MR. SHADY MANSOUR KAMAL (Orcid ID : 0000-0002-7565-811X)
PROF. UTE ROMLING (Orcid ID : 0000-0003-3812-6621)

Article type : Research Article

Corresponding author mail id: Ute.Romling@ki.se

A recently isolated human commensal *Escherichia coli* ST10 clone member mediates enhanced thermotolerance and tetrathionate respiration on a P1 phage derived IncY plasmid

running title: characterization of an E. coli K-12 clade member

Shady Mansour Kamal[1,2], Annika Cimdins-Ahne[3], Changhan Lee[1,*], Fengyang Li[1], Alberto J. Martin-Rodriguez[1], Zaira Seferbekova[4,5], Robert Afasizhev[4], Haleluya Tesfaye Wami[3], Panagiotis Katikaridis[6], Lena Meins[6], Heinrich Lünsdorf[7], Ulrich Dobrindt[3], Axel Mogk[6], Ute Römling[1]

[1]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
[2]Department of Microbiology and Immunology, Faculty of Pharmaceutical Sciences & Pharmaceutical Industries, Future University in Egypt, Cairo, 11835, Egypt
[3]Institute of Hygiene, University of Münster, Münster, Germany
[4]Kharkevich Institute for Information Transmission Problems, RAS, Moscow, Russia

[5]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
[6]Center for Molecular Biology of the University of Heidelberg (ZMBH) and German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, 69120 Heidelberg, Germany
[7]Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124, Braunschweig, Germany

[*]Current address: Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

Keywords: *Escherichia coli*, disaggregase ClpG, IncY plasmid, phylogenetic analysis, thermotolerance, tetrathionate respiration

Abstract

The ubiquitous human commensal *Escherichia coli* has been well investigated through its model representative *E. coli* K-12. In this work, we initially characterized *E. coli* Fec10, a recently isolated human commensal strain of phylogroup A/sequence type ST10. Compared to *E. coli* K-12, the 4.88 Mbp Fec10 genome is characterized by distinct single nucleotide polymorphisms and acquisition of genomic islands. In addition, *E. coli* Fec10 possesses a 155.86 kbp IncY plasmid, a composite element based on phage P1. pFec10 codes for a variety of cargo genes such as a tetrathionate reductase and its corresponding regulatory two-component system. Among cargo gene products is also the Transmissible Locus of Protein Quality Control (TLPQC), which mediates tolerance to lethal temperatures in bacteria. The disaggregase $ClpG_{GI}$ of TLPQC constitutes a major determinant of thermotolerance of *E. coli* Fec10. We confirm stand-alone disaggregation activity, but observe distinct biochemical characteristics of $ClpG_{GI\text{-}Fec10}$ compared to the nearly identical *Pseudomonas aeruginosa* $ClpG_{GI\text{-}SG17M.}$ Furthermore, we observed a unique contribution of $ClpG_{GI\text{-}Fec10}$ to the exquisite thermotolerance of *E. coli* Fec10 suggesting functional differences between both disaggregases *in vivo*. Detection of thermotolerance in 10% of human commensal *E. coli* isolates suggests successful establishment of food-borne heat resistant strains in the human gut.

Keywords: *Escherichia coli*, disaggregase ClpG, IncY plasmid, phylogenetic analysis, thermotolerance, tetrathionate respiration

1 INTRODUCTION

Escherichia coli colonizes the gastrointestinal tract of almost every human in low numbers as the most predominant commensal facultative anaerobic bacterium. *E. coli* conquers the gastrointestinal tract soon after birth (Bettelheim & Lennox-King, 1976), which contributes to early stimulation of the immune system, strengthening of the epithelial barrier function, production of vitamin B12 and K and provides the physiological basis for the growth of anaerobes (Blount, 2015). However, the human gastrointestinal tract is only one of the habitats of *E. coli.* Members of the highly diverse species *E. coli*, which consists of *E. coli* sensu stricto and several cryptic *E. coli* clades are found in diverse ecological niches including the gastrointestinal tract of wild and domestic animals, sewage, slurry, plants and the environment (Tenaillon *et al.*, 2010, Meric *et al.*, 2013, Jorgensen *et al.*, 2017).

Present epidemiological approaches classify *E. coli* strains into seven phylogroups (Clermont *et al.*, 2000, Beghain *et al.*, 2018) and/or sequence types (ST) based on sequence variability in a restricted number of conserved loci that allow discrimination on the subspecies level (Manges *et al.*, 2019). Epidemiological analyses indicate that commensal *E. coli* can be mainly classified into phylogroup A and B2, with the predominant phylogroups to widely shift temporally and among different populations (Massot *et al.*, 2016, Lescat *et al.*, 2013). Classification according to sequence types unraveled the ubiquitous occurrence of certain sequence types such as the multidrug resistant ST131 in fecal and clinical samples, but also ST10 is found in the gastrointestinal tract of humans, animals, in association with plants and in the environment (Reid *et al.*, 2017, Day *et al.*, 2019, Freitag *et al.*, 2018, Dublan Mde *et al.*, 2014). Other sequence types are more restricted (Manges et al., 2019). Genome sequencing revealed that intestinal and extraintestinal *E. coli* pathogens can be closely related to innocuous commensal *E. coli* strains (Cimdins *et al.*, 2017b). Indeed, enterotoxigenic *E. coli* can readily arise from commensal strains by the acquisition of few genetic elements including the transfer of characteristic virulence plasmids (von Mentzer *et al.*, 2014) indicating that plasmids can readily alter the characteristics of strains.

