Metabolism of L -arabinose converges with virulence regulation to promote enteric pathogen fitness

Curtis Cottam[a], Rhys T White[bc], Lauren C Beck[d], Christopher J Stewart[d], Scott A Beatson[c], Elisabeth C Lowe[a], Rhys Grinter[e] and James PR Connolly[a*]

[a]Newcastle University Biosciences Institute, Newcastle University, Newcastle-upon-Tyne, NE2 4HH, United Kingdom

[b]Institute of Environmental Science and Research, Wellington, New Zealand

[c]Australian Infectious Diseases Research Centre and School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia

[d]Newcastle University Translation and Clinical Research Institute, Newcastle University, Newcastle-upon-Tyne, NE2 4HH, United Kingdom

[e]Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

*Correspondence:

James.Connolly2@newcastle.ac.uk Tel: 0191 208 8866

Key words – *Escherichia coli, Citrobacter rodentium,* EHEC, Type 3 Secretion System, L-arabinose, colonisation resistance

Abstract:

Virulence and metabolism are often interlinked to control the expression of essential colonisation factors in response to host-associated signals. Here, we identified a novel transporter of the dietary monosaccharide L-arabinose that is widely encoded by the zoonotic pathogen enterohaemorrhagic *Escherichia coli* (EHEC), required for full competitive fitness in the mouse gut and highly expressed during human infection. Accordingly, we discovered that L-arabinose induces expression of the EHEC type 3 secretion system, enhancing its ability to attach to host cells, and that the underlying mechanism is dependent on products of its catabolism rather than the sensing of L-arabinose as a signal. Finally, using the murine pathogen *Citrobacter rodentium,* we show that L-arabinose metabolism provides a fitness benefit during enteric infection via virulence factor regulation, as opposed to supporting pathogen growth. This study describes an intrinsic mechanism of integrating central sugar metabolism with virulence regulation and highlights the unexpected impact that nutrient utilisation can have in enteric pathogens.

Introduction

The mammalian gastrointestinal tract poses a formidable barrier to infection by foreign pathogens. Invaders must sidestep a combination of intrinsic host defences as well as overcome colonisation resistance by the native gut microbiota[1]. Accordingly, bacterial pathogens have evolved many unique strategies to effectively compete with the microbiota and cause infection within a favourable host-niche[2]. This includes a combination of pathogen specific virulence mechanisms and metabolic adaptations that increase within-host fitness. Importantly, virulence and fitness factor regulation often coincide and are dynamically controlled in response to the host environment to maximise competitiveness therein[3–5].

Enterohaemorrhagic *Escherichia coli* (EHEC) is a zoonotic pathogen that causes severe diarrhoeal illness in humans and, in extreme cases, renal failure[6,7]. EHEC is a member of the attaching and effacing (A/E) family of pathogens, which intimately colonise the colonic epithelium forming characteristic pedestal-like lesions on the surface of host cells[8,9]. A/E pathogenesis is defined by the activity of a type 3 secretion system (T3SS) encoded on a ~35 kb horizontally acquired island known as the locus of enterocyte effacement (LEE)[10–12]. This T3SS translocates more than 30 effector proteins - encoded on the LEE and several additional horizontally acquired elements, termed O-islands (OIs) - that collectively subvert host cell function[13–16]. The murine pathogen *Citrobacter rodentium* also encodes the LEE and has been adopted as the relevant surrogate model to study EHEC pathogenesis *in vivo* due to its dependency on the T3SS, comparable pathology and lack of requirement to pre-treat mice with antibiotics[17–19].

The LEE-encoded T3SS is essential for overcoming microbiota-dependent colonisation resistance[20,21]. Epithelial attachment does not depend on a tissue receptor interaction. Instead, the LEE is uniquely controlled in response to a multitude of host and microbiota derived signals such as sugars, amino acids, short- and long-chain fatty acids and hormones that are integrated into a regulatory network ensuring correct spatial deployment of the T3SS[3–5]. For example, D-glucose represses the LEE whereas gluconeogenic substrates encountered at the epithelial surface (such as succinate) enhance its expression[22,23]; L-arginine found in abundance in the gut is directly sensed to activate the LEE, whereas the amino acid D-serine, abundant in the urinary tract, represses the LEE thus restricting EHEC to the gut niche[24–26]. A/E pathogenesis therefore hinges on nutrient availability, by promoting competition with the commensal microbiota as well as regulating essential virulence factors.

E. coli as a species relies on cross-feeding of microbiota liberated mono- and disaccharide sugars as it cannot degrade complex dietary polysaccharides[27]. However, EHEC has evolved to utilise a unique hierarchy of sugars compared to commensal *E. coli,* therefore implying a strategy to limit competition with similar members of the microbiota for critical resources[28]. L-arabinose is one such resource and highly abundant in nature as a major constituent of plant cell walls[29]. *E. coli* transports L-arabinose into the cell via the H^+ symporter AraE and the ATP-binding cassette (ABC) transporter AraFGH. The transcription factor AraC then senses cytosolic L-arabinose and upregulates both transporters as well as the AraBAD enzymes essential for its metabolism, a mechanism enhanced by the presence of cyclic AMP that increases during D-glucose depletion[30]. Here, we show that L-arabinose utilisation is exploited by A/E pathogens during infection. Surprisingly, this occurs independently of its role as a source of nutrition. Instead, this advantage is achieved via a mechanism that links metabolism

and virulence regulation. Transcriptome analysis revealed that L-arabinose upregulates expression of the LEE-encoded T3SS. Importantly, this relies on the uptake and breakdown of L-arabinose in the cell, acting as regulatory trigger as opposed to supporting pathogen growth *in vivo*. Finally, we also describe the identification of a novel L-arabinose uptake system that is widely encoded by T3SS-encoding EHEC strains in nature. We hypothesise that this system gives EHEC an enhanced capacity to scavenge L-arabinose in the gut, which therefore provides a fitness advantage via the convergence of virulence regulation and metabolism.

Results

A novel accessory L-arabinose uptake system is widely encoded by EHEC isolates in nature

While analysing our previously determined *C. rodentium in vivo* transcriptome, we noticed that genes related to monosaccharide uptake and metabolism were some of the most highly upregulated during infection[21]. Among these was ROD_24811, an uncharacterised gene displaying ~20-fold induction in the colon and predicted to encode a periplasmic binding protein of a simple sugar ABC transporter. A search for homologues in EHEC did not find an exact match but instead identified a similar uncharacterised 4.4Kb ABC-transporter locus (>60 % nucleotide identity) that appeared to be EHEC-specific, displaying a unique genetic context, being located on OI-island 17. We therefore focused our efforts on the role of this system given its relevance to EHEC and potentially human infection. This locus was predicted to encode a periplasmic binding protein (Z0415), an ATPase (Z0416-7) and two permease subunits comprised of α-helices (Z0418/Z0419), characteristic of Type II ABC transporters[31]. *In silico* modelling of Z0415-9 using AlphaFold2 supported this, by displaying the expected modular structure of an inner membrane ABC transporter (Fig. 1a). InterPro analysis of the associated amino acid sequences identified domain signatures related to pentose monosaccharide substrate specificity and comparison with known *E. coli* systems supported that Z0415 clustered closely with monosaccharide ABC transporters (Extended Data Fig. 1). Presence/absence analysis of Z0415-9 carriage amongst 949 representative *E. coli* genomes revealed that the locus is not completely conserved across the species phylogeny (Fig. 1b). For instance, Z0415-19 is carried predominantly amongst isolates from phylogroups B1 and E (397/445 strains), comprised largely of EHEC strains, whereas phylogroups A and B2 (largely commensal and extraintestinal isolates, respectively) largely lack this system. Strikingly, there was a significant correlation between Z0415-9 and LEE carriage (Odds ratio = 35.9; $P < 0.001$),

with the converse scenario (LEE positive, Z0415-9 negative) being an incredibly rare event (8/948 strains) (Fig. 1c). This suggests that there may be conflating evolutionary pressures for LEE-encoding EHEC strains to acquire Z0415-9, implicating its associated function in pentose sugar scavenging as being potentially beneficial for EHEC infection.

To assess the function of Z0415-9, we generated a transcriptional reporter (EHEC transformed with pMK1/*lux*-$P_{Z0415\text{-}9}$) and cultured this in MEM-HEPES supplemented with either D-xylose, D-ribose or L-arabinose. Z0415-9 promoter activity was induced in a concentration-dependent manner exclusively in response to L-arabinose and we validated this result using RT-qPCR (Fig. 2a; Extended Data Fig. 2a/b). Additionally, we confirmed that Z0415-9 induction occurs exclusively in response to L-arabinose and not D-arabinose, the less common form of this sugar found in nature (Extended Data Fig. 2c). We next hypothesised that Z0415-9 might be regulated similarly to the canonical L-arabinose machinery. Activity of pMK1/*lux*-$P_{Z0415\text{-}9}$ in a Δ*araC* background was completely abolished during growth with L-arabinose. Importantly this could be complemented by expressing *araC in trans*, suggesting that this horizontally acquired system is co-regulated with the canonical L-arabinose utilisation machinery (Fig. 2b). We therefore named this locus Aau, for accessory L-arabinose uptake system.

Deletion of Aau did not result in a significant growth defect during *in vitro* growth on L-arabinose (Extended Data Fig. 3). We therefore hypothesised that it may play a more important role in the host-context during nutrient competition. To test this, we orally inoculated streptomycin treated BALB/c mice with a 1:1 mixture of wild type EHEC and Δ*aau* (Fig. 2c)[28]. While both strains initially colonised equally well, Δ*aau* was significantly attenuated for longer term persistence displaying a 10 to 100-fold decrease in competitive

fitness from day 9 onwards (Fig. 2d). This result implies that Aau provides a fitness benefit to EHEC within the complex and dynamic gut niche.

L-arabinose alters the EHEC transcriptome and enhances expression of the LEE-encoded T3SS

Our discovery of Aau prompted us to investigate the effects of L-arabinose utilisation on EHEC. Nutrients often act as signals that modulate virulence gene expression. We therefore hypothesised that L-arabinose may affect the expression of key EHEC virulence genes, as well as being able to act as a source of nutrition. To test this, we performed RNA-seq on EHEC cultured in MEM-HEPES with and without 1 mg/ml L-arabinose to late exponential phase. Strikingly, we identified 1187 significant (>1.5 fold-change; FDR $P < 0.05$) differentially expressed genes (DEGs) (Fig. 3a; Supplementary Table 1). As expected, the canonical L-arabinose utilisation genes (*araBAD, araE, araFGH*), as well as genes encoding Aau, were among the most strongly upregulated displaying increased expression of up to 42-fold ($P < 0.001$). Functional network analyses identified that the large abundance of DEGs unrelated to L-arabinose clustered into diverse biological pathways including metabolism, quorum sensing, signalling and regulation of biofilm formation (Extended Data Fig. 4a/b). While we anticipated shifts in expression of genes related to metabolism, we also noticed that several genes belonging to the LEE pathogenicity island were significantly upregulated in response to L-arabinose (Fig. 3b).

The LEE is essential for EHEC pathogenicity and is responsive to gut-associated cues. We therefore reasoned that L-arabinose may provide a fitness advantage through regulating T3SS activity either in concert to, or independent from a role in host-associated nutritional

competition. To assess the dynamics of LEE induction by L-arabinose, we engineered a transcriptional reporter of the T3SS (EHEC transformed with pMK1/*ux*-P_{LEE1}). Culture of this strain in MEM-HEPES (LEE-inducing conditions) supplemented with various concentrations of L-arabinose resulted in no increase in growth rate but a significant and prolonged increase in LEE promoter activity from late exponential phase onwards (Fig. 3c). RT-qPCR also confirmed a significant increase in transcription across all five LEE operons (Extended Data Fig. 5a). We detected an increase in LEE expression in the presence of L-arabinose at concentrations as low as 50 µg/ml (Extended Data Fig. 5b), within range of the amount quantified from the luminal content and faeces of mice maintained on a conventional diet (Extended Data Fig. 5c). We next observed an increased accumulation of T3SS associated effector proteins in the cytosol and cell-free supernatant by SDS-PAGE and western blot analysis (Fig. 3d). To test if this increase in T3SS expression and function resulted in enhanced host-cell interaction, we quantified adhesion of EHEC to cultured epithelial cells by fluorescence microscopy (Fig. 3e). Pre-exposing EHEC to L-arabinose before infection of epithelial cells significantly increased the number of attached EHEC and associated A/E lesions (identified as foci of actin accumulation) per infected cell (Fig. 3f). This was in parallel with an >10% increase in the number of individually infected cells in the presence of L-arabinose. These results collectively show that L-arabinose enhances the expression and function of the LEE-encoded T3SS, resulting in a greater capacity to cause A/E lesions on the host epithelial surface.

Metabolism of L-arabinose is essential to enhance T3SS expression

Nutrient sensing typically occurs either from the environmental (via two-component systems) or via a cognate transcription factor in the cytosol[32]. Due to L-arabinose utilisation occurring at three levels (uptake via multiple transporters, sensing by a dedicated transcription factor

and metabolism via AraBAD), this left the mechanism of T3SS regulation by L-arabinose unclear, as in principle it could occur at any of these three stages (Extended Data Fig. 6a)[30]. We therefore generated mutants in each component, first validating their roles in L-arabinose utilisation (Extended Data Fig. 6b), allowing compartmentalisation of each stage to explore the mechanism of T3SS enhancement. We first measured LEE activity in Δ*araC* and observed a complete reversal of L-arabinose enhanced T3SS expression. This was complemented *in trans* suggesting that sensing via AraC may directly regulate LEE expression (Fig. 4a). Next, we measured LEE activity in the Δ*araBAD,* Δ*araFGH* and Δ*araE* backgrounds. This revealed that enhanced LEE expression was also abolished in Δ*araBAD*, but not in the Δ*araFGH* background and only partially in Δ*araE,* likely since there are multiple routes for L-arabinose uptake and in line with the major dependence on AraE for growth on this sugar (Fig. 4b; Extended Data Fig. 6b). To narrow down the mechanism further, we measured LEE expression in Δ*araC* transformed with plasmids constitutively expressing either *araBAD* and *araE* in parallel (pSU-*araBAD/E*) or *araE* alone (pSU-*araE*), allowing functional differentiation between uptake and downstream metabolism in the cytosol (Extended Data Fig. 6b). The logic behind this experiment was that *araBAD* and *araE* expression (and therefore uptake and metabolism) is completely dependent upon AraC[33]. Therefore, bypassing the role of AraC using constitutively expressed *araBAD/araE* allowed us to conclusively determine whether L-arabinose sensing by AraC, uptake or metabolism caused enhanced LEE expression. Complementation of Δ*araC* with pSU-*araBAD/E* resulted in complete restoration of L-arabinose enhanced LEE expression above wild type levels, whereas constitutive expression of AraE alone (pSU-*araE*) had no restorative effect (Fig. 4c). Taken together, these data comprehensively show that L-arabinose metabolism (via AraBAD) in the cytosol is essential to enhance T3SS expression in response to this sugar.

