Rhen, M. Modulation of Biofilm-Formation in
620 *Salmonella enterica* Serovar Typhimurium by the Periplasmic DsbA/DsbB Oxidoreductase
621 System Requires the GGDEF-EAL Domain Protein STM3615. *PLoS One* **9**, e106095,
622 doi:10.1371/journal.pone.0106095 (2014).
623 62 Smith, D. R. *et al.* The Production of Curli Amyloid Fibers Is Deeply Integrated into the
624 Biology of *Escherichia coli*. *Biomolecules* **7**, 75, doi:10.3390/biom7040075 (2017).
625 63 Garavaglia, M., Rossi, E. & Landini, P. The pyrimidine nucleotide biosynthetic pathway
626 modulates production of biofilm determinants in *Escherichia coli*. *PLoS One* **7**, e31252-
627 e31252, doi:10.1371/journal.pone.0031252 (2012).
628 64 Cepas, V. *et al.* Transposon Insertion in the *purL* Gene Induces Biofilm Depletion in
629 *Escherichia coli* ATCC 25922. *Pathogens (Basel, Switzerland)* **9**,
630 doi:10.3390/pathogens9090774 (2020).
631 65 Rodrigues, D. F. & Elimelech, M. Role of type 1 fimbriae and mannose in the development
632 of *Escherichia coli* K12 biofilm: from initial cell adhesion to biofilm formation. *Biofouling* **25**,
633 401-411, doi:10.1080/08927010902833443 (2009).
634 66 Li, S. *et al.* An Osmoregulatory Mechanism Operating through OmpR and LrhA Controls the
635 Motile-Sessile Switch in the Plant Growth-Promoting Bacterium *Pantoea alhagi*. *Appl.*
636 *Environ. Microbiol.* **85**, e00077-00019, doi:10.1128/aem.00077-19 (2019).
637 67 Barrios, A. F., Zuo, R., Ren, D. & Wood, T. K. Hha, YbaJ, and OmpA regulate *Escherichia coli*
638 K12 biofilm formation and conjugation plasmids abolish motility. *Biotechnol. Bioeng.* **93**,
639 188-200, doi:10.1002/bit.20681 (2006).
640 68 Garcia-Contreras, R., Zhang, X. S., Kim, Y. & Wood, T. K. Protein translation and cell death:
641 the role of rare tRNAs in biofilm formation and in activating dormant phage killer genes.
642 *PLoS One* **3**, e2394, doi:10.1371/journal.pone.0002394 (2008).

69 Sharma, V. K. & Bearson, B. L. Hha controls *Escherichia coli* O157:H7 biofilm formation by differential regulation of global transcriptional regulators FlhDC and CsgD. *Appl. Environ. Microbiol.* **79**, 2384-2396, doi:10.1128/AEM.02998-12 (2013).
70 Ogasawara, H., Yamamoto, K. & Ishihama, A. Role of the Biofilm Master Regulator CsgD in Cross-Regulation between Biofilm Formation and Flagellar Synthesis. *J. Bacteriol.* **193**, 2587-2597, doi:10.1128/jb.01468-10 (2011).
71 Fang, X. & Gomelsky, M. A post-translational, c-di-GMP-dependent mechanism regulating flagellar motility. *Mol. Microbiol.* **76**, 1295-1305, doi:10.1111/j.1365-2958.2010.07179.x (2010).
72 Nieto, V. *et al.* Under Elevated c-di-GMP in *Escherichia coli*, YcgR Alters Flagellar Motor Bias and Speed Sequentially, with Additional Negative Control of the Flagellar Regulon via the Adaptor Protein RssB. *J. Bacteriol.* **202**, e00578-00519, doi:10.1128/JB.00578-19 (2019).
73 Paul, K., Nieto, V., Carlquist, W. C., Blair, D. F. & Harshey, R. M. The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. *Mol. Cell* **38**, 128-139, doi:10.1016/j.molcel.2010.03.001 (2010).
74 Pesavento, C. *et al.* Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. *Genes Dev.* **22**, 2434-2446, doi:10.1101/gad.475808 (2008).
75 Wood, T. K., González Barrios, A. F., Herzberg, M. & Lee, J. Motility influences biofilm architecture in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **72**, 361-367, doi:10.1007/s00253-005-0263-8 (2006).
76 Prouty, A. M. & Gunn, J. S. Comparative analysis of *Salmonella enterica* serovar Typhimurium biofilm formation on gallstones and on glass. *Infect. Immun.* **71**, 7154-7158 (2003).
77 Samanta, P., Clark, E. R., Knutson, K., Horne, S. M. & Prüß, B. M. OmpR and RcsB abolish temporal and spatial changes in expression of *flhD* in *Escherichia coli* Biofilm. *BMC Microbiol.* **13**, 182, doi:10.1186/1471-2180-13-182 (2013).
78 Trampari, E. *et al.* Antibiotics select for novel pathways of resistance in biofilms. *bioRxiv*, 605212, doi:10.1101/605212 (2019).
79 Barquist, L. *et al.* The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109-1111, doi:10.1093/bioinformatics/btw022 (2016).
80 Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464-469, doi:10.1093/bioinformatics/btr703 (2011).