Plasmids can evolve from phages (Venturini *et al.*, 2019, Sternberg & Hoess, 1983), some of which exist as circular plasmids in the lysogenic state. The contribution of plasmids towards genomic plasticity reaches beyond the provision of pathogenicity factors, antimicrobial resistance determinants and degradation pathways (Shintani *et al.*, 2010, Pilla & Tang, 2018, Koraimann, 2018). One plasmid-encoded feature is the up to 19 kbp long Transmissible Locus of Protein Quality Control (TLPQC) which mediates elevated tolerance against lethal temperature, pressure and oxidatives (Boll *et al.*, 2017, Li & Ganzle, 2016, Li *et al.*, 2020,

Wang *et al.*, 2020, Lee *et al.*, 2016). The TLPQC locus, originally derived from an environmental species, has also been obtained by clinically relevant bacteria (Bojer *et al.*, 2012). The three core genes *dna-shsp20*$_{GI}$-*clpG*$_{GI}$ encoding a transcription factor, a holdase chaperone and a disaggregase are the most conserved loci of TLPQC. ClpG disaggregases with the horizontally transferred ClpG$_{GI}$ as a distinct subgroup, constitute stand-alone disaggregases with high intrinsic ATPase activity determinative for the tolerance phenotype to lethal temperatures in *Klebsiella pneumoniae*, *P. aeruginosa* and *E. coli* (Lee *et al.*, 2018, Bojer *et al.*, 2010). ClpG members surpass the disaggregation activity of the canonical Hsp70 (DnaK)/ClpB bi-chaperone disaggregase, which mediates basal thermotolerance in bacteria (Katikaridis, 2019).

The commensal *E. coli* K-12 strain, isolated in 1921, has been the genetic reference strain since then (Bachmann, 1972). Sequence type 10 *E. coli* K-12 is closely related to NCTC86, the *E. coli* strain originally identified by Theodor Escherich in 1886 and has been considered as a typical commensal strain (Khetrapal *et al.*, 2017). In this work, we initially characterized the biofilm forming commensal strain Fec10, a member of the ST10 clonal complex, closely related to *E. coli* K-12 and NCTC86, recently isolated from a healthy human being (Bokranz *et al.*, 2005). Fec10 contains the phage P1 derived IncY plasmid pFec10 bearing distinct cargo gene sequences, which indicate adaptation of Fec10 to adverse growth conditions. One of the distinct features of pFec10 is the acquisition of an operon for tetrathionate respiration, which is involved in regulation of rdar biofilm expression in the presence of the electron acceptor. Another cargo gene cluster is the TLPQC locus that provides enhanced tolerance to severe lethal temperature stress. We confirm autonomous disaggregation activity of thermotolerance mediating ClpG$_{GI}$ and show that ClpG$_{GI\text{-}Fec10}$ withstands higher temperatures than the wide-spread ClpB/DnaK bi-chaperone system in agreement with the overall exquisite temperature-tolerance phenotype of *E. coli* Fec10 and other commensal *E. coli* isolates.

2 RESULTS

2.1 The commensal isolate Fec10 is a modern member of the *E. coli* K-12 clonal group

Recently, we isolated and characterized commensal *E. coli* strains from the gastrointestinal tract of healthy humans (Bokranz et al., 2005). Thereby, initial phylogenetic analysis identified our internal reference strain ST10 *E. coli* Fec10 as closely related to *E. coli* K-12 (Cimdins et al., 2017b). As a first step towards a more detailed phylogenetic analysis of Fec10, we determined its genome sequence by PacBio RSII sequencing combined with Illumina sequencing. Subsequently, single nucleotide polymorphism (SNP) analysis of the core genomes of 521 fully sequenced and representative *E. coli* strains from different phylotypes including *E. coli* K-12 derivatives such as the MG1655 reference strain; the Theodor Escherich isolate NCTC86; a

recently identified Swine-derived ETEC strain closely related to K-12 (Shepard *et al.*, 2012) and strains most closely related to *E. coli* K-12 and Fec10 in the NCBI and NCTC database including strains from the ECOR collection demonstrated that Fec10 indeed is closely related to *E. coli* K-12 (Figure 1; Figure S1A,B; see below). Unlike *E. coli* K-12, Fec10 displays a full red, dry and rough (rdar) morphotype, a distinct biofilm colony phenotype characterized by the production of the exopolysaccharide cellulose and amyloid curli fimbriae as extracellular matrix components at ambient temperature ((Figure S2); (Cimdins et al., 2017b)). Fec10 is a member of the ST10 type strain complex, which opens up the opportunity to document adaptation of a member of the abundant phylogroup A/ST10 *E. coli* subgroup to post-industrial human beings.

2.2 Description of Fec10 genome features

The genome of strain Fec10 is composed of one circular chromosome of 4.88 Mbp in size and at least two circular entities, a plasmid (see below) and a pro483-like *E. coli* phage (NC_028943). The chromosome of *E. coli* Fec10 is colinear with the *E. coli* K-12 MG1655 reference genome (Figure S1B). With in total 4828 (and 4598 chromosomal) annotated coding sequences, Fec10 has 3926 (85.4%) coding sequences (CDS) in synteny with *E. coli* K-12 MG1655. Comparing Fec10 with six *E. coli* K-12 derivatives (whereby each isolate differs in gene composition), Fec10 has 3853 core, 975 variable and 636 strain specific coding sequences. Common regions of *E. coli* Fec10 and closely related NCTC86 (Figure 1, S1 A,B) comprise ~80% of the genomes, the level of sequence identity being 95%, and the vertically inherited fraction 86%. These numbers are similar when Fec10 is compared to other closely related strains such as VR50, FORC 064, NCTC9102, RR1, AG100, D4, D9, TO73, 26561 with the common genome parts ranging from 77 to 84%, the sequence identity from 95-96%, and the vertically inherited fraction of the common genome from 86-90%. With respect to the *E. coli* population (1080 genomes), Fec10 comprises a core genome of 1321 (27.4%) CDSs and 3567 (72.6%) variable CDSs (Vallenet *et al.*, 2020). Among the acquired distinct genomic islands and regions of plasticity are several phage-like genomic island, but also the yersiniabactin operon, designated "high pathogenicity island", which is a virulence determinant in several enterobacterial genera (Figure S1B).

2.3 *E. coli* Fec10 harbors a phage P1 derived IncY plasmid with the TLPQC locus

E. coli Fec10 harbors a 155.9 kbp plasmid (termed pFec10) with an IncY origin of replication (Figure 2; Table S1). The rare IncY replicon originates from the P1 phage (Sternberg & Hoess, 1983). Several IncY/P1 phage derivatives have recently been described (Venturini et al., 2019).