To understand the ecological context of these results, we measured the temporal dynamics of *araBAD* transcription compared to the LEE, showing that induction of L-arabinose metabolism under these conditions significantly correlates (R = 0.92; $P < 0.001$) with LEE expression dynamics (Extended Data Fig. 7a). MEM-HEPES contains D-glucose, which suppresses L-arabinose utilisation genes by catabolite repression[30,33]. Glycolytic growth is also known to reduce maximal LEE expression[22]. We therefore reasoned that the temporal dynamics of our mechanism likely represented the switch to L-arabinose catabolism as D-glucose became depleted from the medium. Using thin layer chromatography, we measured the relative abundance of monosaccharides in cell-free supernatant of EHEC cultured in MEM-HEPES alone or supplemented with L-arabinose (Extended Data Fig. 7b). The data confirmed that D-glucose depletion coincided with activation of AraBAD (and as such, the LEE). This offered a logical explanation as to why enhanced LEE expression in response to L-arabinose metabolism was observed only at this stage of growth, when the repressive effects of D-glucose on *araBAD* and the LEE would no longer be a factor (Extended Data Fig. 7c). This suggests that dynamic changes in nutrient availability that likely occur within the complex host environment could have important downstream effects on the mechanisms of virulence factor regulation.

An intrinsic, generalised mechanism of T3SS regulation via pentose sugar metabolism

L-arabinose metabolism, like other aldopentose sugars such as D-ribose and D-xylose, can generate cellular energy via the pentose phosphate shunt[30]. While each sugar can support *E. coli* growth using dedicated genes, their metabolism converges at the generation of D-xylulose-5-phosphate, which enters the Embden-Meyerhof pathway and ultimately produces

pyruvate to generate cellular energy (Fig. 5a). It has been previously shown that exogenous pyruvate can enhance LEE expression[34]. We therefore reasoned that the observed effect of L-arabinose on expression of the T3SS could be due to excess pyruvate generation and hypothesised that D-ribose or D-xylose may have a similar effect, given their common cellular fate. Strikingly, enhanced LEE expression dynamics were also observed when EHEC was cultured in the presence of either sugar (Fig. 5b). Importantly, deletion of the D-ribose utilisation genes (Δ*rbsDACBKR*) eliminated the ability to grow on this sugar as a sole carbon source and, as predicted, abolished the observed increase in T3SS expression, suggesting that the common features of cellular pentose sugar metabolism determine the regulatory effect on the LEE (Fig. 5c). This is in line with our finding that L-arabinose must be metabolised to exert its T3SS regulatory effect. Finally, we confirmed that addition of pyruvate to the medium enhanced LEE expression and demonstrated that growth in the presence of both exogenous pyruvate and L-arabinose displayed an additive effect, further enhancing T3SS expression levels (Fig. 5d). This suggests that the common cellular fate of pentose sugar metabolism and generation of key downstream metabolites is a generalised mechanism by which A/E pathogens can regulate virulence gene expression.

L-arabinose metabolism is required for maximal fitness during enteric infection

While L-arabinose metabolism has been reported to provide a fitness advantage to *E. coli* during colonisation of streptomycin-treated mice, this model does not reflect A/E pathogenesis and therefore the role of L-arabinose utilisation during infection was unknown[28]. To test this, we employed *C. rodentium* as a model A/E pathogen, first confirming that L-arabinose also enhanced its T3SS expression (Extended Data Fig. 8). We next compared the ability of wild type *C. rodentium* and a Δ*araBAD* derivative to colonise BALB/c mice (n =

10). During the early stages of infection, bacterial burdens in faeces were similar between the two strains (days 1 to 9 post-infection). However, during the infectious peak at day 13 and onwards into the resolving phase, we observed a significantly more rapid clearance of Δ*araBAD* from the mice when compared to wild type (Fig. 6a). To directly test if L-arabinose utilisation conferred a competitive fitness advantage, we orally infected mice with a 1:1 mixture of wild type *C. rodentium* and Δ*araBAD* (n = 10). The Δ*araBAD* mutant was more significantly outcompeted by the wild type in a competitive infection, displaying an increasing fitness defect from day 9 until termination at day 21 (Fig. 6b). To determine whether the fitness defect was associated with tissue adhered or luminal *C. rodentium,* we quantified the bacterial burden of colon sections from mono-infected mice that were cleared of the luminal content. In agreement with the faecal counts, colon-associated Δ*araBAD* were recovered in significantly fewer numbers than wild type *C. rodentium* suggesting that L-arabinose metabolism actively promotes colonisation of host-tissue (Fig. 6c).

Finally, we aimed to determine whether the observed fitness defect of Δ*araBAD* was driven by the inability to utilise L-arabinose as nutrient or lack of its input as a positive stimulus for regulation of the LEE. To test this, we generated Δ*araBAD* in *C. rodentium*$^{P\textit{ler}\text{-const}}$, a strain that expresses the LEE constitutively via a single base deletion in the -30 element of the *ler* promoter[21,35], and tested its colonisation dynamics in mice (n=9). While *C. rodentium*$^{P\textit{ler}\text{-const}}$ displayed similar colonisation dynamics to that of the wild type strain, Δ*araBAD*$^{P\textit{ler}\text{-const}}$ no longer displayed a significant decrease in within-host fitness (Fig. 6d). This suggested that constitutive expression of the T3SS can overcome the fitness defect of Δ*araBAD* by mimicking the stimulatory effect that L-arabinose metabolism imparts on the LEE and suggesting that during a natural infection *C. rodentium* does not require L-arabinose as a source of nutrition.

Discussion

Successful cross-feeding on nutrients derived from the host, microbiota or diet is essential for bacterial pathogens to overcome colonisation resistance and replicate within their preferred host-niche[1,2]. Moreover, nutrients that support growth can often be "sensed" as stimuli that trigger virulence gene expression, therefore coupling virulence and metabolism to maximise fitness[2–5]. Here, we have found that metabolism of L-arabinose by A/E pathogens promotes within-host fitness by generating central metabolites that converge with regulation of virulence. While L-arabinose supports the growth of EHEC and *C. rodentium in vitro,* it is not required as a source of nutrition during infection of mice. Instead, its catabolism to pyruvate stimulates expression of the LEE-encoded T3SS, a virulence factor essential for colonisation, which provides an advantage during host-colonisation. We suggest that this underlying mechanism revolving around the generation of central metabolic products that enhance LEE expression is generalisable for other pentose sugars with a similar cellular fate. We also identified a novel ABC transporter, termed Aau, that is selectively induced in response to L-arabinose exclusively, suggesting that EHEC has an enhanced ability to scavenge L-arabinose in the gut and maximise its competitiveness through the convergence of virulence and metabolism.

Freter's nutrient-niche hypothesis suggests that within the complex, multi-species environment of the gut, an invading pathogen must be able to utilise at least one limiting nutrient better than the commensals it competes with[27,36]. L-arabinose-containing polysaccharides are a major component of dietary fibre, with free L-arabinose being highly enriched in the colon of humans and animals[37,38]. As such, non-starch hydrolysis and degradation of L-arabinose-containing polysaccharides by saccharolytic members of the gut

306 microbiota could liberate free L-arabinose that is subsequently exploited by species such as
307 *E. coli* during the process of cross-feeding[39–43]. Indeed, research from the labs of Conway and
308 Cohen describing carbon nutrition of *E. coli in vivo*, has shown that L-arabinose is utilised by
309 both commensal and EHEC strains in the streptomycin-treated mouse intestine[28].
310 Furthermore, they described how EHEC metabolises L-arabinose earlier in the hierarchy of
311 carbon utilisation than a commensal strain, when grown on a mixture of sugars. This suggests
312 that L-arabinose is an important nutrient for *E. coli* to sustain growth within the mammalian
313 intestine and that EHEC has evolved an enhanced ability to do this.
314
315 An enhanced ability of EHEC to utilise L-arabinose may be achieved by more efficient
316 breakdown within or uptake into the cell. Our discovery of the novel L-arabinose ABC
317 transporter Aau suggests that the latter may be particularly important in providing EHEC with
318 a competitive edge in nature. Indeed, despite Aau being encoded on an O-island without a
319 dedicated regulatory system[15], we have found that its activation relies on the core genome-
320 encoded AraC in a manner akin to the canonical L-arabinose utilisation machinery[33]. This
321 results in a rapid and co-ordinated response to the sugar by inducing the expression of
322 multiple uptake systems. While we are currently investigating the underlying mechanisms
323 (affinity and kinetics of uptake in comparison to the canonical AraE/FGH uptake systems) of
324 this transporter, there have been other reports of Aau induction in ecologically relevant
325 scenarios that support its likely role in EHEC infection. Transcriptome studies of EHEC in
326 response to spinach and lettuce lysates found that the canonical L-arabinose genes and those
327 encoding Aau were significantly upregulated, indicating a potential dietary source of L-
328 arabinose[29,44]. Furthermore, the permease was identified as being one of the most highly
329 expressed EHEC proteins during human infection using *in vivo* antigen detection technology[45],

which is particularly noteworthy considering that *C. rodentium* (a dedicated murine pathogen) does not encode this system[46]. This therefore provides compelling evidence that Aau, and its role in L-arabinose scavenging, likely represents a crucial element of the real-world, human infectious context.

Our discovery that L-arabinose metabolism is required for regulation of the T3SS reveals new insights into how we perceive virulence factor control in the host context. Traditionally, nutrients or small metabolites are thought to be "sensed" either in the cytosol post-uptake by a cognate transcription factor, or from the environment by periplasmic interactions with two-component systems. This is indeed true of pyruvate, which has been described as a virulence-inducing signal in many species beyond *E. coli.* For example, both *Staphylococcus aureus* and *Salmonella* Typhimurium upregulate virulence gene expression in response to host-derived pyruvate via two-component systems, the latter of which is particularly noteworthy here as this regulation occurred independently of any growth advantage[47,48]. Our proposed mechanism provides new evidence that the endogenous metabolism of certain molecules, that do not in themselves act as "signals", can cause dramatic shifts in the pathogen response. This follows on from our recent study where we found that metabolism of microbiota-derived 1,2-propanediol generates the short chain fatty acid propionate, which in turn provides a signal that activates LEE expression in *C. rodentium* independently of a growth advantage[21]. In a comparable manner, uropathogenic *E. coli* produces excess pyruvate via serine deamination, which in terms results in increased sensing of pyruvate via a two-component system and subsequent uptake[49]. These studies collectively point to alternative mechanisms of virulence and fitness regulation whereby "signals" are generated endogenously by the pathogen based on the biochemical status of the environment.

Importantly, it also alludes to a more complex view of the consequences of niche adaptation, whereby metabolic systems can provide additional underappreciated benefits to a pathogen through the regulation of genetically unlinked virulence factors. The mechanism of L-arabinose induced LEE expression also reinforces the benefit of being able to dynamically shift between available nutrients *in vivo* that are likely to be in constant flux and thus would aid in limiting any associated repressive effects of catabolite repression on virulence gene expression[22,27,28].

Finally, while the canonical L-arabinose utilisation pathway has been understood and exploited in bacterial genetics for decades, the relevance of L-arabinose metabolism in promoting pathogenesis has only recently emerged. Dietary L-arabinose utilisation by *S.* Typhimurium promotes expansion in the gastrointestinal tract of mice[50]. This occurs via a mechanism involving an intrinsic alpha-*N*-arabinofuranosidase that liberates the sugar from dietary polysaccharides. Dietary fibre is traditionally thought to provide a benefit in maintaining colonisation resistance via maintenance of microbiota composition, the mucosal barrier and short chain fatty acid production[51–53]. The data from Ruddle *et al.* and ours therefore highlight the diverse strategies that enteric pathogens can also use to take advantage of fibre-derived sources of nutrition. Indeed, EHEC utilises several strategies to achieve this. For example, sensing of pectin-derived galacturonic acid promotes initial expansion in the mouse gut while also directly regulating LEE expression throughout the infection lifespan[54]. These studies collectively highlight the diverse strategies that enteric pathogens use to exploit the host diet and aid in overcoming colonisation resistance.

In summary, we have identified a new mechanism by which A/E enteric pathogens regulate an essential virulence factor in response to host associated L-arabinose and identified novel genes involved in its uptake from the environment. Our data reveal that important pathogenic processes can be controlled via central metabolism of sugars independently of any effects on growth, warranting a new perspective on this process and its potential impact in bacterial pathogens. We anticipate that by expanding our view of how nutrient "sensing" occurs via metabolism we have the potential to discover new mechanisms of bacterial virulence regulation that are highly relevant to the host context. While our studies using *C. rodentium* cannot rule out that L-arabinose metabolism may benefit EHEC on a nutritional level in certain contexts (such as during human infection), the results highlight how virulence factor regulation can provide additional benefits for overcoming colonisation resistance. Lastly, by identifying further functional distinctions between EHEC and *C. rodentium* (exemplified here by the EHEC-specific Aau), we may reveal genetic factors that dictate host-range of these human and murine pathogens exclusively.

