MICROBIOLOGY SOCIETY

A novel *hemA* mutation is responsible for a small–colony-variant phenotype in *Escherichia coli*

Alasdair T. M. Hubbard, Issra Bulgasim and Adam P. Roberts*

Abstract

We identified a small colony variant (SCV) of an amoxicillin/clavulanic acid-resistant derivative of a clinical isolate of *Escherichia coli* from Malawi, which was selected for *in vitro* in a subinhibitory concentration of gentamicin. The SCV was auxotrophic for hemin and had impaired biofilm formation compared to the ancestral isolates. A single novel nucleotide polymorphism (SNP) in *hemA*, which encodes a glutamyl-tRNA reductase that catalyses the initial step of porphyrin biosynthesis leading to the production of haem, was responsible for the SCV phenotype. We showed the SNP in *hemA* resulted in a significant fitness cost to the isolate, which persisted even in the presence of hemin. However, the phenotype quickly reverted during sequential sub-culturing in liquid growth media. As *hemA* is not found in mammalian cells, and disruption of the gene results in a significant fitness cost, it represents a potential target for novel drug development specifically for the treatment of catheter-associated urinary tract infections caused by *E. coli*.

INTRODUCTION

Small colony variants (SCVs) are slower growing mutants observed in the study of multiple bacterial species. They have altered, smaller than normal, colony morphology, which is usually due to mutations within genes involved in metabolic processes, for example in genes involved in the electron transport chain of *Staphylococcus aureus* [1]. SCVs have been isolated from multiple infection sites, including tissue, bone and device-associated infections [2] and can result in increased resistance to aminoglycosides [3, 4], which can also be used to select for SCVs *in vitro* [5]. The mechanism of decreased susceptibility to aminoglycosides has not been extensively investigated but in *Staphylococcus aureus* and *Escherichia coli* it has been suggested to be due to reduced uptake caused by alterations in membrane potential [6, 7]. Investigations of SCVs of *E. coli* are relatively rare but have also been associated with mutations within genes encoding proteins involved in the electron transport chain, including *yigP* [4], *hemB* [8] and *lipA* [9]. SCVs of *E. coli* have been isolated from urinary tract infections [10], bacteraemia [11] and chronic hip infections [8] and are therefore increasingly considered clinically relevant [12].

An important enzyme involved in the electron transport chain is HemA. HemA is a glutamyl-tRNA reductase that catalyses the initial step of porphyrin biosynthesis leading to the production of haem, which is required for respiration. Single nucleotide polymorphisms (SNPs) identified within *hemA* have been found to result in a SCV in several bacterial genera, including *Salmonella* [3, 13] and *Listeria* [14, 15], as well as in *E. coli* [16]. The morphology of three SCVs of *Listeria monocytogenes,* all with a common SNP present in *hemA* resulting in early termination of the protein, were rescued when grown in the presence of exogenous hemin, suggesting they were auxotrophic for haem [15].

Here we report the identification of a novel SNP in *hemA* resulting in a SCV phenotype of an amoxicillin/clavulanic acid (AMC)-resistant derivative of clinical isolate of *E. coli* from Malawi, which was selected for *in vitro* in a subinhibitory concentration of gentamicin. We assessed the effect of this SNP on the fitness of the SCV compared to the ancestral isolates and, as biofilm-producing *E. coli* are common etiological agents of catheter-associated UTIs [17], we sought to determine the biofilm-forming ability of the SCV.

METHODS

Bacterial strains, media and antibiotics

E. coli 10129 is a susceptible clinical isolate from Malawi isolated from cerebral spinal fluid [18] and *E. coli* 10129 LB_2 is an *in vitro* derived isolate with resistance to amoxicillin/ clavulanic acid (AMC) [19]. Both isolates were previously whole-genome sequenced and are available in GenBank under the following accession numbers: assembly of *E. coli* 10129; SIJF00000000 and sequencing reads of *E. coli* 10129 LB_2; SAMN10963734.

All bacterial cultures were initially grown on LB (Lennox) agar (Sigma, UK) for 18 h at 37 °C then subcultured into cation-adjusted Mueller–Hinton Broth (CA-MHB), LB (Lennox) broth (LB) (both Sigma, UK) or Iso-sensitest broth (ISO, Oxoid, UK) and incubated at 37 °C, 200 r.p.m. for 18 h unless otherwise stated.

Gentamicin were solubilized in molecular grade water (all Sigma, UK) and filter sterilized through a 0.22 µM polyethersulfone filter unit (Millipore, USA).

Selection of gentamicin-resistant colonies

An initial culture of *E. coli* 10129 LB_2 was diluted 1/1000 in LB plus 1 µg ml^{-1} gentamicin and incubated for 24 h at 37 °C, 200 r.p.m. Following incubation, the culture was serially diluted 1 in 10 to the 10^{-2} dilution and 50 µl plated out on LB agar plus 16 µg ml^{-1} gentamicin and incubated at 37 °C for 18 h. A SCV, designated *E. coli* 10129 LB_2.6 and a normal sized colony, designated *E. coli* 10129 LB_2.5 were immediately stored at −80 °C.

Antimicrobial susceptibility testing

MIC of gentamicin towards *E. coli* 10129, *E. coli* 10129 LB_2 and *E. coli* 10129 2.6 were determined using the microdilution broth method following the CLSI guidelines [20] using CA-MHB and performed in triplicate. An additional MIC of gentamicin towards *E. coli* 10129 LB_2.6 was also performed in triplicate supplemented with 20 µg ml^{-1} hemin (Sigma, UK).