Indeed, BLAST search with standard parameters indicated 25 and 15% query cover by *E. coli* phage P1 (NC_005856.1) and *Salmonella* phage SJ46 (NC_031129.1), respectively, which thus cover a significant part of the plasmid backbone (Figure S1C; Table S1; (Arndt *et al.*, 2016) (Altschul *et al.*, 1990)). However, major head and sheat proteins are missing, making the assembly of a functional phage unlikely.

The closest plasmid homolog of pFec10 is pSSE which is a multireplicon IncHI2/IncHI2A plasmid from the thermo-resistant *Salmonella enterica* subsp. *enterica* serovar Senftenberg strain 775W (ATCC42845) isolated from Chinese egg powder in 1941. pFec10 is 99.96% identical over 50% coverage to pSSE, partly due to the presence of the highly conserved TLPQC locus present on pFec10 and pSEE, which harbors two distinct TLPQC loci (Figure S1D, S3; Table S1; (Nguyen *et al.*, 2017)).

Assessment of characteristic elements for plasmid maintenance showed that the pFec10 plasmid does not encode genetic elements that mediate antibiotic resistance (Hendriksen *et al.*, 2019). Plasmid stability is most likely maintained by the type II toxin-anti-toxin system YdeE-YdeD (Prevent host death-death on curing (Phd-Doc)), whereby *ydeD* encodes a 'death on curing' toxin (Lehnherr *et al.*, 1993, Liu *et al.*, 2008). Death on curing toxins arrest elongation factor EF-Tu in a non-functional state by phosphorylation. Within the pFec10 sequence, there is also a region representing an origin of transfer (Figure 2).

2.4 pFec10 is an unconventional plasmid

Distinct features of pFec10 include a multitude of distinct IS elements (Figure 2; Table S1, S2). In addition, a variety of cargo gene products are encoded on pFec10. With 11586 bp, the longest open reading frame encodes a 3862 amino acid long tandem-95-repeat protein with the closest homolog (50.8% identity) in *Butticuxella* spp. Another distinct feature is the presence of a type I restriction-modification system with the modification, sensitivity and restriction subunit. Besides the TLPQC locus involved in stress resistance, cargo gene modules e.g. for detoxification, nutrient acquisition and redox balance are encoded on pFec10. For example, the plasmid codes for genes involved in arsenate resistance, a DyP-type decolorizing peroxidase and its encapsulating protein, a lactate utilization gene cluster and a tetrathionate reductase operon and its respective sensing/regulatory two-component system (Figure 2; Table S1). Those elements are embedded into the phage-derived plasmid scaffold (Figure 2, Table S1; (Arndt et al., 2016)).

The plasmid encodes a DyP-decolorizing peroxidase unit involved in detoxification. DyP decolorizing peroxidases of fungi and bacteria have unknown physiological function, but wide biotechnological applicability as they degrade lignin and oxidize a variety of aromatic compounds, dyes and other small molecules (Singh *et al.*, 2013, Lin *et al.*, 2019). A 32% identical DyP homologue is encoded on the Fec10 and *E. coli* K-12 chromosome. Furthermore, an encapsulin, which encapsulates peroxidases via a recognized C-terminus in nanocompartments, is encoded downstream of the peroxidase (Tracey *et al.*, 2019) and a NAD-dependent formate dehydrogenase is encoded upstream.

As another element of detoxification, the plasmid encodes two gene clusters involved in arsenate reduction. These newly arranged gene clusters encode gene products closely related to gene products from the Fec10 and *E. coli* K-12 core genome, but also gene products with no close homologues. The four-gene cluster encodes a protein with partial homology to the ArsA efflux transporter ATPase subunit, an ArsC reductase, an ArsF efflux pump membrane protein and an ArsA efflux transporter ATPase subunit. ArsC and ArsF have close homologues with >90% identity on the chromosome of Fec10 and *E. coli* K-12. The three-gene cluster encodes an ArsR transcriptional regulator, a ArsD transacting repressor and another ArsA efflux transporter ATPase subunit. Only the ArsR regulator possesses a 74.1% identical homologue on the core genome.

Furthermore, several components required for D- and L-Lactate catabolism are encoded on pFec10. The gene cluster contains an LctP L-lactate permease, which has no counterpart on the chromosome of Fec10 and *E. coli* K-12 (MG1655 strain). On the other hand, a homolog of the D-lactate dehydrogenase Dld with >85% identity is present in Fec10 and K-12. Adjacently located is the *ykgEFG* gene cluster with a corresponding highly similar *ykgEFG* operon in Fec10 and K-12. For example, the iron-sulphur oxidoreductase subunit YkgE of the L-lactate dehydrogenase shows 94% identity. *YkgEFG* is involved in L-lactate catabolism and biofilm formation in *Bacillus subtilis* (Chai *et al.*, 2009), but has no L-lactate catabolism phenotype in *E. coli* K-12, but an alternative operon for lactate catabolism is present (Chai et al., 2009). Therefore, the evolutionary driving force for the acquisition of a highly homologous second copy of *ykgEFG* remains to be determined.

The species *E. coli* usually does not encode the genetic information to respire tetrathionate. However, BLAST search indicates that an almost identical tetrathionate operon is present in selected *E. coli* and *Klebsiella pneumoniae* strains (data not shown) suggesting that acquisition of the tetrathionate operon by a commensal *E. coli* strain aids adaptation to altered

conditions in the gastrointestinal tract. The tetrathionate operon might have originated from *Citrobacter* spp. (homology can be as high as >99%), as tetrathionate reduction is a characteristic feature of, e.g. the species *Citrobacter freundii* (Kapralek, 1972). In order to get a first insight into the biological impact and physiological contribution of the plasmid under certain environmental conditions, we chose to initially investigate the biological consequences of two cargo operons on the plasmid, the tetrathionate operon and the core *dna-shsp20*$_{GI}$-*clpG*$_{GI}$ operon of the TLPQC locus for elevated temperature tolerance.