Acknowledgements

We are very grateful to Dr Nicky O'Boyle for his critical appraisal of the work. We also would like to thank the Comparative Biology Centre staff for their support with the animal studies and the Newcastle University Genomics Core Facility for RNA-sequencing. This work was supported by a Springboard Award from the Academy of Medical Sciences/Wellcome Trust [SBF005\1029], a Royal Society Research Grant [RGS\R2\202100], a Medical Research Council Career Development Award [MR/X007197/1] and a Faculty Fellowship (Newcastle University) awarded to J.P.R.C. E.C.L was supported by a Wellcome Trust Discovery Award [226644/Z/22/Z]. C.C. was supported by a PhD studentship from the Barbour Foundation.

Author contributions

C.C. and J.P.R.C conceptualised and designed the research. C.C. and J.P.R.C. performed the research. C.C., R.T.W., L.C.B., R.G. and J.P.R.C. analysed the data. C.J.S., S.A.B., R.G., E.C.L. and J.P.R.C. contributed reagents/analytical tools. C.C. and J.P.R.C. wrote the paper with input from all other authors.

Competing interest

The authors declare no competing interest

Materials and methods

Bacterial growth conditions

All bacterial strains used are listed in Supplementary Table 2. Single colonies were used to inoculate 5 mL LB media before overnight culture at 37 °C and back dilution into LB or MEM-HEPES (T3SS-inducing conditions) to an OD_{600} of 0.05 for specific assays or growth analysis. For sole carbon source experiments, overnight cultures were first washed three times in PBS to remove trace LB and used to inoculate M9 minimal media supplemented with a single carbon source (D-glucose, L-arabinose, D-ribose or D-xylose) at the concentrations indicated. All experiments were performed at 37 °C with shaking at 200 rpm. When necessary, antibiotics were used at the following concentrations: 100 µg/mL ampicillin, 50 µg/mL kanamycin, 20 µg/mL chloramphenicol, 100 µg/mL streptomycin and 50 µg/mL nalidixic acid. All chemicals and media were purchased from Merck or Thermo Fisher Scientific.

Lambda red recombineering

Non-polar single-gene deletions were generated using the Lambda Red recombineering system[55]. Briefly, the FRT-Kanamycin or FRT-Chloramphenicol cassette was amplified from pKD4 or pKD3 respectively, using primers containing 50 bp overhangs homologous to the regions directly up- and downstream regions of the gene of interest. Parental strains carrying pKD46 were cultured in SOB media (Ampicillin; 30 °C) containing 10 mM L-arabinose to an OD_{600} of 0.4 before electroporation with 100 ng of the PCR product described above. Subsequent recovery was carried out in SOC media at 37 °C prior to plating the transformations on LB-agar containing the appropriate antibiotic for selection of recombinants. Positive mutants were identified by colony PCR and verified by Sanger sequencing. To remove the resistance cassette, positive mutants were subsequently

433 transformed with pCP20, plated on LB-agar containing Ampicillin and grown overnight at 30
434 °C to induce FLP recombinase activity. Colonies were re-streaked non-selectively at 42 °C to
435 cure the strains of pCP20. Plasmids used are listed in Supplementary Table 3. All primers used
436 are listed in Supplementary Table 4.
437
438 **Plasmid construction**
439 All complementation and reporter plasmids were generated by standard restriction enzyme
440 cloning except for pSU-*araBADE,* which was generated using the NEBuilder® HiFi DNA
441 Assembly Cloning kit as per the manufacturer's instructions. All plasmids generated are listed
442 in Supplementary Table 3 and all primer sequences are listed in Supplementary Table 4. For
443 reporter plasmids, promoter regions comprising of approximately 300 bp upstream of the
444 gene of interest were PCR amplified from genomic DNA using primers containing EcoRI (5')
445 and BamHI (3') restriction sites. PCR products were gel extracted, digested, phosphatase
446 treated and ligated into pMK1*lux*. This resulted in a transcriptional reporter whereby the
447 respective promoter for a gene of interest was fused to the *luxCDABE* operon from
448 *Photorhabdus luminescens*. For complementation constructs, the gene/operon of interest
449 was PCR amplified from genomic DNA using primers containing BamHI (5') and XbaI (3')
450 restriction sites. PCR products were ligated into pSU-PROM or pACYC184. All restriction
451 enzymes, Q5 high fidelity polymerase, Antarctic Phosphatase and T4 ligase were purchased
452 from New England Biolabs. Cloning was confirmed by sequencing of plasmid inserts.
453
454 **LUX-promoter fusion reporter assays**
455 Promoter activity was determined by measuring cell density (OD_{600}) and absolute
456 luminescence of cultures carrying LUX reporter fusions using a FLUOstar Omega microplate

457 reader (BMG Labtech). Relative luminescence units were calculated by dividing absolute
458 luminescence values by the OD_{600} at each given timepoint. Assays were performed as a time-
459 course in 200 µL wells of white walled, clear flat-bottom, 96-well polystyrene microtiter
460 plates. or taken as single measurements. Alternatively, endpoint luminescence
461 measurements were taken from reporter strains grown in MEM-HEPES for 2 h before being
462 spiked with the indicated sugar and grown for a further 5 h. Control cultures were spiked with
463 the equivalent volume of sterile PBS. Individual culture volumes were inoculated with the
464 strain of interest at a ratio of 1:100.
465
466 **Secreted protein SDS-PAGE profiling**
467 T3SS-mediated protein secretion was profiled by culturing EHEC in 50 mL MEM-HEPES at
468 37 °C to late exponential phase (OD_{600} of 0.8-0.9). The cell-free supernatant was separated
469 from the cellular fraction by centrifugation and filtration using a 0.4 µm filter. Total secreted
470 protein was precipitated from the cell-free supernatant with 10% ice-cold trichloroacetic acid
471 at 4 °C overnight. Secreted proteins were concentrated by centrifugation for 1 h at 4000 *g*.
472 The supernatant was discarded, and pellets resuspended in 150 µL of 1 x lithium dodecyl
473 sulphate buffer and boiled at 95 °C for 10 min. Samples were normalised by OD_{600} at the point
474 of harvesting. All protein samples were separated by SDS-PAGE using a 4-12 % Bis-Tris
475 NuPAGE gel (Invitrogen) and running at 150 V for 90 min, before staining with Coomassie
476 blue.
477
478 **Immunoblot analysis**
479 Secreted protein fractions and corresponding whole-cell lysates (obtained by boiling the cell
480 pellet in sample buffer as above) were used for immunoblot analysis. Samples were separated

by SDS-PAGE and were transferred from a 4-12 % Bis-Tris NuPAGE gel to a 0.45 µM nitrocellulose membrane (GE Healthcare) using an XCell II Blot module (Invitrogen) at 30 V for 90 mins. Membranes were then blocked with 5 % skim milk powder in PBS-Tween at room temperature for one hour before being incubated with primary antibodies for one hour. Membranes were washed three times in PBS-Tween for 10 minutes each before incubating for one hour with secondary antibodies. Primary antibodies used were anti-EspD (1:2500) and anti-GroEL (1:25000). Secondary antibodies used were anti-mouse and anti-rabbit HRP-conjugated (1:20000). Immunoblots were incubated with SuperSignal West Pico chemiluminescent substrate (Pierce) for five minutes before imaging using a G:Box Chemi system (Syngene).

HeLa cell adhesion assays and fluorescence microscopy

For adhesion assays, HeLa cells were seeded onto sterile coverslips coated with rat tail collagen (10^4 cells per coverslip in 12-well plates) in DMEM supplemented with 10 % foetal calf serum and 1 % Penicillin/Streptomycin. Cells were incubated overnight at 37 °C with 5 % CO_2. Prior to infection, cells were washed twice with PBS and fresh MEM-HEPES supplemented with or without 5 mg/mL L-arabinose was added. A 40 µL volume of MEM-HEPES containing EHEC grown to OD_{600} 0.9 and back-diluted to 0.1 was added to each coverslip. Plates were centrifuged at 400 rpm for 3 min and incubated at 37 °C with 5 % CO_2 for two hours. Wells were then washed with fresh media and incubated for a further 3 hours. The wells were washed three times and fixed with 4 % paraformaldehyde for 15 minutes at room temperature. Wells were washed an additional two times and permeabilised with 0.1 % Triton X-100 for 5 minutes. The wells were washed twice more before incubation with ActinRed™ 555 ReadyProbes™ reagent (Rhodamine phalloidin; Invitrogen). Tissues were washed for a

504 final time and mounted onto glass slides using Fluoroshield™ with DAPI. All washes were
505 done using sterile PBS. For all assays, EHEC transformed with an *rpsM*:GFP reporter plasmid
506 were used. Slides were imaged using a Zeiss Axioimager at x40 magnification. Data were
507 collected by imaging 20 random fields of view across three coverslips prepared on
508 independent occasions.
509
510 **Animal experiments**
511 Strains of interest were used to mono- and co-infect BALB/C mice as previously described[19].
512 In brief, groups of five adult female mice (18-20 g) were inoculated by oral gavage using 200
513 µL of PBS suspension containing 2 x 10^9 CFU of either a single strain or two strains mixed at a
514 ratio of 1:1. For the analysis of bacterial colonisation, stool samples were aseptically
515 recovered, weighed, and homogenised in PBS before being serially diluted. The number of
516 CFU per gram of stool was determined by plating onto LB-agar with the appropriate antibiotic.
517 Experiments were typically performed on two independent occasions and CFU counts were
518 analysed using the Mann-Whitney *U*-test. For experiments involving *E. coli*, streptomycin
519 resistant strains encoding an intact Aau locus were used and drinking water was
520 supplemented with streptomycin (5 g/L) two-days prior to oral gavage and maintained for the
521 duration of the experiment[28]. For experiments involving *C. rodentium,* animals were given
522 normal drinking water.
523
524 **Ethics statement**
525 All animal experiments were performed in strict accordance with the United Kingdom Home
526 Office Animals Scientific Procedures Act of 1986 under the personal project licence number
527 PP8850146. The experiments were subject to local ethical approval and consideration given

to the refine, reduce and replace principals wherever possible so as all efforts were made to minimize animal suffering.

RNA extraction

EHEC cultures were mixed with 2 volumes of RNAprotect reagent (Qiagen), incubated at room temperature for 5 minutes and harvested by centrifugation. Total RNA extraction was done using a Monarch® Total RNA Miniprep Kit (New England Biolabs) prior to treatment with TURBO DNAse (Ambion). DNA-free RNA samples were analyzed by Qubit (ThermoFisher Scientific) and assessed for degradation using agarose gel electrophoresis.

Quantitative real time PCR (RT-qPCR)

RNA samples were normalised, and cDNA synthesis performed using a LunaScript RT SuperMix kit (New England Biolabs). RT-qPCR was performed on the resulting cDNA using a LightCycler 96 Real-Time PCR system (Roche) and Luna Universal qPCR Master Mix kit (New England Biolabs). Reactions were performed as technical replicates within each of three biological replicates. All genes were normalised against the housekeeping gene, *groEL*. All primers used in RT-qPCR were checked for efficiency (90-110%) using standards made from template gDNA (100, 20, 4, 0.8 and 0.16 ng/µl). The data was analysed by the $2^{-\Delta\Delta CT}$ method[56]. Primer sequences are listed in Supplementary Table 4

RNA-sequencing and transcriptome analysis

Ribosomal depletion and library assembly of DNA-free total RNA samples was carried out using an Illumina Ribo-Zero Tru-seq kit according to the manufacturer's specifications. Samples were processed by the Newcastle Genomics Core Facility. Sequencing was carried

552 out using a mid-range run on the Illumina Next-seq platform, generating 75-bp single-end
553 reads. Raw read quality was checked using FastQC
554 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). To estimate transcript
555 abundance, SALMON was used under the default parameters for the mapping of reads to the
556 *E. coli* O157:H7 strain EDL933 reference sequence (Accession GCA_000006665), retrieved
557 from Ensembl. Transcript level counts outputted from SALMON were then summarised at the
558 gene-level using tximport[57]. DESeq2 (v.1.28.1) was used to normalise RNA-seq count data and
559 identify differentially expressed genes between conditions[58]. Differentially expressed genes
560 displayed an absolute fold change of >1.5 and were a false discovery rate of 5 % (adjusted *P*
561 < 0.05). Volcano plots were generated in R studio using the enhanced volcano package
562 (v.1.6.0). Functional enrichment and gene ontology analysis was performed using STRING to
563 identify protein-protein interactions related to the differentially expressed genes identified
564 using default parameters and excluding any disconnected nodes[59].
565
566 **L-arabinose quantification**
567 L-arabinose was quantified from homogenised and filtered luminal and colon tissue samples
568 of BALB/c mice using an L-arabinose/D-galactose assay kit (Megazyme) as per the
569 manufacturer's instructions.
570
571 **Thin layer chromatography (TLC)**
572 Cell-free supernatants of EHEC cultured in MEM-HEPES alone or supplemented with L-
573 arabinose were prepared by removing 1 mL of culture every hour, removing the supernatant
574 by centrifugation and passing through a 0.2 µM filter. Using TLC aluminium plates (Merck), 6
575 µL of sample for each time point was spotted 1 cm from the bottom and allowed to dry. The

plates were run in solvent (1-butanol:acetic acid:water at a ratio of 2:1:1), dried and sugars visualised by immersion in Orcinol stain.