Figures

Figure 1: Genes involved in biofilm formation over time in *E. coli*. Expression of genes in green were found to be beneficial for, and those in red were found to be detrimental to, biofilm formation.

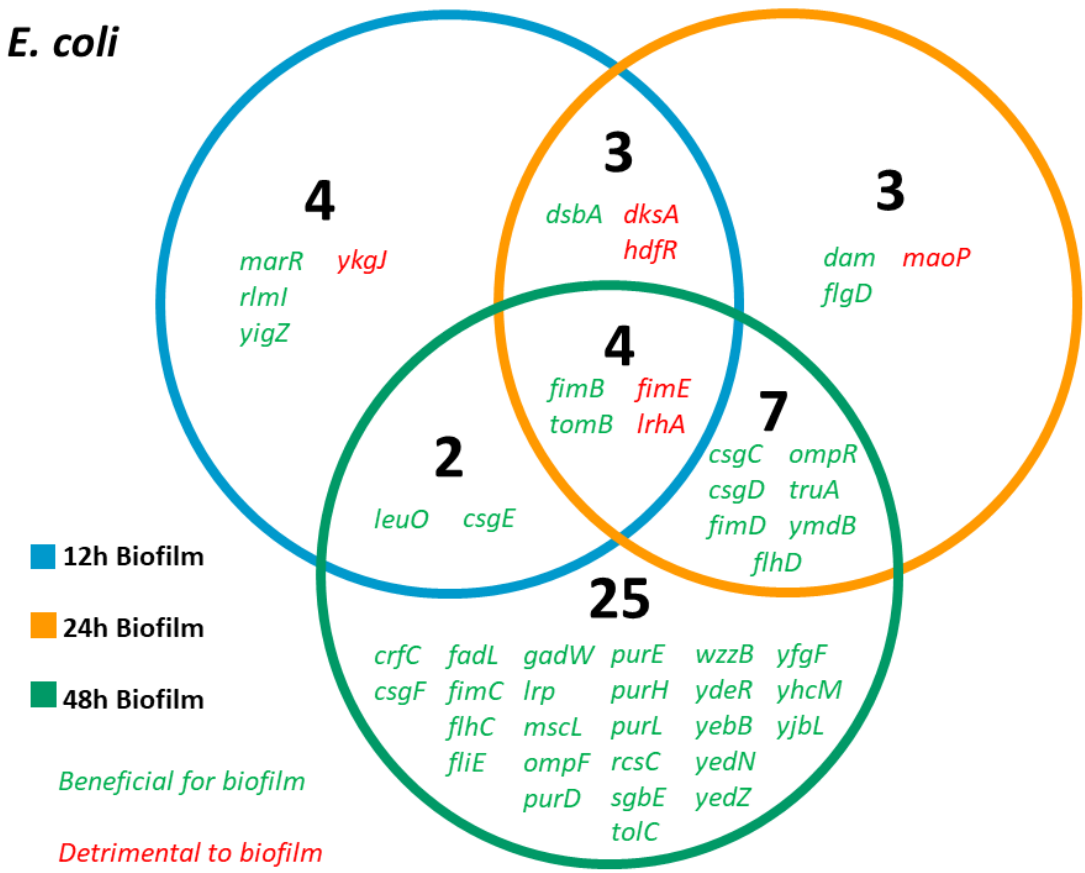


Figure 2: Mapped reads from TraDIS-Xpress data, plotted with BioTraDIS in Artemis. **a)** Insertion sites in and around *fimB* and *fimE* in planktonic and biofilm conditions after 12 and 48 hours growth. **b)** Insertion sites in and around *hdfR*, *lrhA* and *tomB* in planktonic and biofilm conditions after 24 hours growth. **c)** Insertion sites in and around *dksA* and *dsbA* in planktonic and biofilm conditions after 12 and 48 hours growth. For all plot files, one of two independent replicates is shown and the transposon-located promoter is induced with 1mM IPTG in all conditions shown.