2.5 A Biofilm phenotype is associated with tetrathionate sensing and respiration

Tetrathionate is used as an alternative electron acceptor by *Salmonella typhimurium* (Winter & Baumler, 2011, Hensel *et al.*, 1999), but not *E. coli.* A combination of tetrathionate versus thiosulfate, the product of tetrathionate respiration, is even toxic for *E. coli* (Palumbo & Alford, 1970). We have recently shown that the reduction of alternative electron acceptors triggers alternations in the amount of biofilm formation in bacteria (Martin-Rodriguez and Römling, unpublished data; (Martin-Rodriguez *et al.*, 2020)). To this end, we deleted the tetrathionate operon genes on the pFec10 plasmid. We constructed three different mutants; 1. a deletion mutant of *ttrSR* encoding the two-component system histidine kinase and response regulator required for tetrathionate sensing and activation of the reductase operon; 2. a deletion mutant of *ttrBCA* encoding the tetrathionate reductase in combination with a short uncharacterized upstream open reading frame of 144 bps encoding a hypothetical protein; 3. a mutant with a deletion of the two divergently transcribed operons; to test whether exposure to tetrathionate has any effect on rdar biofilm formation (Figure S2E; (Römling *et al.*, 1998b)). When grown on Congo Red (CR) agar plate to display biofilm formation, the mutants did not show a change in the rdar biofilm morphotype. We observed, however, a mutant and temperature dependent effect of tetrathionate on colony morphology appearance when cells were grown on a CR agar plate with 1.5 and 15 mM tetrathionate (Figure 3). At 28 °C, the rdar morphotype was diminished strongest upon deletion of the *ttrBCA* operon. At 37 °C, the colony appearance of the wild type turned pale at 15 mM tetrathionate, while deletion of the *ttrBCA* operon led to an upregulation of the rdar morphotype. Thus tetrathionate has a differential temperature dependent effect on colony morphology.

2.6 *ClpG*$_{GI}$ and *dna-hsp20*$_{GI}$-*clpG*$_{GI}$ mediate elevated tolerance to lethal temperature

High tolerance to lethal heat treatment has been associated with the TLPQC locus (alternatively called LHR locus) in various species with the three-gene operon *dna-shsp20*$_{GI}$-*clpG*$_{GI}$ to play a

determinative role in extended heat tolerance (Bojer et al., 2010, Lee *et al.*, 2015). In particular, horizontally transferred $ClpG_{GI}$ (ClpK) has been shown to mediate increased tolerance to lethal temperature in *K. pneumoniae, P. aeruginosa* and food-derived and ESBL clinical *E. coli* isolates (Lee et al., 2018, Boll et al., 2017, Bojer *et al.*, 2011). In order to investigate whether and to which extent $clpG_{GI}$ and/or the entire $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ operon contribute to tolerance against lethal temperatures, we deleted solely $clpG_{GI}$ and the entire $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ operon in the commensal *E. coli* strain Fec10. Deletion of $clpG_{GI}$ and $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ dramatically reduced the tolerance to lethal heat treatment >3-fold $\log_{10}$ upon exposure to 60 °C for 15 min with a major contribution derived from $clpG_{GI}$, whereas the wild type Fec10 remained almost unaffected (Figure 4A; Figure S4). Similarly, exposure to 65 °C for 2 min reduced survival of both deletion mutants >4-fold $\log_{10}$ with Fec10 wild type not affected (Figure S4). The loss of tolerance to lethal heat treatment could, however, be partially restored by the provision of $clpG_{GI}$ and $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ on plasmid pBAD30 under the control of the L-arabinose inducible P_{BAD} promoter (Figure 4). These results show that in commensal isolates of *E. coli* provision of tolerance to elevated heat shock is largely mediated by the $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ operon with $clpG_{GI}$ constituting the major determining factor. In *P. aeruginosa* clone C, gene products of the chromosomal $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ locus are predominantly expressed in the stationary phase of growth (Lee et al., 2018, Lee et al., 2015). Investigation of $ClpG_{GI}$ production of *E. coli* Fec10 in the logarithmic and stationary phase of growth revealed production of the protein at 37 °C in log phase cells with $ClpG_{GI}$ levels to increase during stationary phase (Figure 4B). Notably, the canonical disaggregase ClpB is expressed at detectable levels in Fec10 at this temperature mainly in the stationary phase of growth.

2.7 $ClpG_{GI\text{-}Fec10}$ can complement *clpB* and *dnaK103* mutants in *E. coli* MC4100

We have previously shown that *P. aeruginosa* $ClpG_{GI\text{-}SG17M}$ expressed from a plasmid restores thermotolerance to *E. coli* MC4100 Δ*clpB* and *dnaK103* mutants, lacking individual components of the canonical bi-chaperone disaggregase (Lee et al., 2018). $ClpG_{GI\text{-}Fec10}$ is 95.8% identical to $ClpG_{GI\text{-}SG17M}$ with a conserved N-terminal domain and two highly conserved AAA+ ATPase domains containing conserved catalytic motifs such as the Walker A/Walker B ATP binding motif, while amino acid changes accumulate also at the C-terminal end of the N2 domain and the highly charged C-terminus ((Figure S5, S6); (Lee et al., 2018)). $ClpG_{GI\text{-}Fec10}$ restored heat tolerance at 50 °C to the same extent as $ClpG_{GI\text{-}SG17M}$ in MC4100 *clpB* and *dnaK103* mutants. Both $ClpG_{GI}$ variants were superior as compared to plasmid-expressed *clpB*,

confirming that enhanced heat resistance can be transferred to other bacteria by sole expression of *clpG*$_{GI}$ (Figure 4C). Expression of *clpG*$_{GI\text{-}Fec10}$ also effectively rescued *dnaK103* null mutants, though lack of the central Hsp70 chaperone causes massive protein aggregation upon heat stress (Figure 4C) (Mogk *et al.*, 1999). This indicates that *clpG*$_{GI}$ covers and expands the functionality of *dnaK* and *clpB* in tolerance to lethal temperatures.