Gene carriage analysis

To assess the frequency of carriage of the Z0415-9 locus and LEE island across the *E. coli* lineage, paired-end sequence read data for 1,067 strains previous described[25] were retrieved from the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (accessed 16th September 2021), using the 'prefetch' and 'fastq-dump' tools within the SRA Toolkit v2.9.0-mac64 (http://ncbi.github.io/sra-tools). FastQC v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to generate quality statistics for the paired-end reads, which were aggregated into a single report and visualised using MultiQC v1.11[60]. Kraken v2.0.8-beta was used to screen the raw Illumina sequencing data for contamination against the NCBI RefSeq Database[61]. Raw reads were filtered using Trimmomatic v0.36 by removing low-quality bases and read pairs together with Illumina adaptor sequences (settings: LEADING:3, TRAILING:3, MINLEN:36, SLIDINGWINDOW:4:15)[62]. The complete chromosomes of the 56 genomes were used to simulate error free reads. This was done using the software package ART (version ART-MountRainier-2016-06-05) which simulated paired-end Illumina reads to 60x coverage with an insert size of 340 ± 40 bp[63]. The average sequence coverage depth was estimated using the Burrows–Wheeler Aligner (BWA) v0.7.15[64]; SAMtools v1.2[65]; Picard v2.7.1 (https://github.com/broadinstitute/picard); the Genome Analysis Tool Kit v3.2-2 (GATK)[66,67]; BEDTools v2.18.2[68]; and SNPEff v4.1 as implemented in SPANDx v3.2[69,70]. In brief, the quality-trimmed paired-end reads were mapped to the complete chromosome of strain EDL933 (GenBank: AE005174)[15]. Trimmed reads for the draft genomes were *de novo*

assembled using Shovill v1.0.4 (https://github.com/tseemann/shovill) (--gsize 5M, --minlen 200, --mincov 10, --opts "--sc"), which implements: Seqtk v1.3-r106 (https://github.com/lh3/seqtk); Lighter v1.1.2[71]; FLASH v1.2.11[72]; SPAdes v3.13.1[73]; Samclip v0.2 (https://github.com/tseemann/samclip); SAMtools v1.8[65]; BWA-MEM v0.7.17-r1188[64]; and Pilon v1.22[74]. QUAST v4.5 was used to assess the *de novo* assembly metrics generated from Shovill by comparing each draft assembly to strain EDL933[75]. We identified and excluded the sequence data for certain strains from further analysis based on sequencing coverage being below 20-fold, the presence of contaminants, and the genome length from *de novo* assembly falling outside the upper (Q3) and lower (Q1) 1.5 x interquartile range (i.e., < 4,095,491 bp or > 5,889,907 bp) (*n*=157). To assess the publicly available strains for the presence of strain mixtures, paired-end reads were mapped onto the chromosome of EDL933 using SPANDx to generate annotated single-nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) matrices. Heterozygous SNPs in each genome were identified from GATK UnifiedGenotyper VCF output. MLST v2.19.0 (https://github.com/tseemann/mlst) with default settings was used to characterise the multi-locus sequence type (MLST) of the remaining 949 strains (*n*=948 excluding EDL933) against the *E. coli* MLST allelic profiles hosted on PubMLST[76]. High-resolution analysis of genetic variants was performed using SPANDx with EDL933 as a reference[70]. A maximum likelihood phylogenetic tree from the non-recombinant SNP alignment was generated using RAxML v8.2.10 (GTR-GAMMA correction) thorough optimisation of 10 distinct, randomized maximum parsimony trees[77]. The resulting phylogenetic tree was visualised using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Protein bioinformatics

AlphaFold modelling was performed using AlphaFold version 2.1.1 implemented on the Monash University MASSIVE M3 computing cluster[78,79]. The amino acid sequences of the periplasmic binding protein (Z0415) (minus the predicted signal peptide), ATPase (Z0416-7) and two permease subunits (Z0418/Z0419) of Aau were provided and modelling was run in multimer mode, with a single molecule of each subunit requested. The quality of five ranked models produced by AlphaFold were assessed based on pLDDT score and compared for consistency with the top-ranked model used for further analysis and figure generation.

To identify ABC transporters closely related to Z0415-19, a phylogenetic analysis was conducted based on the associate amino acid sequence. The ABC_tran_pfam domain (pfam00005) was searched against EHEC EDL933 using the IMG/M server 'find function' tool. The returned hits were then filtered based on the predicted substrates to select only those predicted to transport sugar substates. The amino acid sequences were then exported and aligned in MEGAX (V 10.1.8) using MUSCLE[80]. The evolutionary history was inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model. The tree with the highest log likelihood (-11697.05) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.6636)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis

involved 12 amino acid sequences. There were a total of 546 positions in the final dataset. Evolutionary analyses were conducted in MEGA X[81].

Statistical analyses

Graphs were generated and statistical analyses were performed using GraphPad Prism version 8. Test details for each experiment can be found within the associated figure legends. The student's *t*-test was used for comparing two groups with normally distributed data and equal variance. The Mann-Whitney U test was used to compare two groups that were not normally distributed and had unequal variance (mono-infections). The Wilcoxon signed-rank test was used to compare matched samples (co-infections or co-colonisations). The Fisher's exact test was used in the analysis of Odd's ratio contingency tables. *P*-values of less than or equal to 0.05 were considered statistically significant.

References

1. Caballero-Flores, G., Pickard, J. M. & Núñez, G. Microbiota-mediated colonization resistance: mechanisms and regulation. *Nature Reviews Microbiology* **21**, 347–360 (2023).

2. Baümler, A. J. & Sperandio, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).

3. Connolly, J. P. R., Brett Finlay, B. & Roe, A. J. From ingestion to colonization: The influence of the host environment on regulation of the LEE encoded type III secretion system in enterohaemorrhagic *Escherichia coli*. *Frontiers in Microbiology* **6,** 568 (2015).

4. Turner, N. C. A., Connolly, J. P. R. & Roe, A. J. Control freaks—signals and cues governing the regulation of virulence in attaching and effacing pathogens. *Biochemical Society Transactions* **47**, 229–238 (2018).

5. Wale, K. R., Cottam, C., Connolly, J. P. & Roe, A. J. Transcriptional and metabolic regulation of EHEC and *Citrobacter rodentium* pathogenesis. *Current Opinion in Microbiology* **63**, 70–75 (2021).

6. Kaper, J. B., Nataro, J. P. & Mobley, H. L. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2**, 123–40 (2004).

7. Croxen, M. A. & Finlay, B. B. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Reviews Microbiology* **8**, 26–38 (2010).

8. Moon, H. W., Whipp, S. C., Argenzio, R. A., Levine, M. M. & Giannella, R. A. Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. *Infect. Immun.* **41**, 1340–1351 (1983).

9. Kenny, B. *et al.* Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* **91**, 511–520 (1997).

10. Jerse, A. E., Yu, J., Tall, B. D. & Kaper, J. B. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7839–7843 (1990).

11. Mcdaniel, T. K., Jarvis, K. G., Donnenberg, M. S. & Kaper, J. B. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1664–1668 (1995).

12. Deng, W. *et al.* Dissecting virulence: Systematic and functional analyses of a pathogenicity island. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3597–3602 (2004).

13. Dean, P. & Kenny, B. The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. *Current Opinion in Microbiology* **12**, 101–109 (2009).

14. Tobe, T. *et al.* An extensive repetoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14941–14946 (2006).

15. Perna, N. T. *et al.* Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533 (2001).

16. Wong, A. R. C. *et al.* Enteropathogenic and enterohaemorrhagic *Escherichia coli:* Even more subversive elements. *Molecular Microbiology* **80**, 1420–1438 (2011).

17. Collins, J. W. *et al. Citrobacter rodentium*: Infection, inflammation and the microbiota. *Nature Reviews Microbiology* **12**, 612–623 (2014).

18. Mullineaux-Sanders, C. *et al. Citrobacter rodentium*–host–microbiota interactions: immunity, bioenergetics and metabolism. *Nature Reviews Microbiology* **17**, 701–715 (2019).

19. Crepin, V. F., Collins, J. W., Habibzay, M. & Frankel, G. *Citrobacter rodentium* mouse model of bacterial infection. *Nat. Protoc.* **11**, 1851–1876 (2016).

20. Kamada, N. *et al.* Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* **336**, 1325–1329 (2012).

21. Connolly, J. P. R. *et al.* Host-associated niche metabolism controls enteric infection through fine-tuning the regulation of type 3 secretion. *Nat. Commun.* **9**, 4187 (2018).

22. Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G. & Sperandio, V. Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli. MBio* **3**, e00280-00212 (2012).

23. Curtis, M. M. *et al.* The gut commensal *Bacteroides thetaiotaomicron* exacerbates enteric infection through modification of the metabolic landscape. *Cell Host Microbe* **16**, 759–769 (2014).

24. Menezes-Garcia, Z., Kumar, A., Zhu, W., Winter, S. E. & Sperandio, V. L-Arginine sensing regulates virulence gene expression and disease progression in enteric pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12387–12393 (2020).

25. Connolly, J. P. R. *et al.* The host metabolite D-serine contributes to bacterial niche specificity through gene selection. *ISME J.* **9**, 1039–1051 (2015).

26. O'Boyle, N., Connolly, J. P. R., Tucker, N. P. & Roe, A. J. Genomic plasticity of pathogenic *Escherichia coli* mediates D-serine tolerance via multiple adaptive mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22484–22493 (2020).

27. Conway, T. & Cohen, P. S. Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut. *Microbiol. Spectr.* **3** MBP-0006-2014 (2015).

28. Fabich, A. J. *et al.* Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect. Immun.* **76**, 1143–1152 (2008).

29. Crozier, L. *et al.* The role of L-arabinose metabolism for *Escherichia coli* o157:H7 in edible plants. *Microbiol. (United Kingdom)* **167,** 1-12 (2021).

729 30. Mayer, C. & Boos, W. Hexose/Pentose and Hexitol/Pentitol Metabolism. *EcoSal Plus* **1**
730 10.1128/ecosalplus.3.4.1 (2005).
731 31. Rice, A. J., Park, A. & Pinkett, H. W. Diversity in ABC transporters: Type I, II and III
732 importers. *Critical Reviews in Biochemistry and Molecular Biology* **49**, 426–437 (2014).
733 32. Zschiedrich, C. P., Keidel, V. & Szurmant, H. Molecular Mechanisms of Two-Component
734 Signal Transduction. *Journal of Molecular Biology* **428**, 3752–3775 (2016).
735 33. Schleif, R. AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and
736 the light switch mechanism of AraC action. *FEMS Microbiology Reviews* **34**, 779–796
737 (2010).
738 34. Carlson-Banning, K. M. & Sperandio, V. Catabolite and oxygen regulation of
739 enterohemorrhagic *Escherichia coli* virulence. *MBio* **7,** e01852-16 (2016).
740 35. Mullineaux-Sanders, C. *et al.* Citrobacter rodentium Relies on Commensals for
741 Colonization of the Colonic Mucosa. *Cell Rep.* **21**, 3381–3389 (2017).
742 36. Freter, R., Brickner, H., Fekete, J., Vickerman, M. M. & Carey, K. E. Survival and
743 implantation of *Escherichia coli* in the intestinal tract. *Infect. Immun.* **39**, 686–703
744 (1983).
745 37. Yip, A. Y. G. *et al.* Antibiotics promote intestinal growth of carbapenem-resistant
746 Enterobacteriaceae by enriching nutrients and depleting microbial metabolites. *Nat.*
747 *Commun.* **14**, 5094 (2023).
748 38. Liu, B. *et al.* Enterohaemorrhagic *E. coli* utilizes host- and microbiota-derived L-malate
749 as a signaling molecule for intestinal colonization. *Nat. Commun.* **14**, 7227 (2023).
750 39. Schutte, J. B., de Jong, J., van Weerden, E. J. & Tamminga, S. Nutritional implications of
751 l -arabinose in pigs. *Br. J. Nutr.* **68**, 195–207 (1992).
752 40. Schwalm, N. D., Townsend, G. E. & Groisman, E. A. Multiple signals govern utilization

of a polysaccharide in the gut bacterium *Bacteroides thetaiotaomicron*. *MBio* **7**, 10.1128/mbio.01342-16 (2016).

41. Pereira, G. V. *et al.* Degradation of complex arabinoxylans by human colonic Bacteroidetes. *Nat. Commun.* **12**, 459 (2021).

42. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).

43. Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).

44. Kyle, J. L., Parker, C. T., Goudeau, D. & Brandl, M. T. Transcriptome analysis of *Escherichia coli* O157:H7 exposed to lysates of lettuce leaves. *Appl. Environ. Microbiol.* **76**, 1375–1387 (2010).

45. John, M. *et al.* Use of in vivo-induced antigen technology for identification of *Escherichia coli* O157:H7 proteins expressed during human infection. *Infect. Immun.* **73**, 2665–2679 (2005).

46. Petty, N. K. *et al.* The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *J. Bacteriol.* **192**, 525–538 (2010).

47. Harper, L. *et al. Staphylococcus aureus* responds to the central metabolite pyruvate to regulate virulence. *MBio* **9**, 10.1128/mbio.02272-17 (2018).

48. Jiang, L. *et al. Salmonella* Typhimurium reprograms macrophage metabolism via T3SS effector SopE2 to promote intracellular replication and virulence. *Nat. Commun.* **12**, 879 (2021).

49. Wiebe, M. A. *et al.* Serine Deamination Is a New Acid Tolerance Mechanism Observed in Uropathogenic *Escherichia coli*. *MBio* **13**, e02963-22 (2022).

50. Ruddle, S. J., Massis, L. M., Cutter, A. C. & Monack, D. M. *Salmonella*-liberated dietary

L-arabinose promotes expansion in superspreaders. *Cell Host Microbe* **31**, 405-417 (2023).

51. Neumann, M. *et al.* Deprivation of dietary fiber in specific-pathogen-free mice promotes susceptibility to the intestinal mucosal pathogen *Citrobacter rodentium*. *Gut Microbes* **13**, 1949-0976 (2021).

52. Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353 (2016).

53. An, J. *et al.* Western-style diet impedes colonization and clearance of *Citrobacter rodentium*. *PLoS Pathog.* **17**, e1009497 (2021).

54. Jimenez, A. G., Ellermann, M., Abbott, W. & Sperandio, V. Diet-derived galacturonic acid regulates virulence and intestinal colonization in enterohaemorrhagic *Escherichia coli* and *Citrobacter rodentium*. *Nat. Microbiol.* **5**, 368–378 (2020).

55. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

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

57. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

59. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

60. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

61. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).

62. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

63. Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: A next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).

64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

65. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

66. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

67. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).

68. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

69. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6**, 80–92 (2012).

70. Sarovich, D. S. & Price, E. P. SPANDx: A genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC Res. Notes* **7**, (2014).

71. Song, L., Florea, L. & Langmead, B. Lighter: fast and memory-efficient sequencing error

825 correction without counting. *Genome Biol.* **15**, 509 (2014).

826 72. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve
827 genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

828 73. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to
829 single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

830 74. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant
831 detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

832 75. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for
833 genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

834 76. Larsen, M. V. *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J.*
835 *Clin. Microbiol.* **50**, 1355–1361 (2012).

836 77. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of
837 large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

838 78. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*
839 **596**, 583–589 (2021).