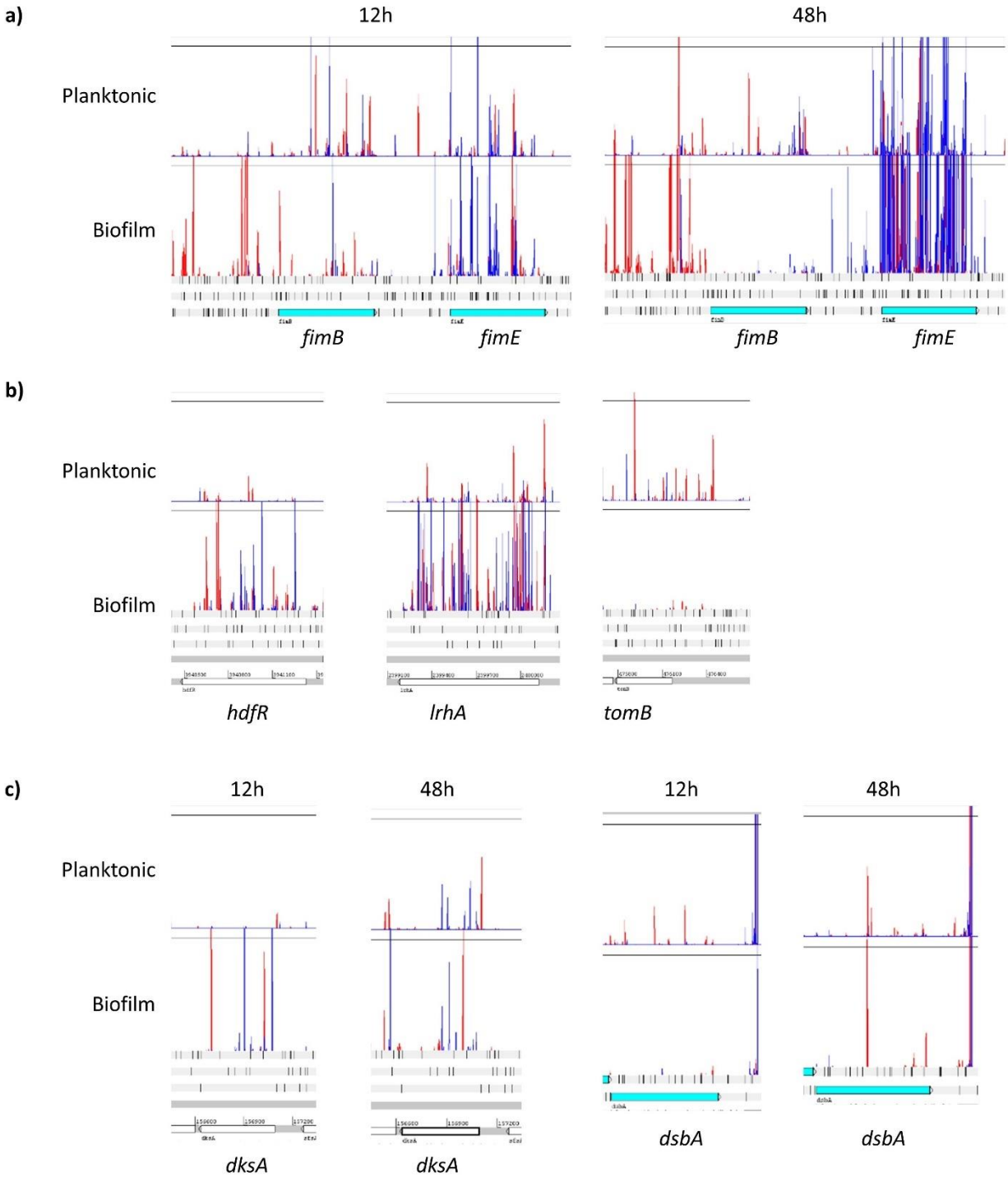
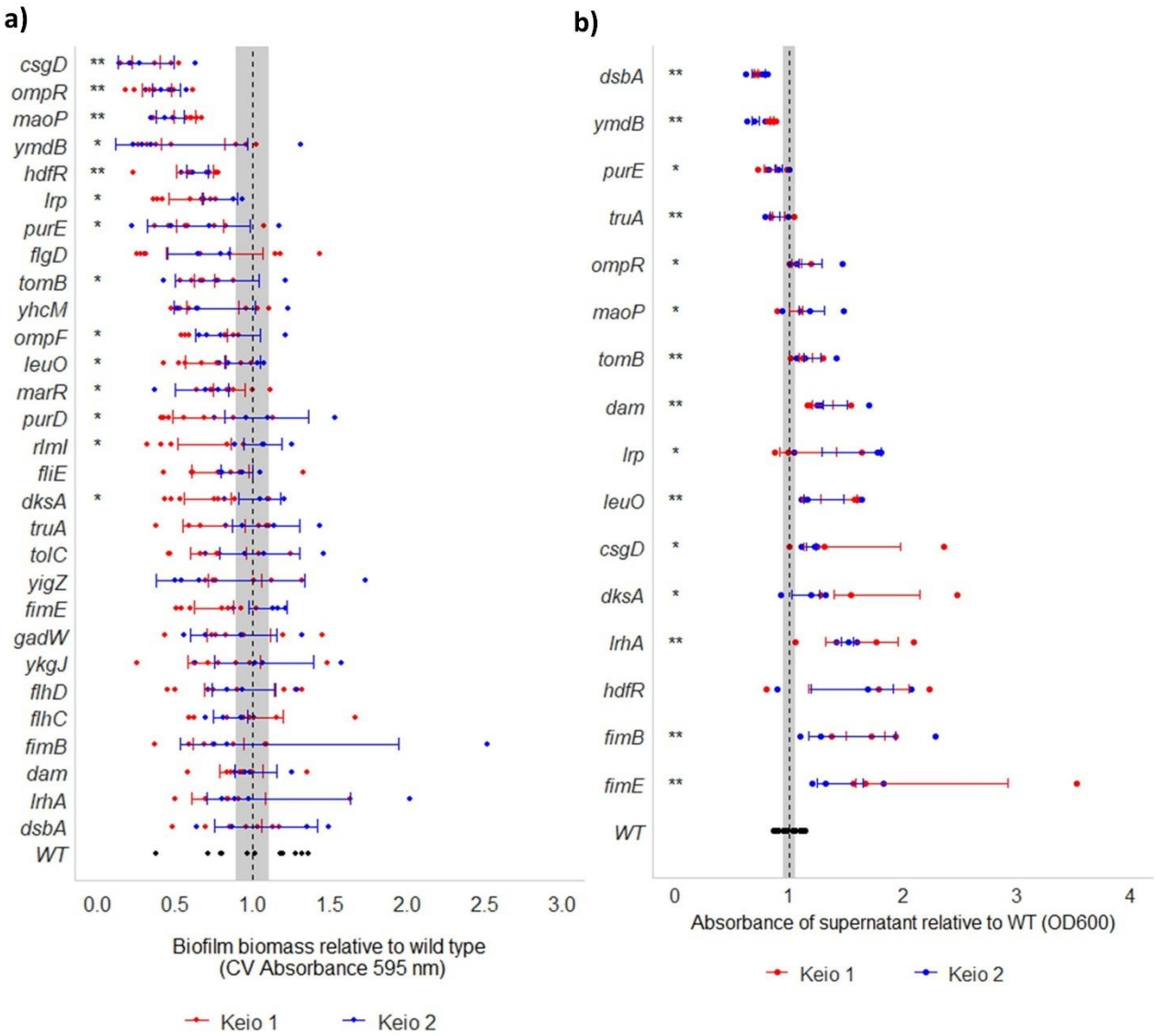
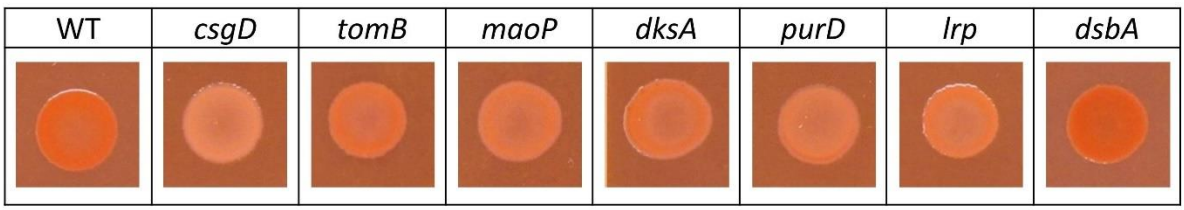