Characterization of colony and cell morphology

E. coli 10129, *E. coli* 10129 LB_2, *E. coli* 10129 LB_2.5 and *E. coli* 10129 LB_2.6 were all grown on LB agar, both with and without supplementation of 20 µg ml^{-1} hemin and CHROMagar Orientation agar for 18 h at 37 °C in triplicate then photographed under a Celestron MicroDirect 1080 p HD Handheld Digital microscope (Celestron, USA). Gram stain of each isolate was performed using a standard protocol (Sigma, UK) in triplicate and visualized under a Motic Trinocular light microscope mounted with a Moticam X Lite camera (Motic, Hong Kong).

Stability of the small-colony phenotype

E. coli 10129, *E. coli* 10129 LB_2, *E. coli* 10129 LB_2.5 and *E. coli* 10129 LB_2.6 were grown in 10 ml LB for 18 h at 37 °C, 200 r.p.m. Following growth, 10 µl was passaged in to fresh 10 ml LB and incubated at 37 °C, 200 r.p.m. for 24 h and repeated for a total of 3 days. Each culture on each day was subcultured on to LB agar and incubated for 18 h at 37 °C, the colonies on each plate where photographed under a Celestron MicroDirect 1080 p HD Handheld Digital microscope.

Whole-genome sequencing and bioinformatics analysis

Short-read (2×250 bp paired end) genome sequencing of *E. coli* 10129 LB_2.6 using the Illumina MiSeq platform, and subsequent read trimming, was provided by MicrobesNG (http://www.microbesng.uk). The genome was *de novo* assembled using SPAdes (v3.12.0 [21]) and annotated using Prokka (v1.12 [22]). The sequencing reads of *E. coli* 10129 LB_2.6 were aligned to the annotated assembled genomes of *E. coli* 10129 and *E. coli* 10129 LB_2 to identify statistically relevant SNPs, indels, duplications and amplifications using Breseq (v0.32.0 [23]). Protein structure predictions were carried out using I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [24] and comparisons carried out using Tm-align (https://zhanglab.ccmb.med.umich.edu/TM-align/) [25]. Protein models used in structural comparisons were selected based on the highest C-score and TM-score.

DNA extraction

DNA extraction was performed using the PureGene Yeast/ Bact kit B (Qiagen, Germany) following the manufacturer's instructions for extraction of DNA from Gram-negative bacteria and eluted in molecular grade water.

PCR

All PCR was performed using 0.02 U/µl Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA), 1×Q5 Reaction Buffer, 0.5 µM primers and 200 µM dNTPs in a total volume of 25 µl and, unless otherwise stated, all PCR products were purified using the Monarch PCR and DNA Clean-up Kit (New England Biolabs, USA) following the manufacturer's instructions and eluted in molecular grade water unless stated.

Amplification of 819 bp of *hemA* was performed the primers HemA_F (TGTCGACGTGTAACCGCACA) and HemA_R, (CCACAGCAGCAGCTTTCCGTTG) and the following protocol: denaturation at 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 70 °C for 30 s and elongation at 72 °C for 50 s, followed by a single, final extension of 2 min at 72 °C. The entire 1257 bp of *hemA* from *E. coli* 10129 was amplified using the primers HA1_F (ATGACCCTTTTAGCACTCGG) and HA1_R (CTACTC-CAGCCCGAGGCT) in a total volume of 50 µl and the following protocol: denaturation at 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 66 °C for 30 s and elongation at 72 °C for 13 s, followed by a single, final extension step 72 °C for 2 min. A band at approximately 1 kb was excised and DNA extracted using the Monarch DNA Gel Extraction Kit (New England Biolabs, USA) following the manufacturer's instructions. The restriction sites for the restriction enzymes KpnI were added to the 5′ ends of the

1257 bp double-stranded amplicon by a second round of PCR, using the primers KpnIHA_F1 (AAAAAGGTACCATGACC-CTTTTAGCACTCGG) and KpnIHA_R1 (AAAAAGGTACC CTACTCCAGCCCGAGGCT) (KpnI site is underlined). Amplification was performed using the same protocol as the initial amplification of *hemA*.

Construction of complement plasmid

The resulting PCR product of the PCR amplifying the wild-type *hemA* with added KpnI sites and the recipient plasmid pHSG396 were digested with the restriction enzyme KpnI (New England Biolabs, USA). PCR clean-up was then performed using the Monarch PCR and DNA Clean-up Kit following the manufacturer's instructions. The digested *hemA* PCR product and pHSG396 were ligated together using T4 ligase (New England Biolabs, USA), at a 3:1 insert to vector mass ratio, incubated at room temperature for 10 min followed by 65 °C for 10 min.

The ligated, recombinant plasmid (2 μl of the ligation mix) was transformed into NEB 5-alpha competent *E. coli* (New England Biolabs, USA) using the following protocol: the reaction was incubated on ice for 30 min, then heat shocked at 42 °C for 30 s. Following a second incubation on ice for 5 min, 950 μl of SOC outgrowth medium was added and incubated at 37 °C, 250 r.p.m. for 1 h. The transformed cells were plated out onto LB agar supplemented with 35 μg ml⁻¹ chloramphenicol and Isopropyl β-D-1-thiogalactopyranoside (IPTG)/X-gal (Fisher Scientific, USA) and incubated at 37 °C for 18 h and the plasmid was extracted using the Monarch Plasmid Miniprep Kit (New England Biolabs, USA) following the manufacturer's instructions. *hemA* in pHSG396 was Sanger sequenced by GeneWiz (Takely, UK) using commercially available M13 primers. *E. coli* 10129 LB_2.6 was made competent following the protocol set out in Chung *et al.* [26] and pHSG396 and pHSG396:*hemA* was transformed following the above protocol. *E. coli* 10129 LB_2.6 with pHSG396 and *E. coli* 10129 LB_2.6 with pHSG396:*hemA* were both plated out on to LB agar supplemented with 35 μg ml⁻¹ chloramphenicol and 1 mM IPTG in triplicate and photographed under a Celestron MicroDirect 1080 p HD Handheld Digital microscope.