840 79. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*
841 2021.10.04.463034 (2022).

842 80. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high
843 throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

844 81. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary
845 genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

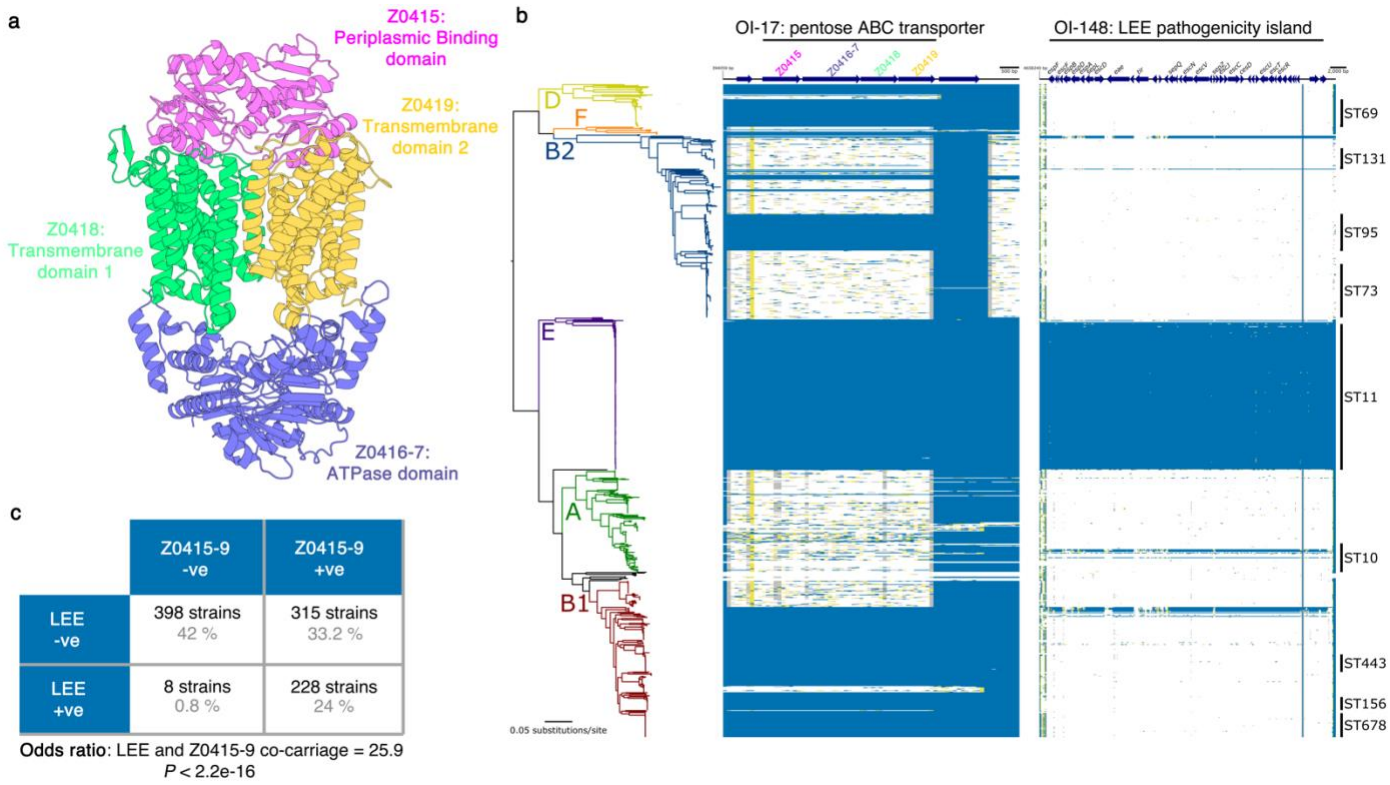


Fig. 1. A horizontally acquired pentose sugar ABC transporter widely encoded by enteric EHEC strains. a, AlphaFold2 model of the predicted Z0415-9 structure assembled at the EHEC inner membrane. **b**, Maximum likelihood analysis depicting the core-genome phylogeny of 949 *E. coli* isolates built from 245,518 core-genome single-nucleotide polymorphisms called against the reference chromosome EDL933. Phylogeny is rooted according to the actual root by *Escherichia fergusonii* ATCC 35469, which has been omitted for visualisation. Branch colours indicate the six main phylogenetic groups. Branch lengths and scale bar represent number of nucleotide substitutions per site. The presence/absence of Z0415-9 or the LEE is based on the uniform coverage at each 100 bp window size. Coverage is shown as a heat map where ≥80% identity is highlighted in blue, ≥50% identity is highlighted in yellow, and ≥1% is highlighted in grey. White plots indicate absent regions. **c**, 2 x 2 contigency matrix for carriage of the LEE and Z0415-9, based on the 949 genomes analysed in panbel **b**. A '+ve' indicates presence whereas a '-ve' indicates absence. The number of strains for each scenario and the

859 percentage distribution is given in black and grey respectively. Statistical significance of the

860 associated odd's ratio was calculated using a Fisher's exact test.

861

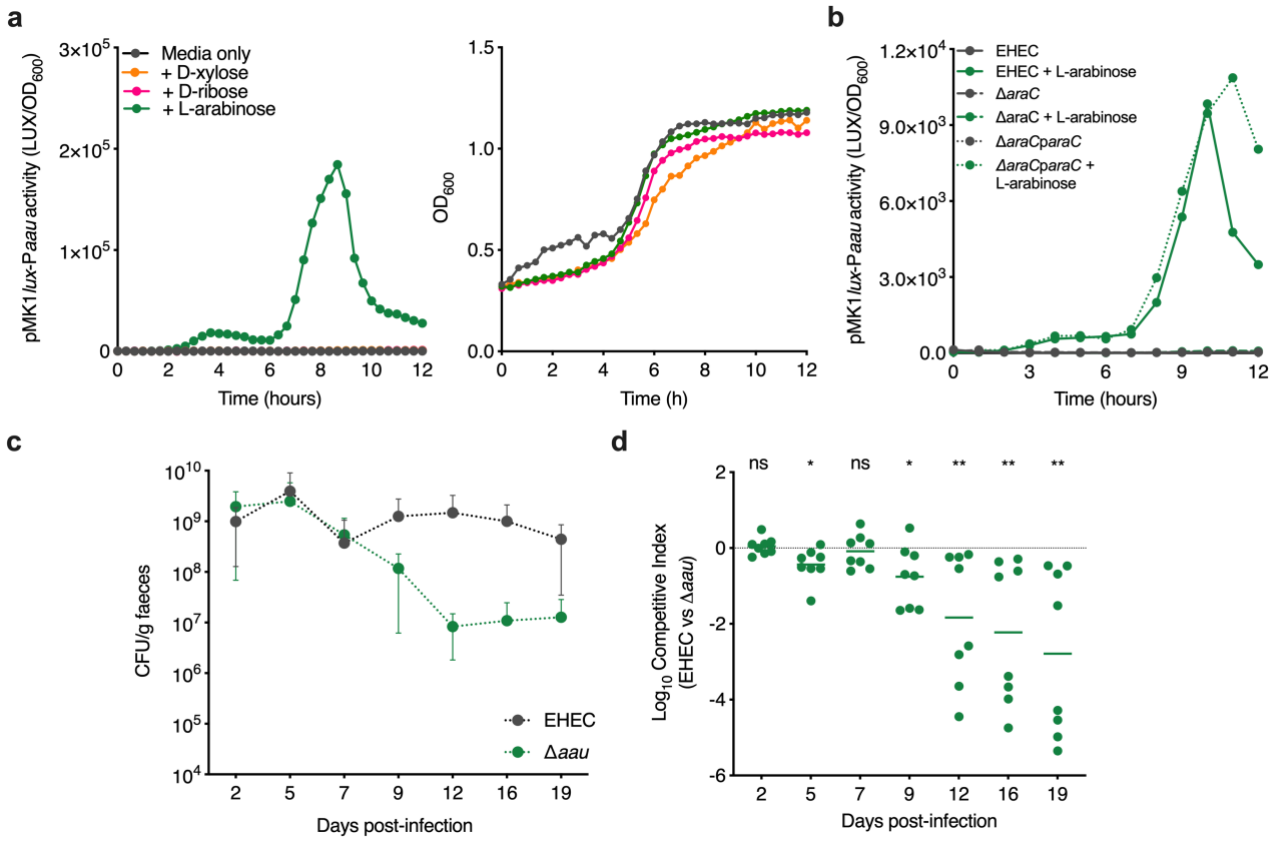


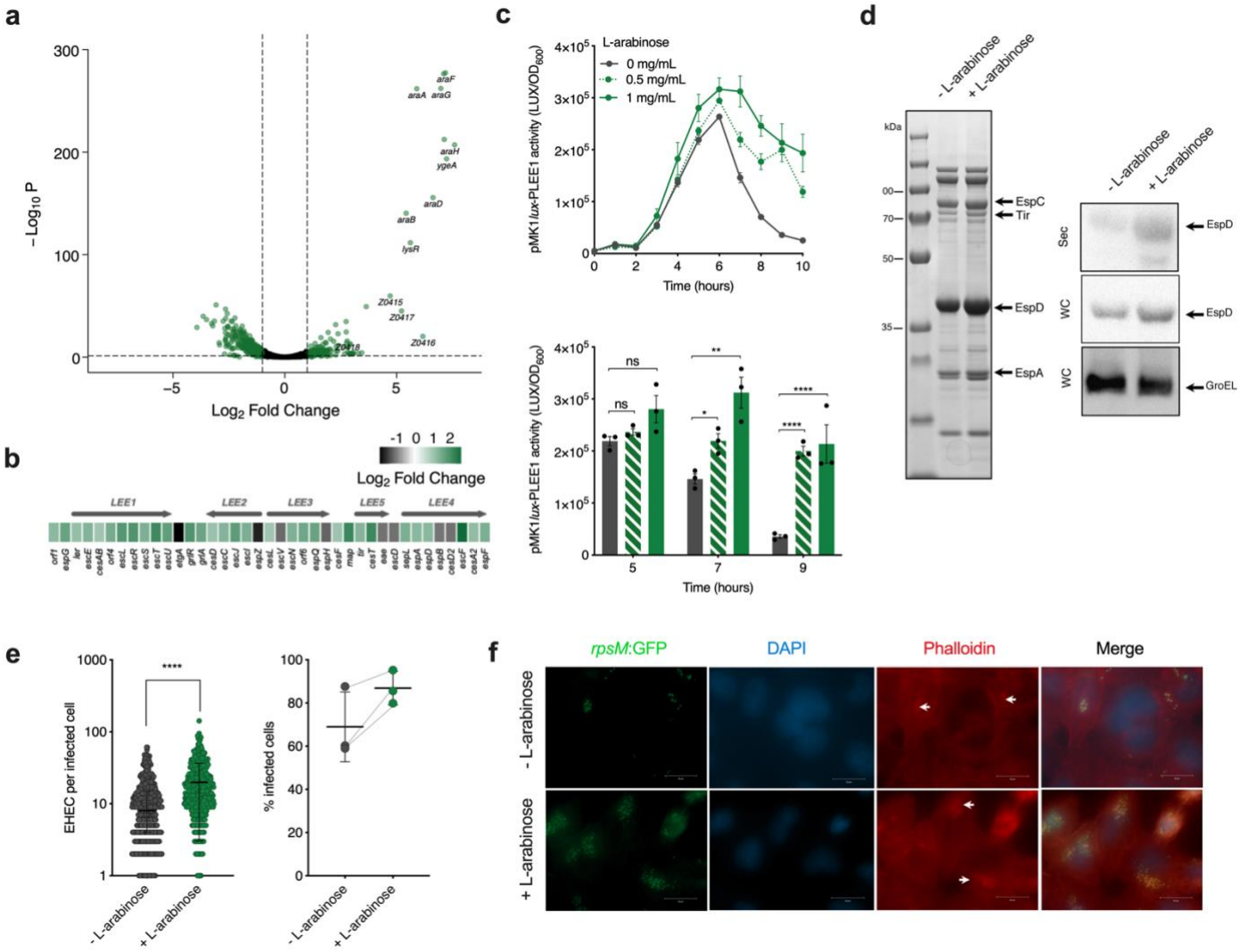
Fig. 2. Aau is induced by L-arabinose and required for EHEC fitness in the mouse gut. a, Transcriptional reporter assay of EHEC transformed with pMK1*lux*-P*aau* cultured in MEM-HEPES alone or supplemented with 0.5 mg/ml of the indicated sugar. Data are depicted as luminescence units (LUX) divided by optical density (OD_{600}) of the culture at each timepoint. The growth curve in the right panel was generated from the associated OD_{600} values. **b**, pMK1*lux*-P*aau* reporter assay as described in panel **a** utilising wild type (WT) EHEC, Δ*araC* and Δ*araC* + p*araC* derivative strains grown in MEM-HEPES alone or supplemented with L-arabinose. Graphs in panels **a** and **b** are representative of three independent repeats. **c**, Faecal shedding dynamics of Streptomycin-treated BALB/c mice (n = 8) colonised with a 1:1 mixture of EHEC and Δ*aau.* Error bars represent the standard deviation. **d**, Competitive index of EHEC versus Δ*aau* during colonisation of Streptomycin-treated BALB/c mice. Data points indicate the fold-decrease in Δ*aau* CFU recovered per faecal sample in comparison to wild

874 type EHEC. Statistical significance was determined by Wilcoxon signed-rank test. *, ** and

875 ns indicate $P < 0.05$, $P < 0.01$ or not significant respectively.