2.8 ClpG$_{GI\text{-}Fec10}$ is a stand-alone disaggregase with high intrinsic ATPase activity

The complementation experiments indicated that ClpG$_{GI\text{-}Fec10}$ has a similar functionality as ClpG$_{GI\text{-}SG17M}$. To assess the activity of ClpG$_{GI\text{-}Fec10}$ against model substrates, we purified ClpG$_{GI\text{-}Fec10}$ according to previously described protocols (Figure S7; (Lee et al., 2018)). To this end, end-point assessment of the restoration of the catalytic activity of heat-aggregated luciferase indicated that both, ClpG$_{GI\text{-}Fec10}$ and ClpG$_{GI\text{-}SG17M}$, recovered >30% of luciferase activity after 120 min, while the 15% recovery rate for ClpB was more than 50% lower (Figure 5). Residual disaggregase activity was observed for the DnaK chaperone system (KJE: including the DnaJ, GrpE cochaperones), which also competes with ClpG$_{GI\text{-}Fec10}$ for substrate binding, as 1:1 co-incubation of DnaKJE with ClpG$_{GI\text{-}Fec10}$ abolished luciferase recovery above residual levels (Figure 5; Figure S8). As expected, ClpB alone did not show any activity.

Similarly, we monitored recovery of catalytic activity of the model substrate malate dehydrogenase (MDH) (Figure S8). In this case, the chaperonin complex GroES/EL was added to aid refolding of the disaggregated amino acid chain (Lee et al., 2018). As previously reported, we observed that the end-point recovery rate was highest for the ubiquitous ClpB/DnaKJE system with >80% MDH recovered activity after 120 min. ClpG$_{GI\text{-}Fec10}$, and to a lesser extent ClpG$_{GI\text{-}SG17M}$, showed a trend to display a reduced recovery rate (Figure S8). Lack of disaggregase activity was observed for ClpB and the DnaKJE system, which competed with ClpG$_{GI\text{-}Fec10}$ for substrate binding, as 1:1 co-incubation of DnaKJE with ClpG$_{GI\text{-}Fec10}$ abolished recovery of MDH activity (Figure S8).

Interaction of ClpB with its aggregate-loaded DnaK co-chaperon does not only deliver aggregates to ClpB, but also activates the ATPase activity of ClpB dedicated to disaggregation (Zolkiewski, 1999). We have recently shown that stand-alone ClpG$_{GI\text{-}SG17M}$ has high intrinsic basal ATPase activity, >11 fold higher than the basal activity of ClpB (Figure 5C; (Lee et al., 2018)). The estimated basal ATPase activity of ClpG$_{GI\text{-}Fec10}$ was >15-fold higher than ClpB basal activity and even significantly higher as compared to ClpG$_{GI\text{-}SG17M}$. Addition of the ClpB substrate casein stimulated ATPase activity >6.5-fold, while the ATPase activity of ClpG$_{GI\text{-}Fec10}$ and ClpG$_{GI\text{-}SG17M}$ was inhibited 2.4-fold and >1.2-fold, respectively. This suggests differences in

ATPase regulation and substrate specificities of ClpB and $ClpG_{GI}$ disaggregases, but also between different $ClpG_{GI}$ disaggregases. Of note, $ClpG_{GI}$ protein variability is grossly correlated with the phylogenetic position of the host organism (Figure S5). We assessed differential binding of the disaggregases to fluorescein-labeled FITC-casein by anisotropy measurements (Figure 5D). In the presence of ATPγS ClpB bound with high affinity (K_d: 0.56 ± 0.14 μM) to the substrate, while binding of both $ClpG_{GI}$ disaggregases was much weaker and K_d-values could not be calculated as the binding curves did not saturate.

2.9 Thermal stability of $ClpG_{GI}$ and ClpB

The extreme heat tolerance phenotype of strain Fec10 (Figure 4; Figure S4B), which is comparable to the highest heat tolerant *E. coli* strain isolated after artificial decontamination of meat from slaughterhouse strains (Dlusskaya *et al.*, 2011) also raised the question about the thermodynamic stability of $ClpG_{GI}$ proteins. Circular dichroism (CD) is a straightforward approach to assess changes in the secondary structure of proteins upon temperature upshift as the CD spectra of α-helix, β-sheet and random coil differ substantially (Figure S9A). We therefore recorded the thermal transition of the protein structure by monitoring CD spectra of $ClpG_{GI\text{-}Fec10}$, $ClpG_{GI\text{-}SG17M}$ and ClpB, allowing to compare stabilities of stand-alone $ClpG_{GI}$ and canonical ClpB disaggregases. First, we determined the CD spectrum of $ClpG_{GI\text{-}Fec10}$, $ClpG_{GI\text{-}SG17M}$ and ClpB without and with ATPγS and Mg^{2+}. The CD spectrum of all three proteins without ATPγS and Mg^{2+} resembled an α-helical spectrum (Figure S9A,B). Measurement of the melting temperature of the secondary structure by monitoring the change in absorption of left and right circular polarized light at 222 nm showed that ClpB displayed a far lower melting temperature of 53.2 °C compared to $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$, which showed structural transition at 64.5 and 62.9 °C, respectively (Figure 6, Figure S9B).

We next determined the CD spectra and transition curves in the presence of 1 mM ATPγS and 5 mM Mg^{2+}, which increased melting temperatures for all proteins (e.g. ClpB: 61.9 °C, $ClpG_{GI\text{-}Fec10}$: 69.0 °C and $ClpG_{GI\text{-}SG17M}$: 69.8 °C) (Figure 6, Figure S9D, E). Increase in protein stability is gained by the energy derived from ATPγS binding and ATP-driven oligomerization. T_M-values of $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$ determined in the presence of nucleotide remained >7 °C higher as compared to ClpB, demonstrating higher thermal stability under all conditions tested (Figure 6). Notably, the melting temperatures of ATP-bound $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$ are close to 70 °C which are in congruence with viability of temperature tolerant *E. coli* upon exposure to lethal temperatures (Figure 4A, S4; (Dlusskaya et al., 2011)).