Biofilm assay

Biofilm production by *E. coli* 10129, *E. coli* 10129 LB_2 and *E. coli* 10129 LB_2.6 was quantified as described previously [19]. Briefly, each isolate was initially grown in 10 ml LB, then diluted 1/1000 in M9 [50% (v/v) M9 minimal salts (2×) (Gibco, ThermoFisher Scientific, USA), 0.4% D-glucose, 4 mM magnesium sulphate (both Sigma, UK) and 0.05 mM calcium chloride (Millipore, USA)] with and without supplementation with 20 μg ml⁻¹ hemin. Four technical replicates of each diluted culture were added to a 96-well plate and incubated statically for 18 h at 37 °C. Following incubation, each well was washed with PBS, stained with 0.1% crystal violet and washed again with PBS. The stain was liberated from the 96-well plate with 30% acetic acid (Fisher Scientific, USA) and measured at an OD_{550}. Three biological replicates of this assay

were performed. Biofilm production of *E. coli* 10129 LB_2.6 with pHSG396 and *E. coli* 10129 LB_2.6 with pHSG396:*hemA* was performed as above, except using M9 supplemented with 35 μg ml⁻¹ chloramphenicol and 1 mM IPTG.

Competitive and comparative fitness assays

Competitive fitness of *E. coli* 10129 LB_2.6 was assessed compared to *E. coli* 10129 and *E. coli* 10129 LB_2 in LB and M9 as described previously [19]. Briefly, each culture was diluted to an OD_{600} of 0.1 and then diluted 1/1000 and combined 1:1 in the appropriate media and 150 μl added to a 96-well plate and incubated at 37 °C, 200 r.p.m. for 18 h. Bacterial density of the initial combined culture and after 24 h incubation was enumerated by diluting the culture in PBS and 50 μl of each dilution plated out on LB agar and LB agar plus 20 μg ml⁻¹ gentamicin and incubated at 37 °C for 18 h. The Malthusian equation was used to calculate competitive relative fitness [27]:

$$M = M_m \left(\frac{T_{24}}{T_0} \right) / M_{wt} \left(\frac{T_{24}}{T_0} \right)$$

$W = (M-1) \times 100,$

where W represents relative fitness, M represents the Malthusian parameter, M_{wt} represents the Malthusian parameter of the ancestor isolate, M_m represents the Malthusian parameter of the resistant isolate, T_0 represents log c.f.u. ml⁻¹ count of the initial culture and T_{24} represents the log c.f.u. ml⁻¹ count after 24 h.

Fitness of *E. coli* 10129 LB_2.6 with pHSG396:*hemA* was compared to *E. coli* 10129 LB_2.6 with pHSG396 comparatively. Each culture was diluted to an OD_{600} of 0.1 then further diluted 1/1000 in LB supplemented with 35 μg ml⁻¹ chloramphenicol and 1 mM IPTG. Of each diluted culture, 150 μl was added to a 96-well plate in duplicate. As a negative control, 150 μl LB with 35 μg ml⁻¹ chloramphenicol and 1 mM IPTG was also added to the plate in duplicate. All samples were incubated at 37 °C for 24 h, orbital shaking at 200 r.p.m., in a Clariostar Plus microplate reader (BMG Labtech, Germany) measuring the OD_{600} every 10 min. Relative fitness of *E. coli* 10129 LB_2.6 with pHSG396:*hemA* was compared to *E. coli* 10129 LB_2.6 with pHSG396 using BAT v2.1 [28] with absorbance values between 0.02 and 0.2 and a minimum R value of 0.9902. Both fitness assays were performed in triplicate.

Statistical analysis

Statistical analysis of the competitive fitness and biofilm assays were performed using the ordinary one-way ANOVA plus uncorrected Fisher's LSD test and statistical analysis of comparative fitness and biofilm assay of the complemented *E. coli* 10129 LB_2.6 isolate was performed using the unpaired t-test (two-tailed), all in GraphPad Prism (v8.4.0).

RESULTS

Selection and isolation of a small-colony variant

In a previous study, AMC-resistant derivatives of the clinical isolate *E. coli* 10129 were selected for *in vitro* in LB in the
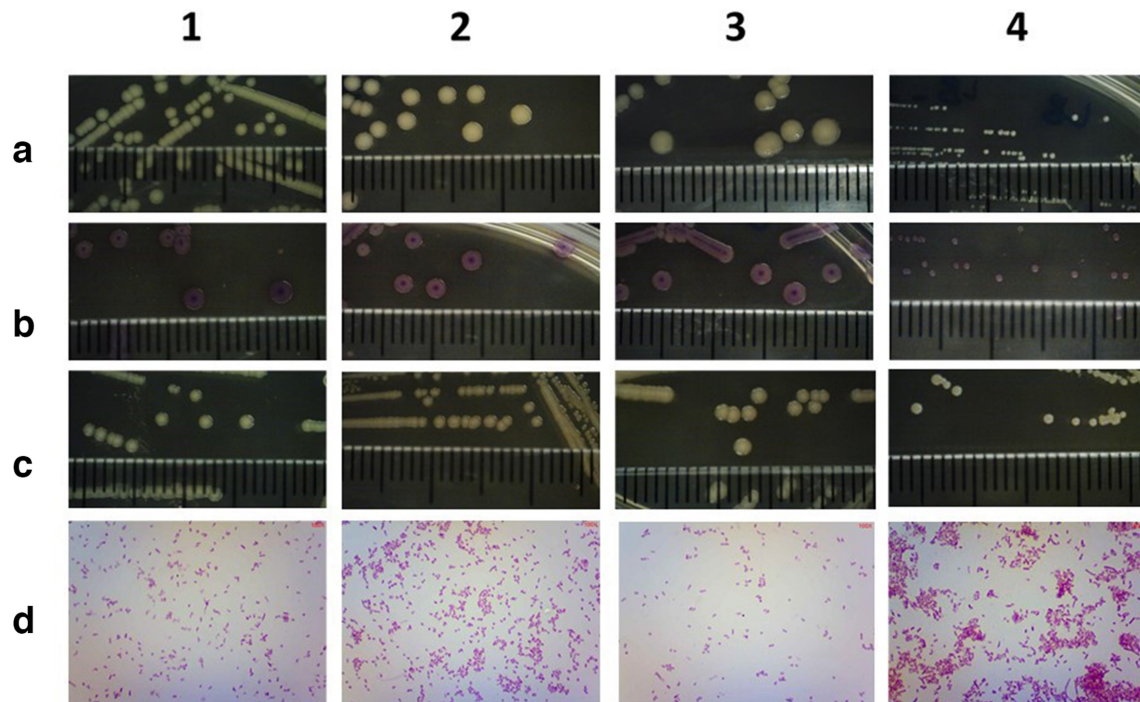
Fig. 1. Photographs of the colonies of (1) *E. coli* 10129, (2) *E. coli* 10129 LB_2, (3) *E. coli* 10129 LB_2.5 and (4) *E. coli* 10129 LB_2.6 (SCV) grown on (a) LB agar, (b) CHROMagar Orientation agar, (c) LB agar plus 20 µg ml⁻¹ hemin (lines on scale are 1 mm increments) and (d) Gram stain of each of the isolates (100x magnification). Each photograph is a representative image of three independent replicates (Fig S1 and S2).