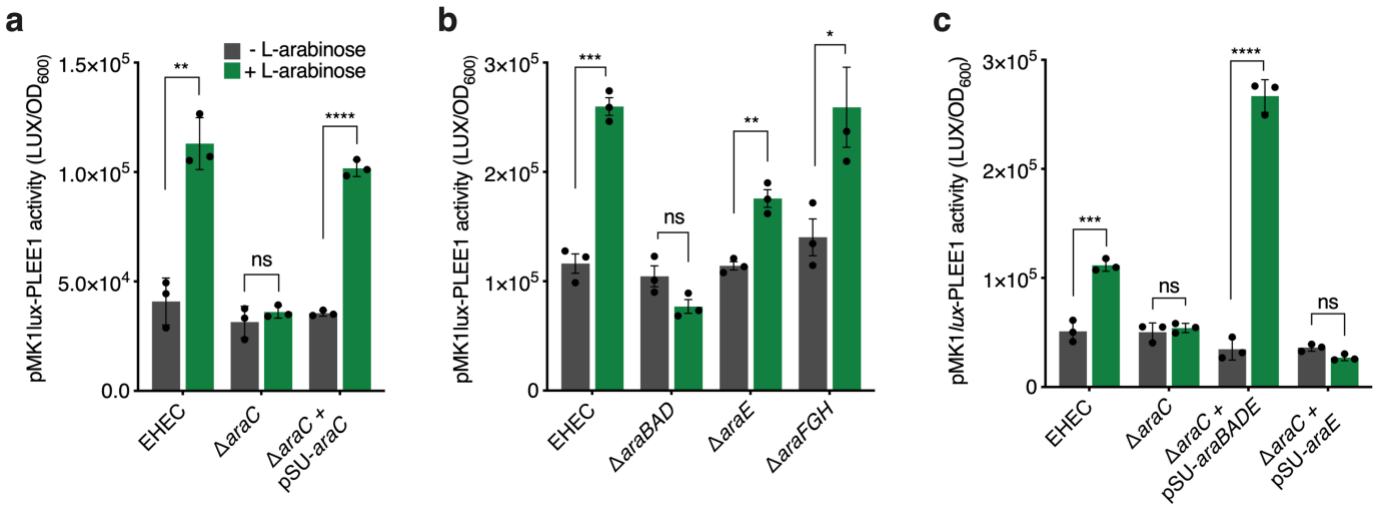
876

877



878 **Figure 3. L-arabinose enhances expression and function of the LEE-encoded T3SS. a**, Volcano
879 plot depicting gene expression patterns of EHEC cultured in MEM-HEPES supplemented with
880 L-arabinose relative to media alone as determined by RNA-seq. Significantly differentially
881 expressed genes (FDR-corrected $P \leq 0.05$) are highlighted in green with select genes related
882 to L-arabinose utilisation labelled. **b**, Expression profile of genes from the LEE island derived
883 from RNA-seq data presented in panel **a**. The scale bar indicates the Log^2 fold-change in
884 expression of genes from EHEC cells in response to L-arabinose supplementation. **c**,
885 Transcriptional reporter assay of EHEC transformed with pMK1*lux*-PLEE1 cultured in MEM-
886 HEPES alone or supplemented with 0.5 or 1 mg/ml of L-arabinose. Data are depicted as
887 luminescence units (LUX) divided by optical density (OD_{600}) of the culture at each timepoint.

888 The bar chart below indicates statistical significance at selected timepoints as determined
889 using a students' *t*-test. *, **, **** and ns indicate $P < 0.05$, $P < 0.01$, $P < 0.0001$ or not
890 significant respectively. Error bars represent standard error of the mean. **d**, SDS-PAGE
891 analysis of secreted proteins identified from cell-free supernatant of EHEC culture in
892 MEM-HEPES alone or supplemented with L-arabinose. The identity of known T3SS
893 associate effector proteins is labelled with arrows. Corresponding immunoblot analysis of
894 EspD (LEE-encoded) or GroEL levels identified in secreted fractions (Sec) in the cell-free
895 supernatant or whole cell pellet (WC). Data is representative of three independent
896 repeats. **e**, Data derived from widefield fluorescence microscopy analysis of HeLa cells
897 infected with EHEC cultured in MEM-HEPES alone (n = 531) or supplemented with L-
898 arabinose (n = 522). Quantification of the number of EHEC lesions per infected HeLa cell
899 was determined from 20 random fields of view. Experiments were performed on three
900 independent occasions. Statistical significance was determined using a Mann-Whitney U
901 test. **** indicates $P < 0.0001$. The graph on the right shows the average percentage of
902 infected HeLa cells determined from three independent experiments. Error bars represent
903 standard deviation. **f**, Representative images of HeLa cells infected with EHEC cultured
904 with and without L-arabinose. Channels are colour coded and labelled to indicate each
905 portion of the merged image. The white arrows indicate the presence of A/E lesions as
906 identified by areas of condensed actin that co-localise with attached EHEC cells.



907 **Figure 4. L-arabinose metabolism is required to enhance T3SS expression. a**, Transcriptional
908 reporter assay of wild type EHEC, Δ*araC* and Δ*araC* +pSU*araC* transformed with pMK1*lux*-
909 PLEE1 cultured in MEM-HEPES alone or supplemented with L-arabinose. Data are depicted as
910 luminescence units (LUX) divided by optical density (OD_{600}) of the sample. All samples were
911 taken at the same point in growth (hour 7). **b**, pMK1*lux*-PLEE1 reporter assays performed in
912 the EHEC, Δ*araBAD,* Δ*araE* and Δ*araFGH* backgrounds. **c**, pMK1*lux*-PLEE1 reporter assays
913 performed in the EHEC and Δ*araC* backgrounds. For complementation, Δ*araC* was
914 transformed with either pSU-*araBADE* or pSU-*araE*. *, **, ***, **** and ns indicate $P <$
915 0.05, $P < 0.01$, $P < 0.001$, $P < 0.0001$ or not significant respectively. Statistical significance
916 was determined using a student's *t*-test. Error bars represent standard error of the mean.
917

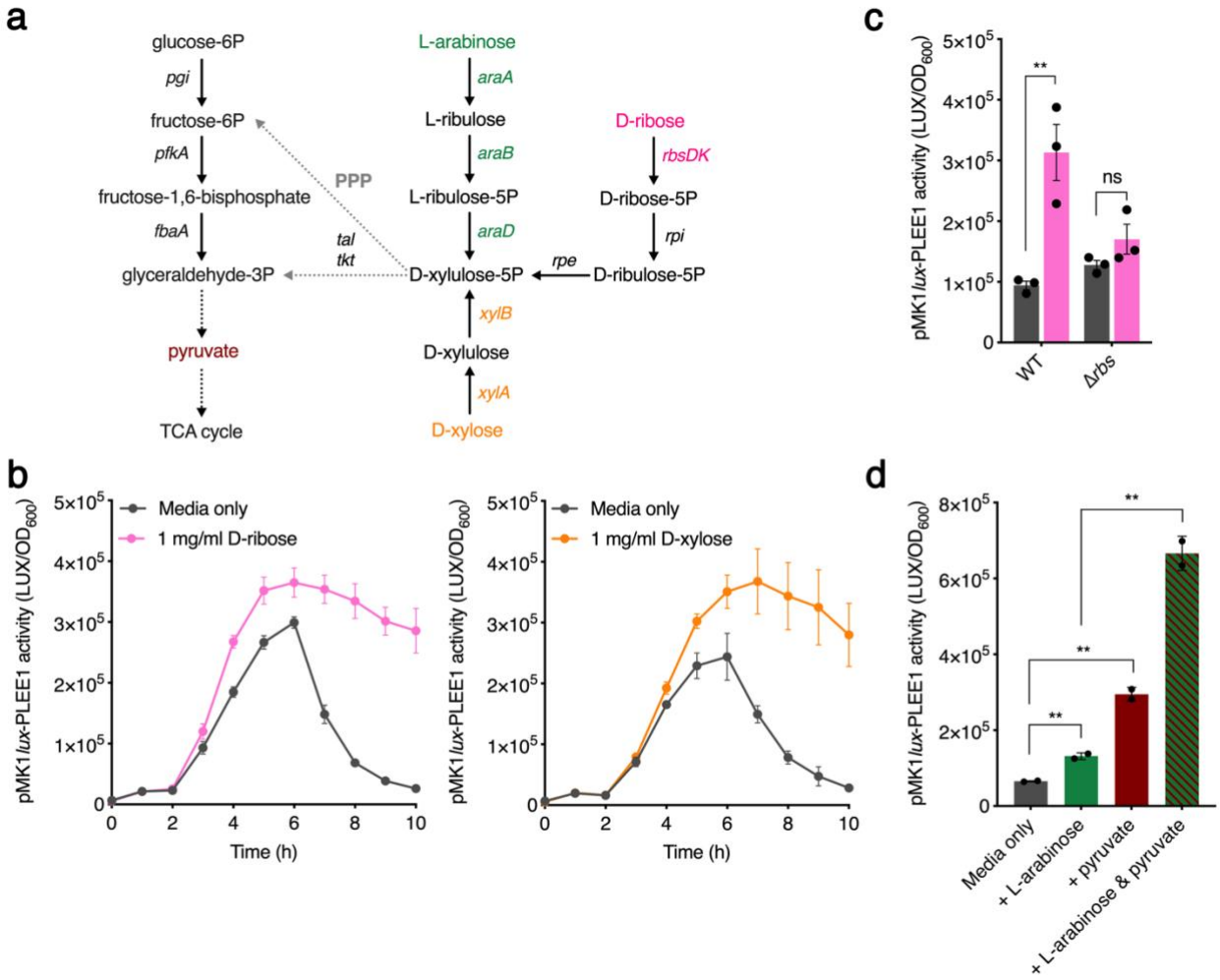
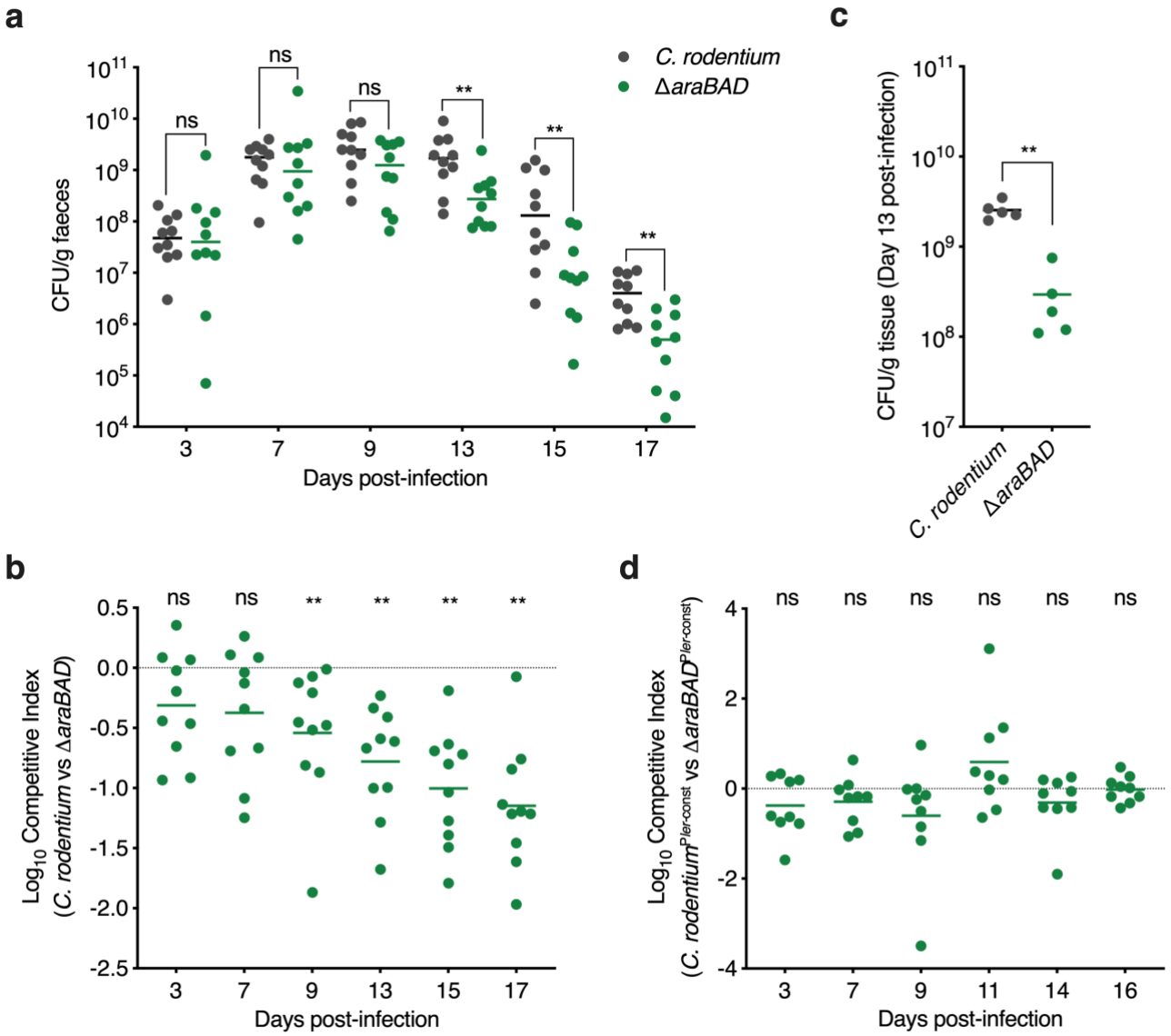


Figure 5. A generalised mechanism for T3SS regulation by pentose sugar metabolism. a, Schematic illustrating where the L-arabinose, D-xylose and D-ribose metabolic pathways converge with glycolysis (left hand side) via the Pentose Phosphate Pathway (PPP). The genes involved at each conversion are illustrated. Single steps are depicted as a solid arrow. Multiple steps are abbreviated to dotted arrows. **b**, Transcriptional reporter assay of EHEC transformed with pMK1/*lux*-PLEE1 cultured in MEM-HEPES alone or supplemented with 1 mg/ml of D-ribose or D-xylose. Data are depicted as luminescence units (LUX) divided by optical density (OD_{600}) of the culture at each timepoint. **c**, pMK1/*lux*-PLEE1 reporter assay of WT EHEC and the Δ*rbs* mutant cultured in MEM-HEPES alone (grey) or supplemented with D-ribose (pink). **d**, pMK1/*lux*-PLEE1 reporter assay of WT EHEC cultured in MEM-HEPES alone or supplemented with 1 mg/ml L-arabinose, 0.2 % pyruvate or a mixture of both. For panels **c**

929 and **d**, ** and ns indicate $P < 0.01$ or not significant respectively. Statistical significance

930 was determined using a student's *t*-test. Error bars represent standard error of the mean.