Figure 3: Phenotypic validation of selected genes involved in biofilm formation. **a**) Biofilm biomass of single knockout mutants relative to wild type *E. coli* BW25113, measured by crystal violet staining. Two biological and a minimum of two technical replicates were performed for each mutant. **b**) Cell aggregation of single knockout mutants relative to wild type *E. coli* BW25113, measured by $OD_{600\ nm}$ of the supernatant of unagitated cultures. Points show the ODs of three independent replicates. For both graphs, red points/bars show data from the first KEIO collection mutant of each gene, and blue points/bars show data from the second mutant. Error bars show 95% confidence intervals and the shaded area shows the 95% confidence interval of the wild type. Single asterisks (*) represent a significant difference between one Keio mutant copy and the wild type, and double asterisks (**) denote a significant difference between both Keio mutant copies and the wild type (Welch's *t*-test, $p < 0.05$). **c**) Colonies grown on agar supplemented with Congo red to compare curli biosynthesis between single knockout mutants and the wild type. Images are representative of 2 biological and 2 technical replicates.


a)
csgD
ompR
maoP
ymdB
hdfR
lrp
purE
flgD
tomB
yhcM
ompF
leuO
marR
purD
rlmI
fliE
dksA
truA
tolC
yigZ
fimE
gadW
ykgJ
flhD
flhC
fimB
dam
lrhA
dsbA
WT
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Biofilm biomass relative to wild type
(CV Absorbance 595 nm)
Keio 1
Keio 2
b)
dsbA
ymdB
purE
truA
ompR
maoP
tomB
dam
lrp
leuO
csgD
dksA
lrhA
hdfR
fimB
fimE
WT
0
1
2
3
4
Absorbance of supernatant relative to WT (OD600)
Keio 1
Keio 2

c)


WT
csgD
tomB
maoP
dksA
purD
lrp
dsbA

708

Figure 4: Biofilm formation of single knockout mutants on glass analysed under flow conditions after 12 and 48 hours growth. 10x Magnification. Images are representative of two independent replicates. Scale bar indicates 5 µm.