presence of a sub-inhibitory concentration of AMC [19]. One such derivative, *E. coli* 1029 LB_2, showed high-level resistance to AMC and was found to contain a SNP in the promoter region of AmpC [19]. In this study, we selected for gentamicin-resistant derivatives of *E. coli* 10129 LB_2 *in vitro* via growth in the presence of a sub-inhibitory concentration of gentamicin in LB, during which we identified a small, pinprick colony on the agar plate, which was immediately stored at −80 °C and designated *E. coli* 10129 LB_2.6. Subsequent regrowth of *E. coli* 10129 LB_2.6 confirmed the SCV morphology persisted when compared to the two ancestral strains (*E. coli* 10129 and LB_2) and another normal sized isolate, *E. coli* 10129 LB_2.5, from the same selective plate (Fig. 1a). The SCV was confirmed to be Gram-negative, however the cell size and shape was not observed to be different than that of the two ancestral isolates or the *E. coli* 10129 LB_2.5 strain (Fig. 1d). To ensure that the SCV was not a Gram-negative contaminant, we confirmed it was *E. coli* through growth on CHROMagar Orientation agar (Fig. 1b). The MIC for gentamicin for *E. coli* 10129 LB_2.6 and *E. coli* 10129 LB_2.5 was 8 µg ml⁻¹ compared to 1–2 µg ml⁻¹ for *E. coli* 10129 and 2–4 µg ml⁻¹ *E. coli* 10129 LB_2.

Mutations identified within the genome of *E. coli* 10129 LB_2.6

By comparing the whole-genome sequence of the *E. coli* 10129 LB_2.6 to the ancestral isolates *E. coli* 10129 and *E. coli* 10129 LB_2, we identified a single novel SNP present in *hemA* in

E. coli 10129 LB_2.6 (T>G at *hemA* nucleotide 426 out of 1257). The SNP was predicted to result in a non-synonymous mutation leading to a change of phenylalanine to leucine atHemA amino acid position 142. This was confirmed to be present in *E. coli* 10129 LB_2.6 but absent in *E. coli* 10129, *E. coli* 10129 LB_2 *and E. coli* 10129 LB_2.5 by PCR and Sanger sequencing of *hemA* in all four isolates (Fig. S7).

Structural changes are predicted in the HemA from the SCV compared to the wild-type

The phenylalanine to leucine change at amino acid 142 is predicted to be within an α-helix that links two globular domains within the tertiary structure of the proteins, the relative positions of which are predicted to change (Fig. 2). This is predicted to change the position of the substrate binding sites for both NAD and glutamyl-tRNA, as determined by Tm-align.

Confirmation of the role of hemA in the small-colony-variant phenotype

As HemA is involved in the biosynthetic pathway of haem, we hypothesized that *E. coli* 10129 LB_2.6 would be auxotrophic for haem due to the mutation in *hemA*. Therefore, we supplied hemin exogenously during growth on agar. We found that there was a reversion of the SCV phenotype to normal size of the *E. coli* 10129 and *E. coli* 10129 LB_2, therefore this mutation is likely to be the cause of the SCV phenotype (Fig. 1c).
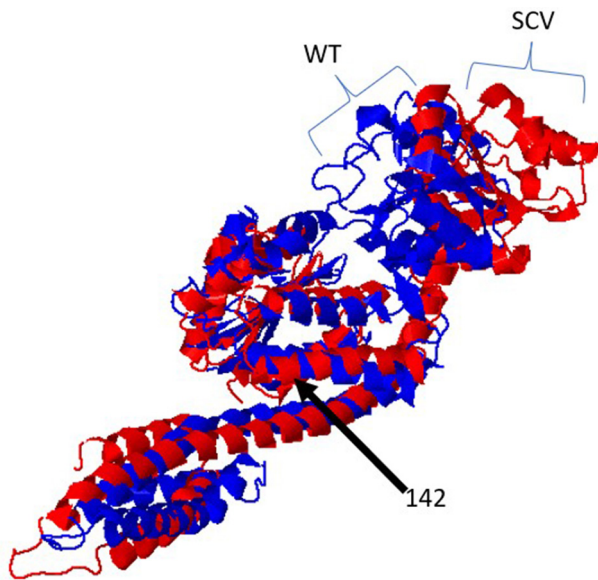
Fig. 2. Structural alignment of WT (blue) and SCV (red) HemA. The changed amino acid at position 142 is indicated within an α-helix. The differential predicted position of the tertiary structure of the HemA is indicated with brackets.