As both $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$ show increased, but not identical thermostability *in vitro* compared to the ClpB disaggregase, we were wondering whether $ClpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$ show a comparable performance *in vivo* upon extreme temperature exposure. To this end, we expressed $clpG_{GI\text{-}Fec10}$ and $clpG_{GI\text{-}SG17M}$ from plasmid pBAD30 in *E. coli* Fec10 $clpG_{GI}$ and *dna-*$shsp20_{GI}$*-*$clpG_{GI}$ mutants (Figure S10). Surprisingly, in contrast to complementation with $clpG_{GI\text{-}Fec10}$, we observed no complementation with $clpG_{GI\text{-}SG17M}$ upon exposure to a lethal temperature of 60 °C. Lowering the temperature to 55 °C, though, allowed complementation of the *E. coli* Fec10 $clpG_{GI}$ mutant partially also by $ClpG_{GI\text{-}SG17M}$. Thus, how *in vivo* complementation capacity of both proteins is reflected by the *in vitro* biochemical characteristics needs to be sorted out. Of note, tolerance to lethal temperatures of *E. coli* Fec10 $clpG_{GI}$ and *dna-*$shsp20_{GI}$*-*$clpG_{GI}$ is superior compared to *E. coli* K-12 and several investigated wild type strains (Figure S4A, S10) suggesting the presence of additional factors that contribute to this phenotype.

2.10 Commensal *E. coli* strains can show high tolerance to lethal temperatures

We were wondering whether gastrointestinal commensal strains commonly harbor the TLPQC locus and show resistance to lethal temperatures. We screened additional 31 representatives of human commensal *E. coli* strains, genetically unrelated as judged by pulsed-field gel electrophoresis (Bokranz et al., 2005), for the presence of the small heat shock protein 20 gene ($shsp20_{GI}$) as an indication for the TLPQC locus (Lee et al., 2015). In total four out of 32 strains (12.5%), Fec6, Fec10, Fec32 and Fec89, contained the TLPQC locus (data not shown). Subsequently, we exposed those strains and four non-TLPQC bearing strains, Fec41, Fec55, Fec75 and Fec113, to a lethal temperature treatment at 55 °C. Indeed, 3 out of four $shsp20_{GI}$–positive strains were more thermotolerant with less than 100-fold reduction in viability count after exposure to 55 °C for 45 min (Figure S4A). Fec6 and Fec10 were the two most thermotolerant strains that substantially survived lethal heat at 60 °C for 15 min (Figure 4A) and exposure to 65 °C for 5 min (Figure S4B).

E. coli strains have previously been classified into seven major phylotypes (Beghain et al., 2018, Clermont et al., 2000). Thereby, commensal strains represent most commonly phylotype A or B2 depending on the investigated human population (Massot et al., 2016, Lescat et al., 2013). All *E. coli* strains, which showed an elevated lethal-heat-tolerance phenotype were phylotype A (Bokranz et al., 2005), an observation that has previously also been reported for food-derived heat–tolerant *E. coli* isolates.

2.11 Two TLPQC loci are present in the high heat tolerant strain *E. coli* Fec6

We determined the whole genome sequence of *E. coli* Fec6, which displayed an extraordinary high tolerance to lethal temperature surviving 65 °C for 5 min with no reduction in viable counts (Figure S4B). The Fec6 genome encodes two TLPQC loci integrated at different locations on the chromosome. TLPQC-1 of Fec6 and TLPQC of Fec10 are grossly identical, align to well-characterized TLPQC loci and might represent the original form of the island (Figure S3). TLPQC-2 of Fec6 has acquired a number of cargo gene products that inserted between a gene of unknown function and the *trx* thioredoxin gene (Figure S3). Truncated inserted genes are found downstream of *kefC* encoding a K^+/H^+ anti-transporter.

The TLPQC locus of *E. coli* Fec10 has 100% query coverage with >99.8% identity to the TLPQC locus of plasmid A of *E. coli* H15. TLPQC-1 of Fec6 has >99% query coverage and >99% identity in several *E. coli* genomes including P12b, NCTC11121, NCTC9966, H1, NCTC9040 and *E. coli* C (NCBI BLAST accessed October 7, 2019). TLPQC-2 of Fec6 has >90% query coverage and >98.8% identity with sequences from *E. coli* VREC0864, S43 and VREC0761. Rapid evolution of this locus is indicated as the island is found with >97% query coverage and >97.5% sequence identity in *E. coli* MEM (BLAST accessed October 7, 2019).

The $ClpG_{GI\text{-}Fec6}$ disaggregases of TLPQC-1 and TLPQC-2 are highly similar, but not identical to $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI}$s from other TLPQC islands (Figure S5, S6). We speculate differential substrate recognition and processing or temperature-dependent activity by $ClpG_{GI\text{-}Fec6\text{-}1}$ and $ClpG_{GI\text{-}Fec6\text{-}2}$ as preliminary indicated for $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$ (Figure 5C), which might contribute to the extended temperature tolerance of strain Fec6. The diversified N- and C-termini eventually provide broadened substrate specificity or direct the disaggregases to specific cellular sites. As the total level of disaggregase is increased in *E. coli* Fec6 compared to Fec10 (Figure 4B), the presence of two $ClpG_{GI}$ copies in strain Fec6 can alternatively explain enhanced stress protection.

3 DISCUSSION

Commensal *E. coli* strains, albeit present at low frequency, are found in the gastrointestinal tract of almost every human being as well as in animals (Rossi *et al.*, 2018, Escobar-Paramo *et al.*, 2006). Despite of the ubiquitous association of commensal *E. coli* with humans and other higher organisms, the breadth and characteristics of the *E. coli* strain population in the ecological niche of the gut has not been covered in detail.

In this work, we initially characterized a commensal representative of the ST10 clonal complex closely related to *E. coli* K-12. This strain produces a rdar biofilm morphotype at 28 °C (Figure S2; (Cimdins et al., 2017b)) and harbors an IncY plasmid with the previously identified

TLPQC/LHR locus that mediates tolerance towards temporal exposure to lethal temperature and additional cargo genes involved in detoxification.