931



932 **Figure 6. L-arabinose metabolism promotes *C. rodentium* fitness *in vivo* by enhancing T3SS**

933 **expression. a**, Faecal shedding dynamics of BALB/c mice (n = 10) orally infected with either

934 *C.rodentium* or Δ*araBAD*. Data points represent the CFU/ml for individual mice as determined

935 from faecal pellets at each timepoint. Statistical significance was determined by Mann-

936 Whitney U test. ** and ns indicate $P < 0.01$ or not significant respectively.

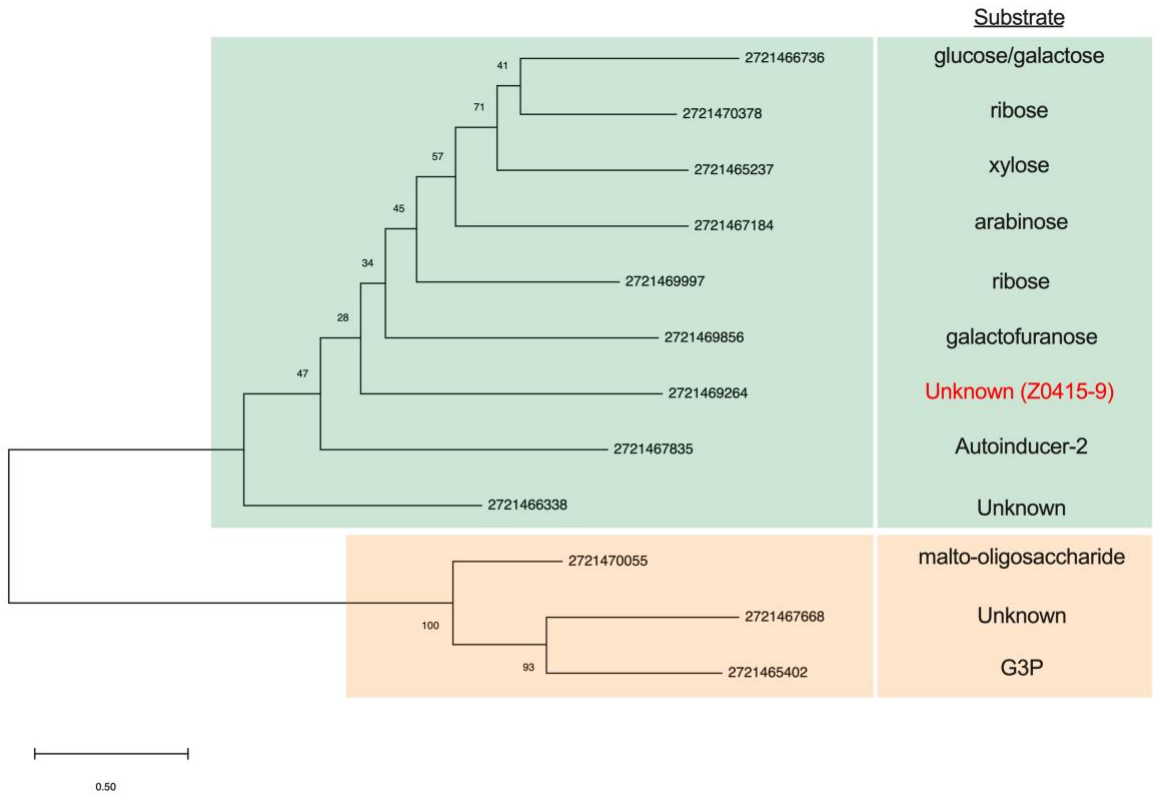
937 **b**, Competitive index of *C.rodentium* versus Δ*araBAD* during infection of BALB/c mice (n =

938 10). Mice were orally infected with a 1:1 mixture of both strains. Data points indicate the fold-

939 decrease in Δ*araBAD* CFU recovered per faecal sample in comparison to wild type *C.*

940 *rodentium*. Statistical significance was determined by Wilcoxon signed-rank test. ** and

941 ns indicate $P < 0.01$ or not significant respectively. **c**, CFU per gram of colon tissue from
942 mice (n = 5) orally infected with either *C.rodentium* or Δ*araBAD.* Statistical significance was
943 determined by Mann-Whitney U test. ** indicates $P < 0.01$. **d**, Competitive index of
944 *C.rodentium*$^{P\textit{ler}\text{-const}}$ versus Δ*araBAD*$^{P\textit{ler}\text{-const}}$ during infection of BALB/c mice (n = 9). Mice were
945 orally infected with a 1:1 mixture of both strains. ns indicates not significant as determined
946 by a Wilcoxon signed-rank test.



947

948 **Extended Data Fig. 1. Phylogenetic analysis of sugar-specific ABC transporters in EDL933.**

949 Phylogeny was inferred using the maximum-likelihood method and Le Gascuel model, with a

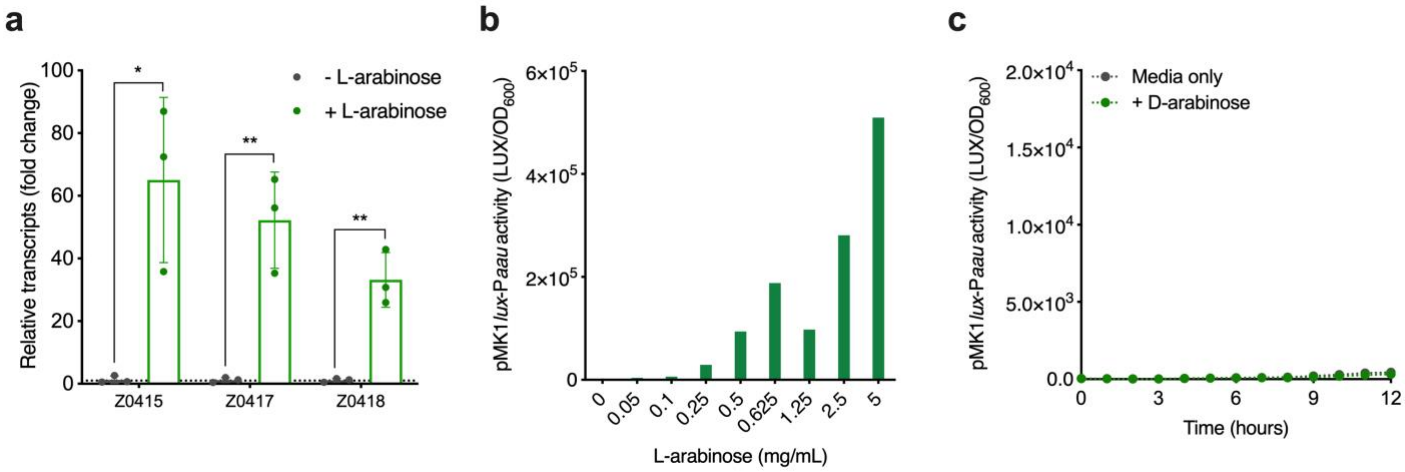
950 Gamma distribution in MegaX. The tree is drawn to scale, with branch lengths measured in

951 the number of substitutions per site. The scale bar represents 0.5 substitutions per site.

952 Bootstrap values are indicated on the respective branches. Clades A and B are coloured green

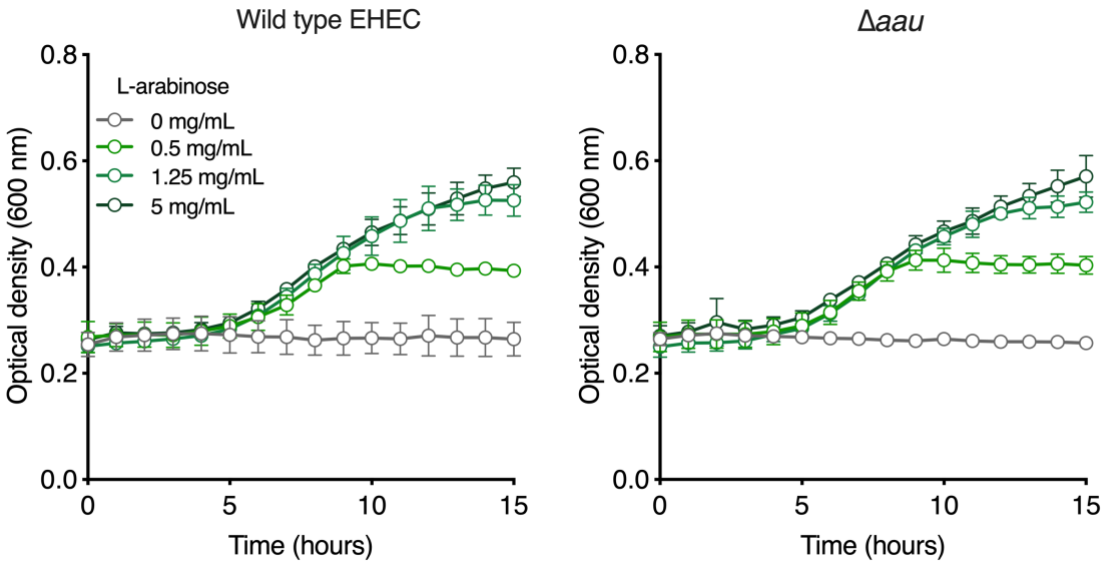
953 and orange respectively. The predicted substrate for each transporter is indicated next to its

954 respective branch in the tree.



955 **Extended Data Fig. 2. The L-isomer of arabinose induces Z0415-9 in a concentration**
956 **dependent manner. a,** RT-qPCR analysis of Z0415/7/8 expression in RNA-derived from EHEC
957 cultured in MEM-HEPES alone or supplemented with L-arabinose. The bars indicate relative
958 fold-increase of L-arabinose treated cultures over media alone. The dotted line indicates
959 baseline expression in comparison to the control. Statistical significance was determined by
960 Student's *t* test. * and ** indicate $P < 0.05$ or $P < 0.01$ respectively. Error bars represent
961 standard deviation from three independent replicates. **b**, Transcriptional reporter assay of
962 EHEC transformed with pMK1/*lux*-P*aau* cultured in MEM-HEPES alone or supplemented with
963 a range of L-arabinose concentrations. Data are depicted as luminescence units (LUX) divided
964 by optical density (OD_{600}) of the culture after 8 hours of growth. **c,** Transcriptional reporter
965 assay of EHEC transformed with pMK1/*lux*-P*aau* cultured in MEM-HEPES alone or
966 supplemented with 5 mg/ml D-arabinose. Data from panels **b** and **c** represent a single
967 replicate from three independent repeats of each experiment.

968



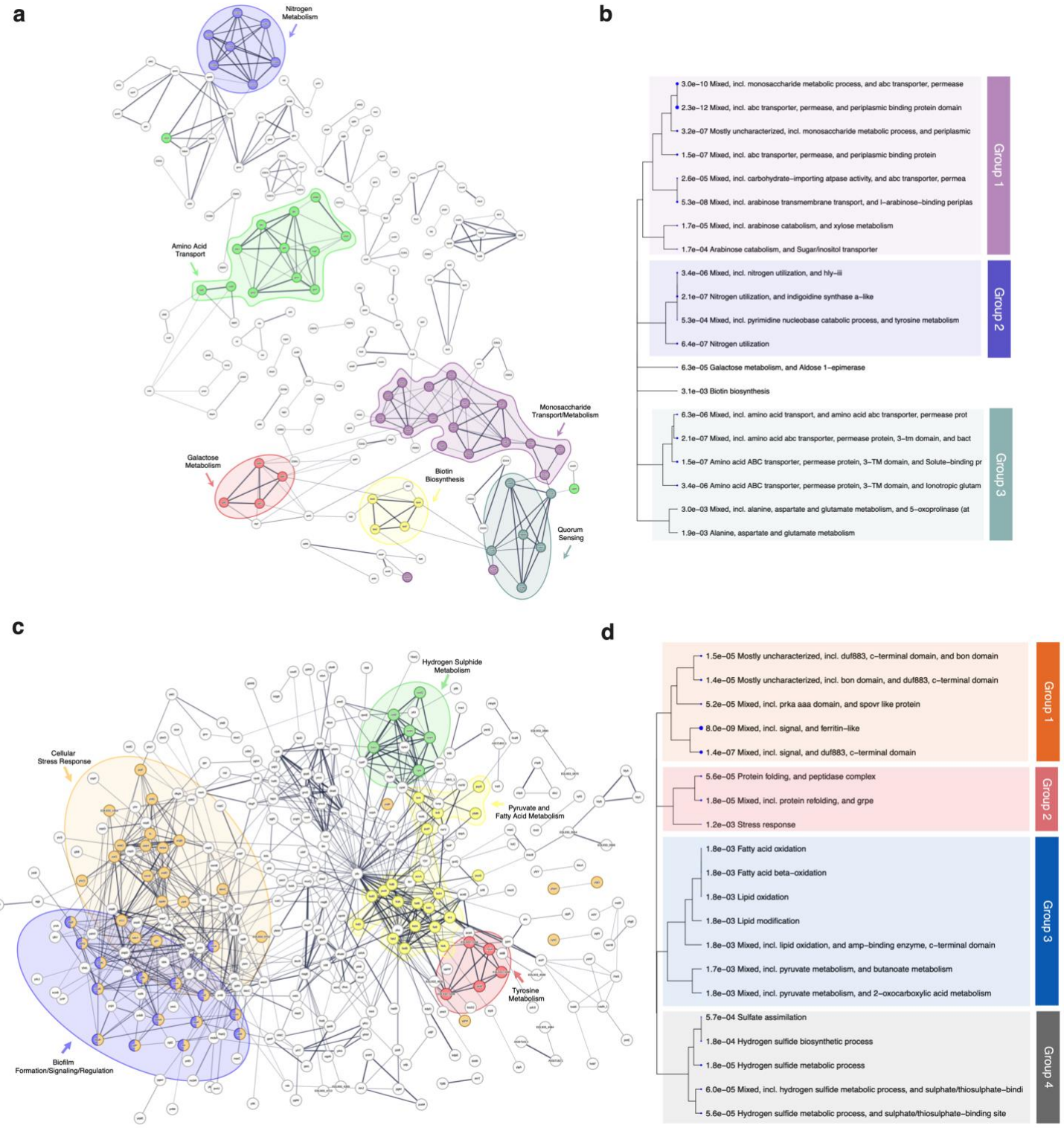
969

970 **Extended Data Fig. 3. Aau is not required for growth on L-arabinose as a sole carbon source.**

971 Growth analysis of parental EHEC and associated Δ*aau* deletion mutant in M9 minimal media

972 supplemented with various concentrations of L-arabinose as a sole carbon source. The error

973 bars indicate standard deviation based on three independent repeats.