Supplementation of 20 μg ml⁻¹ hemin also decreased the MIC of gentamicin for *E. coli* 10129 LB_2.6 from 8 μg ml⁻¹ to 4 μg ml⁻¹; comparable to that of the ancestor isolate *E. coli* 10129 LB_2 (2–4 μg ml⁻¹), suggesting the SNP in *hemA* is also involved in the susceptibility to gentamicin. To confirm this, we rescued the SCV phenotype by transforming a plasmid (pHSG396) with and without the cloned *hemA* from *E. coli* 10129 into *E. coli* 10129 LB_2.6. Using IPTG to induce expression of *hemA* in pHSG396, we found a reversion to a 'normal' colony size in *E. coli* 10129 LB_2.6::pHSG396:*hemA*, while *E. coli* 10129 LB_2.6::pHSG396 did not (Fig. 3, S3 and S4), therefore the SNP at position amino acid position 142 in *hemA* was confirmed to be responsible for a SCV phenotype in *E. coli*.

Stability of the small-colony-variant phenotype

We sought to determine whether *E. coli* 10129 LB_2.6 could revert back to a 'normal' sized colony after multiple generations. We found that the SCV phenotype returned to
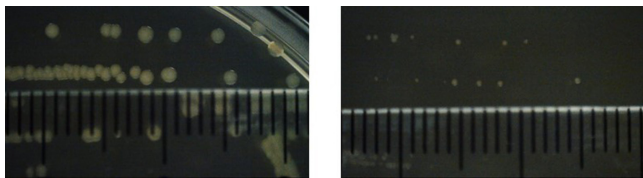

Fig. 3. Photographs of the colonies of *E. coli* 10129 LB_2.6 transformed with (a) pHSG396:*hemA* and (b) pHSG396. Each photograph is a representative image of three independent replicates (Fig S3 and S4). Lines on scale are 1 mm increments.

'normal' sized colony phenotype after a small number (1–3) of overnight passages and therefore the SCV phenotype was not stable (Figs 4, S5 and S6).

Fitness cost of the SNP in *hemA*

The effect of the mutation and subsequent SCV phenotype on fitness compared to the two ancestral isolates *E. coli* 10129 and *E. coli* 10129 LB_2 was assessed competitively. We found that *E. coli* 10129 LB_2.6 carried a significant fitness cost in both LB and ISO compared to *E. coli* 10129 (LB; *P*-value = <0.0001, ISO; *P*-value=0.0026) and *E. coli* 10129 LB_2 (LB; *P*-value = <0.0001, ISO; *P*-value=0.0003) and, despite reversion to a 'normal' sized colony on LB agar with hemin, the fitness cost imposed on *E. coli* 10129 LB_2.6 compared to *E. coli* 10 129 (LB; *P*-value=0.0787, ISO; *P*-value=0.3727) and *E. coli* 10129 LB_2 (LB; *P*-value=0.0784, ISO; *P*-value=0.2317) did not show a statistically significant difference when exogenous hemin was available (Fig. 5a; F statistic 3.539). To confirm that the SNP in *hemA* was responsible for this observed fitness cost, we measured the growth rate of both *E. coli* 10129 LB_2.6 containing pHSG396:*hemA*, induced by IPTG, and *E. coli* 10 129 LB_2.6 containing the empty pHSG396 plasmid in LB plus ITPG and chloramphenicol. We then calculated the relative fitness by comparing the growth rate of *E. coli* 10129 LB_2.6 with pHSG396 to itself, which therefore has a relative fitness of 1, and to *E. coli* 10129 LB_2.6 with pHSG396:*hemA*. Here, we found the expression of the wild-type *hemA* present in pHSG396 increased the relative fitness of *E. coli* 10129 LB_2.6 significantly to 1.709 (*P*-value=0.0422) compared to the pHSG396 without *hemA* (Fig. 5b). Therefore, the SNP in *hemA* was responsible for the significant fitness cost seen in *E. coli* 10129 LB_2.6.

Biofilm production of *E. coli* 10129 LB_2.6

Biofilm production was significantly impaired in *E. coli* 10129 LB_2.6 compared to the immediate ancestor *E. coli* 10129 LB_2 (*P*-value=0.0189) when assessed in M9. However, when biofilm production was assessed in M9 supplemented with hemin none of the isolates produced a biofilm (Fig. 6a). Comparison of biofilm production of *E. coli* 10129 LB_2.6 containing pHSG396 to *E. coli* 10129 LB_2.6 contain pHSG396:*hemA*, we found both produced very little biofilm production and no significant difference between the two isolates.