Among the *E. coli* strains, the historical strain *E. coli* K-12 isolated from a reconvalescent patient is the most well-investigated. However, it has been questioned to which extent *E. coli* K-12 has adapted to laboratory conditions during years of laboratory maintenance and subject to extensive manipulation (Browning *et al.*, 2013). ST10 *E. coli* strains, which belong to phylogroup A, are not only found at high frequency in the gastrointestinal tract of human beings, but also in the environment, plants, animals and the clinical habitat suggesting a superior capacity for survival (Jorgensen et al., 2017, Shepard et al., 2012, Freitag et al., 2018, Manges *et al.*, 2015, Reid et al., 2017, Richter *et al.*, 2018). Whether and, if so, which distinct subgroups exists within the ST10 type, which horizontally transferred genetic elements and single nucleotide polymorphisms contribute to persistence and survival needs to be further explored, but O-antigen and flagella types seem to be highly variable among closely related ST10 strains (Day et al., 2019).

IncY plasmids derive from phage P1, which is present as an episomal copy in its lysogenic cycle, stably maintained by a toxin/antitoxin system (Lobocka *et al.*, 2004). pFec10 clearly is a composite plasmid with a phage gene scaffold harbouring a variety of unconventional cargo genes that mediate e.g. stress resistance and nutrient acquisition. We assume that we observe, with pFec10, the early development of a prophage into a plasmid, which subsequently acquired a high number of IS elements indicative for rapid evolution and cargo genes including a TLPQC locus to provide fitness advantages for strain Fec10. Furthermore, as judged from the majority of cargo genes, we assume that Fec10 may experience inflammatory and/or toxic conditions in the gut. For example, tetrathionate reduction is utilized by the gastrointestinal pathogen *S. typhimurium*, a close relative, to successfully thrive in the inflamed gut (Winter & Baumler, 2011). The ability to reduce tetrathionate could thus contribute to the *in vivo* fitness of commensal Fec10 upon inflammation whereby the produced thiosulfate might be reduced by other microorganisms. Furthermore, the TLPQC locus mediates not only thermotolerance, but also tolerance against the inflammatory gut oxidants chlorine and hydrogen peroxide (Wang et al., 2020).

Interestingly, a collection of minimally processed commensal *E. coli* strains of independent genetic background, previously assessed for their biofilm formation capability (Bokranz et al., 2005), harbored >10% thermotolerant strains, while overall, 2% of sequenced *E. coli* genomes are thermotolerant (Li & Ganzle, 2016). 3 of 4 of those TLPQC bearing strains, including Fec10, were of low abundance in feces of healthy human beings (Bokranz et al., 2005).

Processes that select for thermotolerant strains in the environment and the medical setting include manipulations with temporal temperature upshift such as the processing of raw milk cheese, natural fermentation processes, steam treatment of meat and mild thermosterilization of endoscopes (Wang *et al.*, 2018, Peng, 2012, Bojer et al., 2011, Dlusskaya et al., 2011). We conclude from this work that the consumption of raw milk cheese or other food products does eventually establish *E. coli* strains with enhanced thermotolerance in the human gastrointestinal tract albeit at low frequency. Elevation of this strain population might occur, though, upon inflammation. Thermotolerance, however, is not a recently acquired phenotype of *E. coli* as a highly similar TLPQC locus is already present in *E. coli* C, isolated prior to 1920 (unpublished observation) and encoded on a plasmid of a *S. enterica* serovar Senftenberg strain isolated in 1941 (Nguyen et al., 2017). *S. enterica* probably readily survived in dried egg powder that was a frequent source of outbreaks of salmonellosis in the 1940ies. In this line, the TLPQC locus has recently also been shown to mediate pressure tolerance and tolerance against oxidizing agents used in the food industry (Li et al., 2020, Wang et al., 2020).

The only other TLPQC bearing plasmid in a food-borne *E. coli* belongs to the IncFII incompatibility group (Boll et al., 2017). The presence of the TLPQC locus on IncFII and IncY plasmids indicates that this locus can be readily transferred to other *E. coli* strains, or other bacterial species within or outside of the *Enterobacteriaceae* family. Plasmid transfer is especially effective in the gastrointestinal tract. In combination with a bloom of *E. coli* upon pathogen-driven inflammation, horizontal gene transfer can reach extremely high numbers even in the absence on antibiotic pressure (Stecher *et al.*, 2012).

As observed for *E. coli* strains from different origin (Li & Ganzle, 2016, Dlusskaya et al., 2011), the TLPQC locus can mediate thermotolerance to extremely high temperatures in commensal *E. coli* strains. This physiology has its foundation in the biochemical characteristics of the isolated disaggregase $ClpG_{GI}$, a major determinant of tolerance towards lethal temperatures (this work; (Lee et al., 2018)). $ClpG_{GI}$ exhibits enhanced disaggregation activity as compared to the canonical ClpB disaggregase, enabling the protein to process tight protein aggregates formed at more extreme temperature or higher protein concentration (Katikaridis, 2019). Of note, although highly similar, $ClpG_{GI}$ proteins from different bacteria differ in their biochemical characteristics and their ability to complement exposure to extreme lethal temperature. As such, $ClpG_{GI\text{-}Fec10}$ has a significantly higher ATPase activity than $ClpG_{GI\text{-}SG17M}$. The N2 domain, recently shown to suppress ATPase activity of $ClpG_{GI\text{-}SG17M}$ (Lee et al., 2018), shows extended sequence variability (Figure S6B) and could contribute to the elevated ATPase activity of $ClpG_{GI\text{-}Fec10}$. Our data indicated that $ClpG_{GI}$ proteins are particularly heat stable thus be protected from

unfolding and aggregation at more extreme temperatures. Indeed, both $ClpG_{GI}$ proteins showed increased *in vitro* thermostability compared to the ClpB/DnaK bi-chaperon system, the core genome disaggregase involved in temperature tolerance in *E. coli* and other organisms. The high melting temperature upon binding to ATPγS (T_M-value of 69-70°C) of $ClpG_{GI}$ represents another distinguishing feature from canonical ClpB and rationalizes why $ClpG_{GI}$, but not ClpB (T_M: 62°C) and its essential partner DnaK (T_M: 59°C) (Palleros *et al.*, 1992) protects bacteria from severe heat shock. Loss of canonical ClpB and DnaK at extreme temperatures due to unfolding (and likely aggregation) will eradicate the disaggregation potential of *E. coli* cells that do not encode $ClpG_{GI}$ as an alternative disaggregation system. Nevertheless, only $ClpG_{GI\text{-}Fec10}$ significantly complemented *E. coli* Fec10 $clpG_{GI}$ and $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ mutants *in vivo* upon exposure to 60 °C in contrast to $ClpG_{GI\text{-}SG17M}$. This implies that the slightly different *in vitro* biochemical characteristics are enhanced *in vivo* or that species-specific factors or substrates impact on $ClpG_{GI}$ activity *in vivo*. In any case, our observations provide the biochemical basis that supports the $clpG_{GI}$ phenotype of increased thermotolerance in Fec10 and other commensal *E. coli* strains, but also suggests species-specific adaptation of $ClpG_{GI}$ proteins.