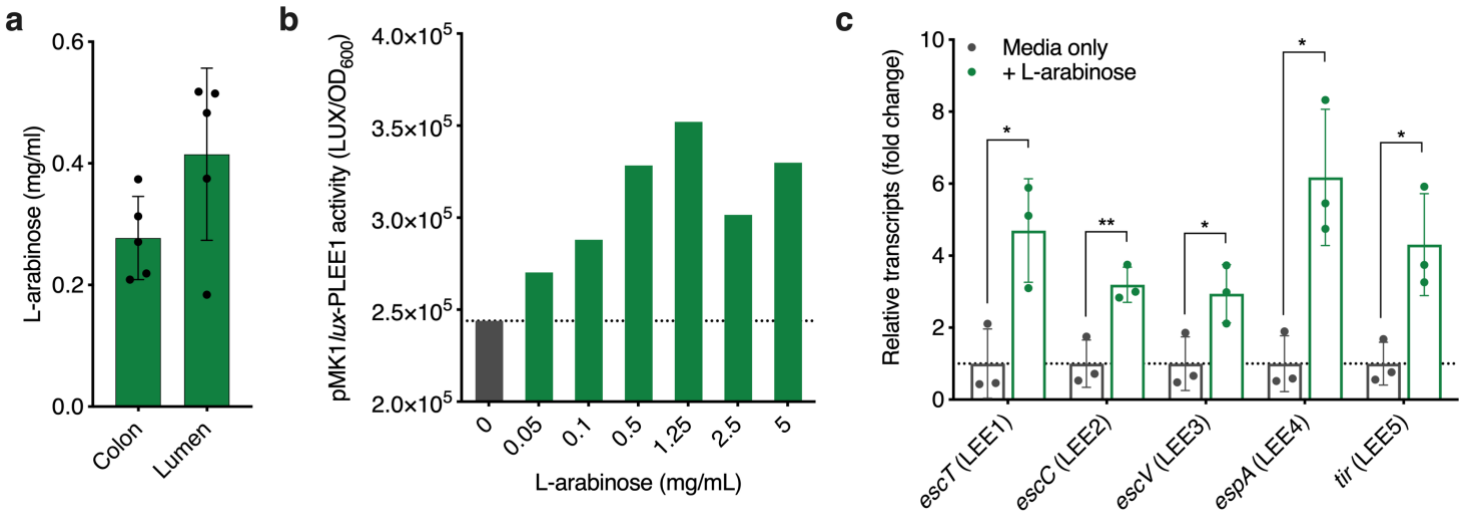
974 **Extended Data Fig. 4. L-arabinose induces a global transcriptional shift in EHEC. a,c**, STRING

975 network analysis of predicted protein-protein interactions based upon upregulated or

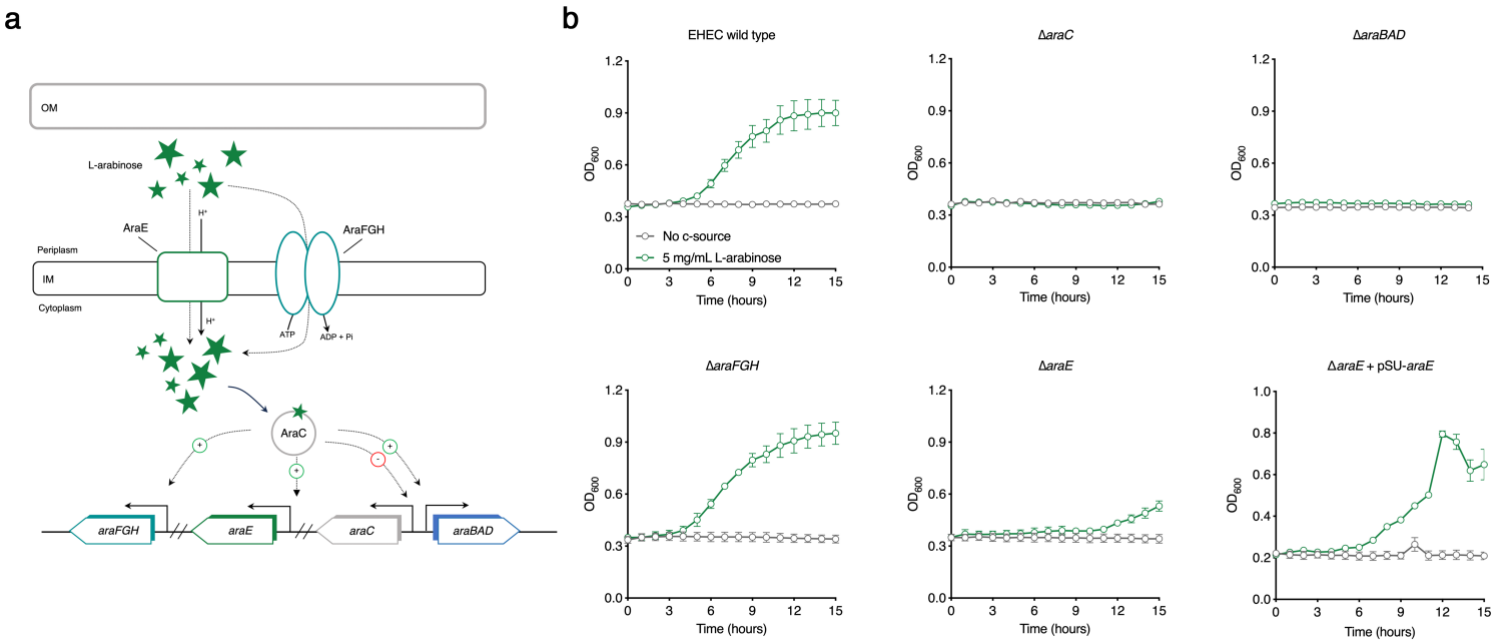
976 downregulated genes identified by RNA-seq analysis of EHEC cultured in MEM-HEPES

977 supplemented with L-arabinose. Line thickness is indicative of confidence in the interactions

between proteins. Disconnected nodes were removed from the network to improve clarity. Distinct functional clusters that are enriched are in coloured bubbles. **b,d**, Hierarchical clustering tree summarising the correlation among significant STRING pathways generated using ShinyGO. Pathways that are enriched with shared genes are grouped together. Larger dots indicate more significant *P*-values.

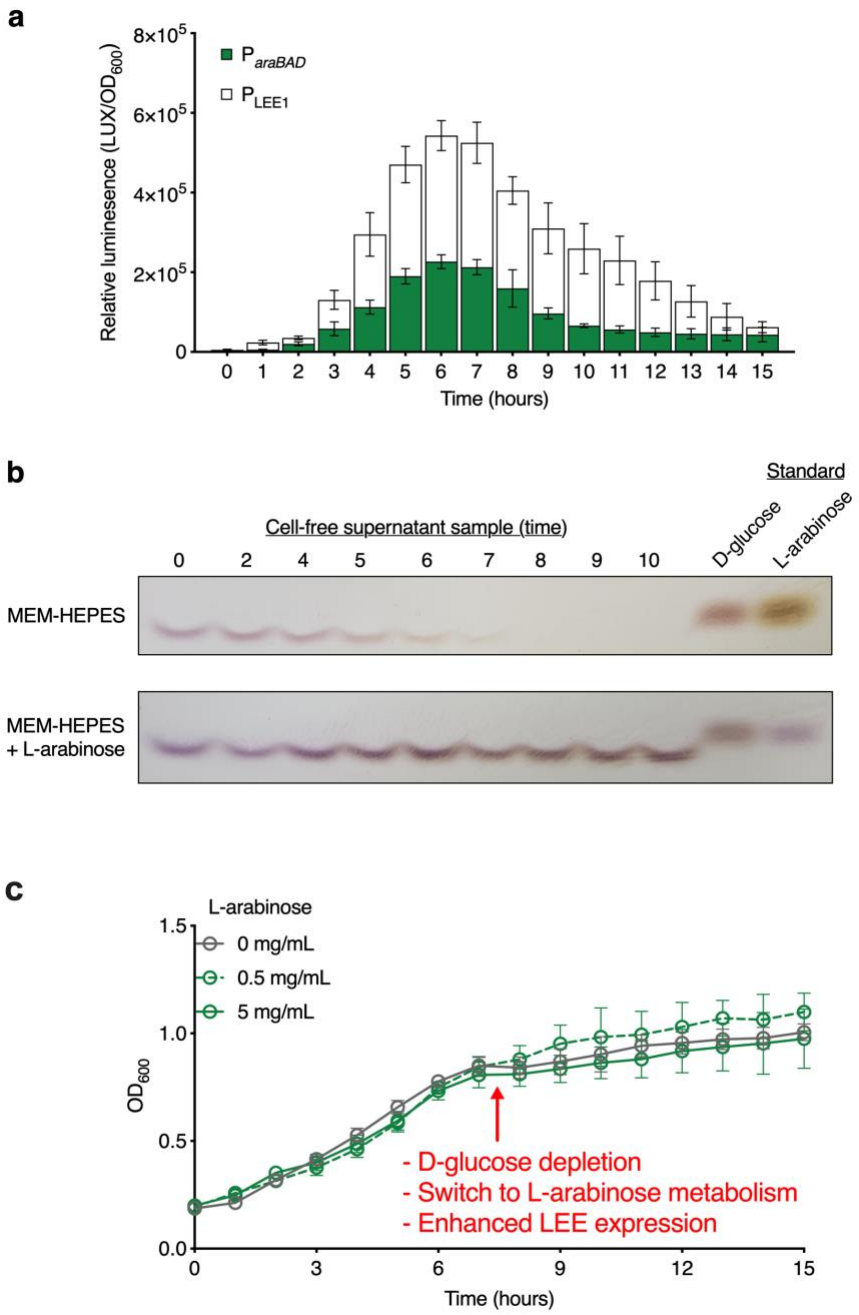


984 **Extended Data Fig. 5. L-arabinose induces LEE expression at concentration relevant to the**
985 **gut. a,** Quantification of L-arabinose from colon tissue and luminal content of BALB/c mice
986 maintained on a conventional diet (n = 5). **b,** Transcriptional reporter assay of EHEC
987 transformed with pMK1/*lux*-P*LEE1* cultured in MEM-HEPES alone or supplemented with a
988 range of L-arabinose concentrations. Data are depicted as luminescence units (LUX) divided
989 by optical density (OD_{600}) of the culture after 8 hours of growth. The data represent a single
990 replicate from three independent repeats of the experiment. **c,** RT-qPCR analysis of relative
991 LEE1-5 expression in RNA-derived from EHEC cultured in MEM-HEPES alone or supplemented
992 with L-arabinose. The bars indicate relative fold-increase of L-arabinose treated cultures over
993 media alone. The dotted line indicates baseline expression in comparison to the control.
994 Statistical significance was determined by Student's *t* test. * and ** indicate $P < 0.05$ or $P <$
995 0.01 respectively. Error bars represent standard deviation from three independent
996 replicates.
997



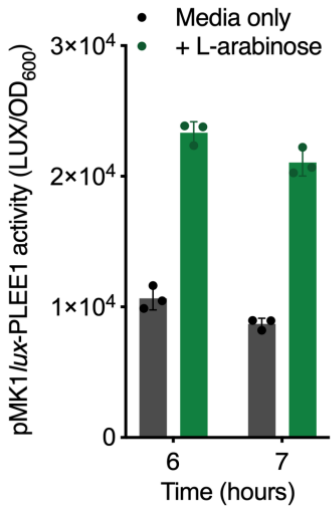
998 **Extended Data Fig. 6. Growth analysis of mutants from each stage of the L-arabinose**
999 **utilisation pathway in EHEC. a,** Schematic illustration of the canonical pathway for L-
1000 arabinose uptake, sensing and catabolism within the cell. **b,** Growth curves of wild type EHEC
1001 and the indicated deletion mutants in M9 minimal media with no carbon source (grey,
1002 control) or supplemented with 5 mg/ml L-arabinose (green). Experiments were performed on
1003 three independent occasions and the error bars represent the standard deviation.

1004



Extended Data Fig. 7. L-arabinose metabolism and enhanced LEE expression are co-ordinated in EHEC. a, Transcriptional reporter assay of EHEC transformed with either pMK1/*lux*-P*LEE1* or pMK1/*lux*-P*araBAD* cultured in MEM-HEPES supplemented with L-arabinose. The experiments were performed under identical growth conditions and samples taken in parallel. The data are depicted as luminescence units (LUX) divided by optical density (OD_{600}) of the culture at each timepoint. Error bars indicate standard error of the mean from

1012 three independent repeats. **b,** Thin layer chromatography analysis of cell-free supernatant
1013 derived from cultures of EHEC grown in MEM-HEPES alone or supplemented with 1 mg/ml L-
1014 arabinose. Samples were taken at the indicated timepoints and the standards for D-glucose
1015 and L-arabinose are positioned on the right. The result represents a single replicate from three
1016 independent experiments. **c,** Representative growth curve of EHEC grown in MEM-HEPES
1017 alone or supplemented with various concentrations of L-arabinose. The red arrow indicates
1018 the point at which D-glucose is depleted from the media (derived from panel **b**) and the
1019 phenotypic switch associated with this event.
1020



Extended Data Fig. 8. L-arabinose induces LEE expression in *C.rodentium*. Transcriptional reporter assay of wild type *C. rodentium* transformed with pMK1*lux*-PLEE1 cultured in MEM-HEPES alone or supplemented with L-arabinose. Data are depicted as luminescence units (LUX) divided by optical density (OD_{600}) of the sample at various timepoints. Error bars represent standard error of the mean derived from three independent experiments.