DISCUSSION

Previous studies have found that SCVs of different bacterial genera have either been developed following selection in, or show resistance to, aminoglycosides [3–5] and in this study we have identified a novel SNP in *hemA* in an isolate of *E. coli,* which has resulted in a SCV phenotype following selection in gentamicin. The uptake of aminoglycosides is dependent on the membrane potential and often requires respiration. Previously, Lazar *et al.* linked a SNP in *hemA* with resistance to aminoglycosides due to disruption of the respiratory transport chain, subsequently altering the membrane potential,
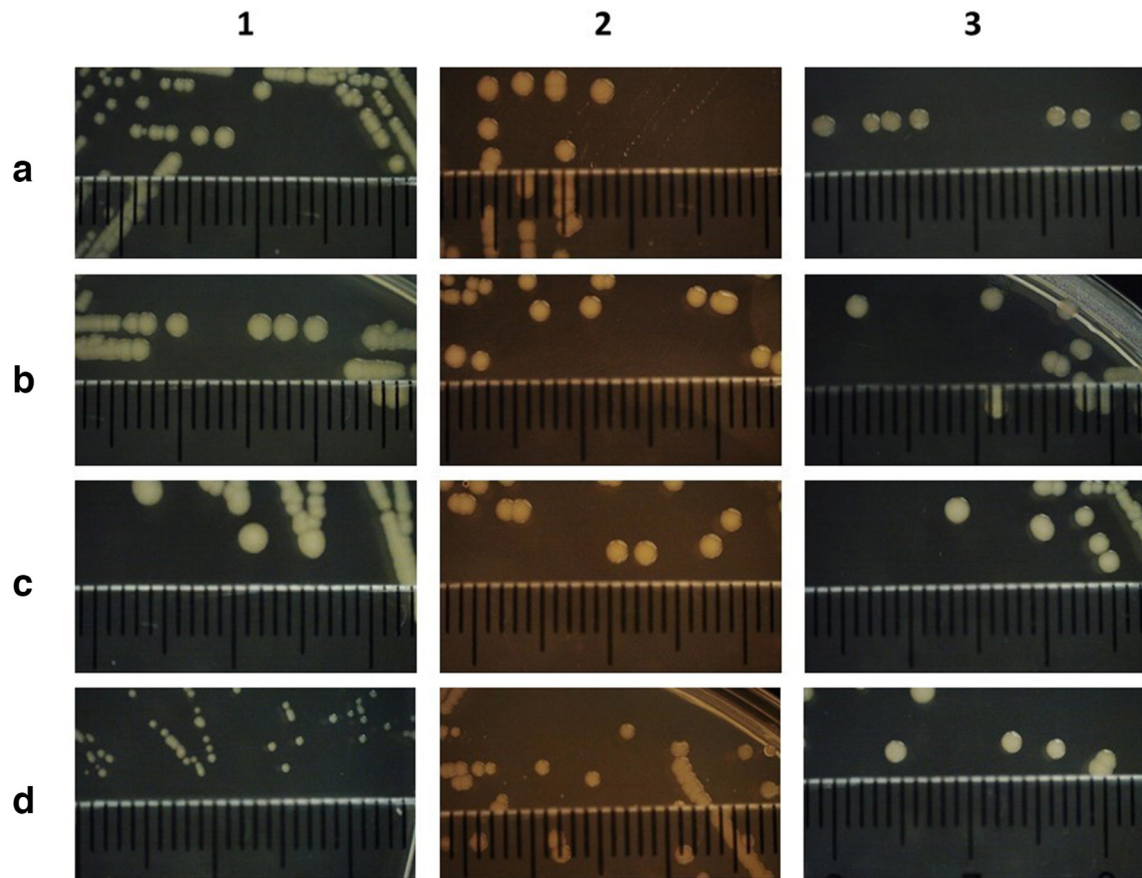
Fig. 4. Photographs of the colonies of (a) *E. coli* 10129, (b) *E. coli* 10129 LB_2 and (c) *E. coli* 10129 LB_2.5 and (d) *E. coli* 10129 LB_2.6 grown on LB agar after 1, 2 and 3 passages. Each photograph is a representative image of three independent replicates (Fig S5 and S6). Lines on scale are 1 mm increments.
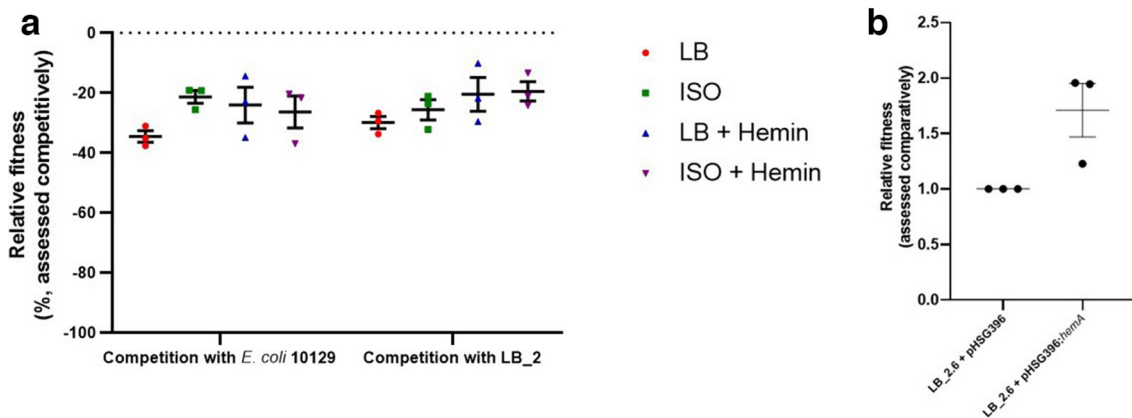