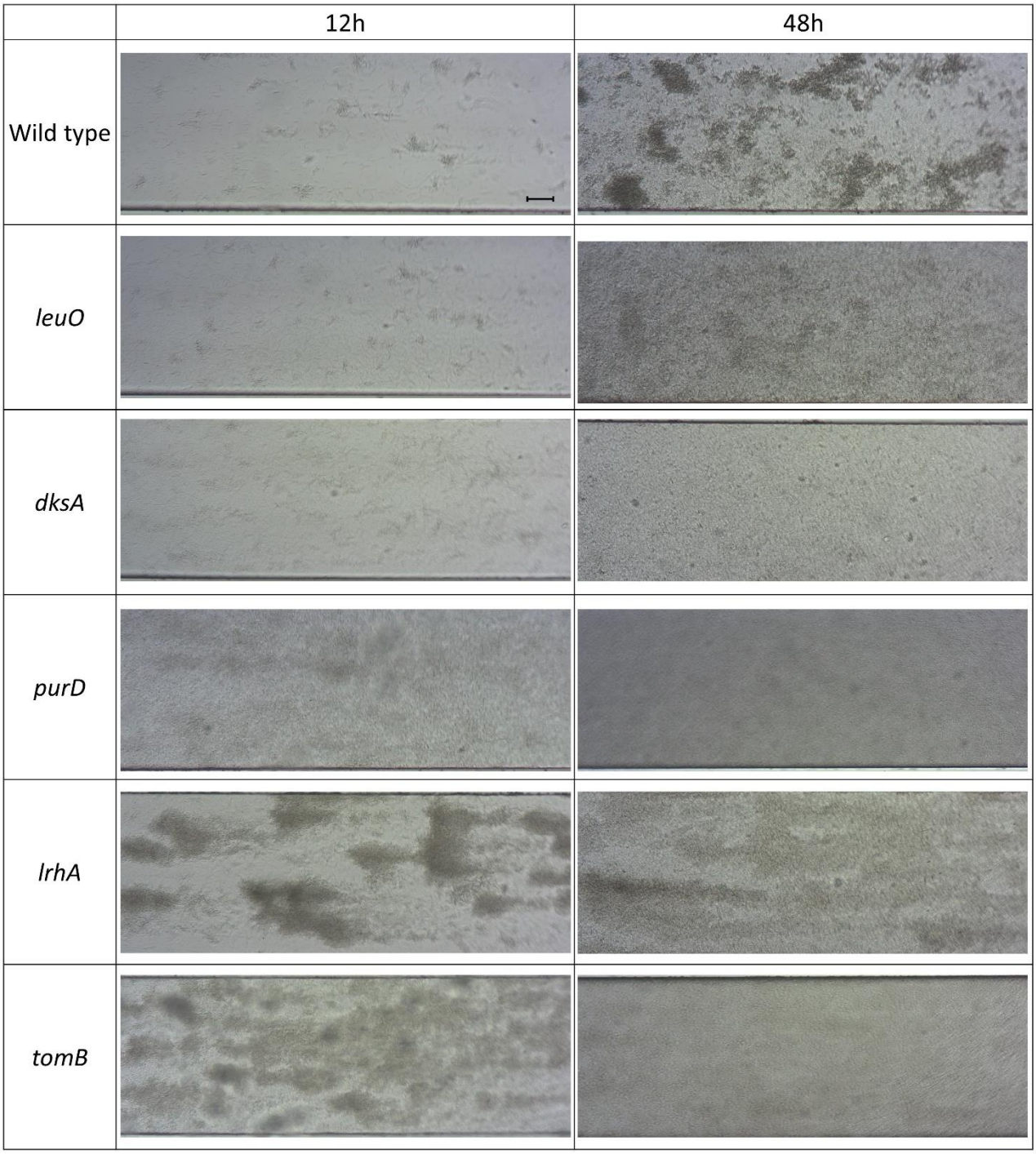
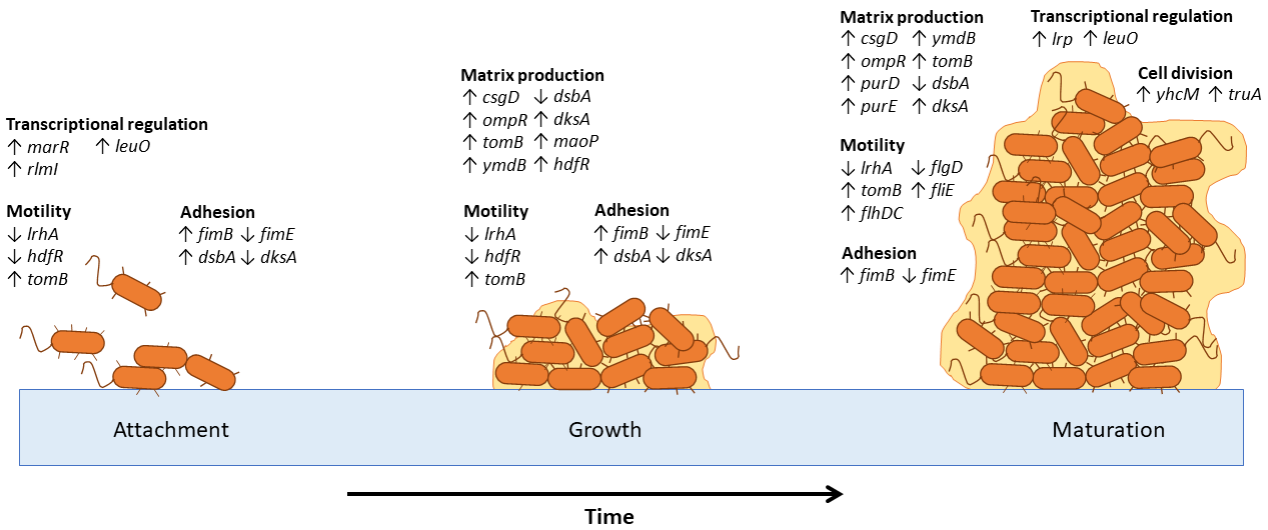



Figure 5. Summary of genes important for biofilm formation by *E. coli* at different stages of development.