Fig. 5. (a) Relative fitness of *E. coli* 10129 LB_2.6 (SCV) to the immediate ancestor *E. coli* 10129 LB_2 and *E. coli* 10129 in ISO and LB, with and without addition of 20 μg ml^{-1} hemin assessed competitively. (b) Relative fitness of *E. coli* 10129 LB_2.6 with pHSG396:*hemA* compared to *E. coli* 10129 LB_2.6 with pHSG396, assessed comparatively.
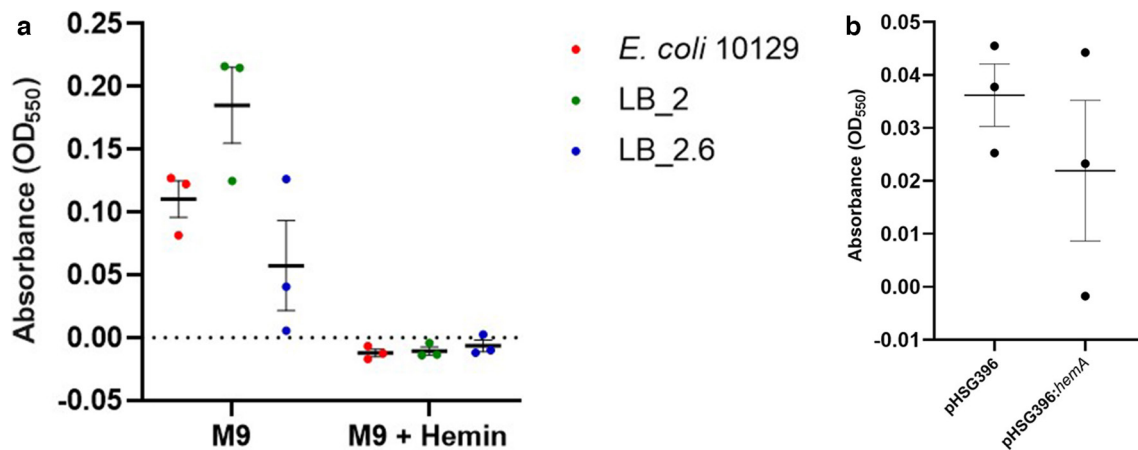
Fig. 6. Biofilm production of (a) *E. coli* 10129, *E. coli* 10129 LB_2 and *E. coli* 10129 LB_2.6 (SCV) in M9 and M9 supplemented with 20 µg ml^{-1} hemin and (b) *E. coli* 10129 LB_2.6 containing pHSG396 or pHSG396:*hemA*.

in *E. coli* [7]. In this study we found that the addition of exogenous hemin reduced the gentamicin MIC of the SCV comparable to that of the ancestor isolate, *E. coli* 10129 LB_2, suggesting that the novel SNP in *hemA* influences resistance to gentamicin. Previous SNPs in *hemA* resulting in the SCV phenotype have been identified, which resulted in amino acid changes at position valine-45-glycine [3, 13], serine-232-stop [3, 13] and glutamine-225-stop [14, 15], with the latter two resulting in an early termination of the protein. The SNP we identified in this *E. coli* isolate resulted in an amino acid substitution at position 142 and did not lead to a premature stop. Mutations in *hem* genes often result in the SCV becoming auxotrophic for hemin [15, 29, 30], and the SNP we identified in *hemA* was confirmed to result in a haem auxotroph phenotype, which also carried a significant fitness cost, which persists even in the presence of exogenous hemin.

Sequential subculture in liquid media showed that the SCV reverted to wild-type colony phenotype with 1–3 of the subcultures suggesting that compensatory mutations are likely occurring resulting in a fitness benefit compared to the SCV. As the SCV is so unfit compared to the ancestral strain any compensatory mutation resulting in even a modest improvement in fitness would quickly be selected in our assay and this can explain the variation in the colony sizes and time of emergence between the replicate evolution experiments (Figs 4, S5 and S6).

The SCV also has impaired biofilm formation compared to the ancestral isolates, which again did not alter in the presence of exogenous hemin. However, in the presence of hemin, all isolates tested in this study were unable to form biofilms. Therefore, while our results indicate that the SNP in *hemA* may affect biofilm formation in the absence of hemin it appears that hemin itself may also interfere with *E. coli* biofilm formation, as has been previously shown for *Porphyromonas gingivalis* [31]. Attempts to determine the biofilm-forming

capability of a complemented strain, plus a strain carrying only the expression plasmid, resulted in little or no biofilm being formed. Recently the effect of IPTG, which is used to induce production of HemA in our complemented strains, has been shown to negatively affect plasmid stability in, and viability of, planktonically growing *E. coli* cells [32]. This makes interpretation of our data difficult as the transition from planktonic to biofilm-growing cells is exactly when we need to induce expression of *hemA* with IPTG to determine its effect on biofilm formation in our complemented strains. Further work will be aimed at determining if complementation will result in wild-type biofilm formation using an alternative expression system.

Even though HemA is non-essential for growth, it has been identified as a potential drug target against *Acinetobacter baumannii* as it is not present in mammalian cells [33]. As the mutation in *hemA* significantly affects the fitness and biofilm production of this *E. coli* isolate, HemA may have target potential for novel compounds, which could reduce fitness of *E. coli*, leading to them being outcompeted by a resident microbiome, and reduced biofilm formation, which, for *E. coli*, are a common cause of catheter-associated UTIs [17].

Conclusions

A novel SNP in *hemA* that emerged upon *in vitro*, sub-inhibitory gentamicin selection of a clinical isolate of *E. coli* results in a SCV phenotype, which is auxotrophic for haem and carries a significant fitness cost even in the presence of exogenous hemin, The SCV also has significantly impaired biofilm production relative to the ancestral isolate, opening up therapeutic possibilities with the development of *hemA* inhibitors.

Data availability

SPAdes assembly of *E. coli* 10129 LB_2.6 was submitted to GenBank under the accession number JAAUVJ000000000.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Vestergaard M, Paulander W, Leng B, Nielsen JB, Westh HT *et al*. Novel pathways for ameliorating the fitness cost of gentamicin resistant small colony variants. *Front Microbiol* 2016;7:7.

2. Kahl BC, Becker K, Löffler B. Clinical significance and pathogenesis of staphylococcal small colony variants in persistent infections. *Clin Microbiol Rev* 2016;29:401–427.

3. Koskiniemi S, Pränting M, Gullberg E, Näsvall J, Andersson DI. Activation of cryptic aminoglycoside resistance in Salmonella enterica. *Mol Microbiol* 2011;80:1464–1478.

4. Xia H, Tang Q, Song J, Ye J, Wu H *et al*. A yigP mutant strain is a small colony variant of *E. coli* and shows pleiotropic antibiotic resistance. *Can J Microbiol* 2017;63:961–969.

5. Musher DM, Baughn RE, Merrell GL. Selection of small-colony variants of Enterobacteriaceae by in vitro exposure to aminoglycosides: pathogenicity for experimental animals. *J Infect Dis* 1979;140:209–214.

6. Vestergaard M, Nøhr-Meldgaard K, Ingmer H. Multiple pathways towards reduced membrane potential and concomitant reduction in aminoglycoside susceptibility in *Staphylococcus aureus*. *Int J Antimicrob Agents* 2018;51:132–135.

7. Lázár V, Pal Singh G, Spohn R, Nagy I, Horváth B *et al*. Bacterial evolution of antibiotic hypersensitivity. *Mol Syst Biol* 2013;9:700.

8. Roggenkamp A, Sing A, Hornef M, Brunner U, Autenrieth IB *et al*. Chronic prosthetic hip infection caused by a small-colony variant of *Escherichia coli*. *J Clin Microbiol* 1998;36:2530–2534.

9. Santos V, Hirshfield I. The physiological and molecular characterization of a small colony variant of *Escherichia coli* and its phenotypic rescue. *PLoS One* 2016;11:e0157578.

10. Park YJ, Le Phuong N, Pinto NA, Kwon MJ, D'Souza R *et al*. Urinary tract infection caused by a small colony variant form of capnophilic *Escherichia coli* leading to misidentification and non-reactions in antimicrobial susceptibility tests. *Antimicrob Resist Infect Control* 2018;7:139.

11. Tappe D, Claus H, Kern J, Marzinzig A, Frosch M *et al*. First case of febrile bacteremia due to a wild type and small-colony variant of *Escherichia coli*. *Eur J Clin Microbiol Infect Dis* 2006;25:31–34.

12. Kahl BC. Small colony variants (SCVs) of *Staphylococcus aureus*--a bacterial survival strategy. *Infect Genet Evol* 2014;21:515–522.

13. Pränting M, Andersson DI. Mechanisms and physiological effects of protamine resistance in *Salmonella enterica* serovar typhimurium LT2. *J Antimicrob Chemother* 2010;65:876–887.

14. Curtis TD, Gram L, Knudsen GM. The small colony variant of Listeria monocytogenes is more tolerant to antibiotics and has altered survival in RAW 264.7 murine macrophages. *Front Microbiol* 2016;7:1056.

15. Kastbjerg VG, Hein-Kristensen L, Gram L. Triclosan-induced aminoglycoside-tolerant Listeria monocytogenes isolates can appear as small-colony variants. *Antimicrob Agents Chemother* 2014;58:3124–3132.

16. Verkamp E, Chelm BK, Isolation CBK. Isolation, nucleotide sequence, and preliminary characterization of the *Escherichia coli* K-12 hemA gene. *J Bacteriol* 1989;171:4728–4735.

17. Sabir N, Ikram A, Zaman G, Satti L, Gardezi A *et al*. Bacterial biofilm-based catheter-associated urinary tract infections: causative pathogens and antibiotic resistance. *Am J Infect Control* 2017;45:1101–1105.

18. Musicha P, Feasey NA, Cain AK, Kallonen T, Chaguza C *et al*. Genomic landscape of extended-spectrum *β*-lactamase resistance in *Escherichia coli* from an urban African setting. *J Antimicrob Chemother* 2017;72:1602–1609.

19. Hubbard ATM, Jafari NV, Feasey N, Rohn JL, Roberts AP. Effect of Environment on the Evolutionary Trajectories and Growth Characteristics of Antibiotic-Resistant *Escherichia coli* Mutants. *Front Microbiol* 2019;10:10.

20. Clinical and Laboratory Standards Institute. CLSI Standard M07. *Methods for Dilution Antimicrobial Susceptibility Testing for Bacteria That Grow Aerobically*, 11th ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.

21. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.

22. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

23. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* 2014;1151:165–188.

24. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 2015;43:W174–W181.

25. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;33:2302–2309.

26. Chung CT, Niemela SL, Miller RH. One-Step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* 1989;86:2172–2175.

27. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term experimental evolution in *Escherichia coli*. I. adaptation and divergence during 2,000 generations. *Am Nat* 1991;138:1315–1341.

28. Thulin M. Bat: an online tool for analysing growth curves 2018.

29. Al Laham N, Rohde H, Sander G, Fischer A, Hussain M *et al*. Augmented expression of polysaccharide intercellular adhesin in a defined Staphylococcus epidermidis mutant with the small-colony-variant phenotype. *J Bacteriol* 2007;189:4494–4501.

30. Ramiro RS, Costa H, Gordo I. Macrophage adaptation leads to parallel evolution of genetically diverse *Escherichia coli* small-colony variants with increased fitness in vivo and antibiotic collateral sensitivity. *Evol Appl* 2016;9:994–1004.

31. Butler CA, Dashper SG, Zhang L, Seers CA, Mitchell HL *et al*. The *Porphyromonas gingivalis* ferric uptake regulator orthologue binds hemin and regulates hemin-responsive biofilm development. *PLoS One* 2020;9:e111168.

32. Gomes L, Monteiro G, Mergulhão F. The impact of IPTG induction on plasmid stability and heterologous protein expression by *Escherichia coli* biofilms. *Int J Mol Sci* 2020;21:576.

33. Nardella C, Boi D, di Salvo ML, Barile A, Stetefeld J *et al*. Isolation of a complex formed between *Acinetobacter baumannii* HemA and HemL, key enzymes of tetrapyrroles biosynthesis. *Front Mol Biosci* 2019;6:6.

Edited by: I. Martin-Verstraete and W. van Schaik