4 EXPERIMENTAL PROCEDURES

4.1 Strains and growth conditions

The commensal *E. coli* strains used in this study are described in Table 1. Strains were routinely cultivated aerobically at 30 °C or 37 °C with shaking at 120-200 rpm in Luria-Bertani (LB) broth (BD Difco), unless otherwise indicated. If required, 100 µg ml^{-1} ampicillin (Amp), 30 µg ml^{-1} kanamycin (Km), 50 µg ml^{-1} spectinomycin (Spec) or 25 µg ml^{-1} chloramphenicol (Cm) were added. *E. coli* TOP10 was used to propagate recombinant plasmids. Genotypically distinct Fec6, Fec9, Fec10, Fec12, Fec17, Fec23, Fec27, Fec32, Fec34, Fec35, Fec41, Fec51, Fec55, Fec56, Fec59, Fec60, Fec61, Fec65, Fec75, Fec81, Fec89, Fec93, Fec95, Fec97, Fec98, Fec99, Fec100, Fec101, Fec108, Fec110, Fec112, Fec113 (Bokranz et al., 2005) were screened for the presence of $shsp20_{GI}$ as a TLPQC marker by PCR (primers in Table S3).

4.2 Determination of the genome sequence and annotation

To sequence the genomes of *E. coli* Fec6 and Fec10, the genomic DNA was isolated using Qiagen Genomic DNA isolation Kit (500/G binding capacity) according to the user's instructions. Briefly, cells were cultured in 30 ml LB broth to OD600=0.4. Cell lysates were obtained by incubation in bacterial lysis buffer containing RNases A (200 µg/ml) and proteinase K (250 µg/ml). The cell lysate was applied onto a Genomic-tip 500/G by gravity flow. After elution with TE buffer pH 8, the genomic DNA was analyzed on a 0.5% agarose gel and the concentration and purity (260/280 nm ratio) measured by Nanodrop.

Genomic DNA sequencing and assembly was done as described by PacBio and Illumina sequencing (Cimdins *et al.*, 2017a). Illumina raw reads were aligned to the PacBio assembly to correct sequencing errors based on PacBio sequencing inaccuracy. Annotatation were performed using Prokka (Seemann, 2014) and RASTtk (Brettin *et al.*, 2015).

The correct plasmid sequence was identified by matching Illumina reads to the plasmid contig and confirmed by conventional DNA Sanger sequencing of isolated plasmid DNA (primers in table S3). Plasmid circularization was done using the circulator minimus2 option (Hunt *et al.*, 2015). Presence of direct repeats was checked with Gepard (Krumsiek *et al.*, 2007). IS elements were identified using ISfinder (https://isfinder.biotoul.fr) with multiple hits and hits with identities below 95% omitted. DNAplotter (Carver *et al.*, 2009) was used to create the plasmid map. Genome sequences of *E. coli* Fec10 and Fec6 were deposited in the NCBI database under GenBank accession number MDLJ00000000.2 and WIPD00000000.2.

4.3 Genome characterization and bioinformatic analyses

Genome annotation and search for genome characteristics was done using RAST and MicroScope (Brettin et al., 2015, Aziz *et al.*, 2008, Vallenet et al., 2020). Initial strain typing was performed using the ANItools (Han *et al.*, 2016) and the Center for Genomic Epidemiology webservices (http://www.genomicepidemiology.org/). ST-type was determined by MLST 2.0 (Larsen *et al.*, 2012). *In silico* analysis of plasmid replicons was done by PlasmidFinder 2.1 ((https://cge.cbs.dtu.dk/services/PlasmidFinder/); (Carattoli *et al.*, 2014)). Phylogenetic groups were identified by Clermont typing (http://clermontyping.iame-research.center/).

Search for phage sequences was done with PHASTER (Arndt et al., 2016). Protein sequences were aligned with Clustal W and MEGA X was used to construct the maximum likelihood (ML) and neighborhood joining (NJ) phylogenetic tree for $ClpG_{GI}$ proteins (Kumar *et al.*, 2018). Jalview was used to visualize the protein alignment and DNAplotter was used to visualize the plasmid map (Carver et al., 2009).

Additional bioinformatic procedures are found in Supporting Information.

4.4 Construction of gene deletion mutants

Construction of the deletion mutants (Δ*clpG*$_{GI}$, Δ*dhc*$_{GI}$ (*dna-shsp20*$_{GI}$-*ClpG*$_{GI}$), Δ*ttrSR* and Δ*ttrBCA*) in *E. coli* Fec10 was performed via λ Red recombination (Datsenko & Wanner, 2000). The chloramphenicol and kanamycin resistant cassette from the template plasmid pKD3 and pKD4, respectively, flanked by 40 bps (50 bps for *ttr* deletion mutants) upstream of the start codon and 40 bp downstream of the stop codon of the respective genes was amplified with primers as stated in Table S3. After purification and DpnI digestion, the PCR products were transformed into Fec10 harboring the pKD46 plasmid. Mutants were selected on 12 μg ml^{-1} chloramphenicol and subsequently verified by PCR for ORF replacement using primers outside the homologous recombination regions. To cure the pKD46 plasmid, mutants were incubated at 42 °C overnight. Colonies were tested by streaking on LB agar with Cm and Amp and Cm^{R} and Amp^{S} colonies were selected. Deletion mutants were verified by PCR.

4.5 Plasmid construction

Plasmids used in this study are described in Table 2. To express the protein in *E. coli* TOP10 for subsequent purification, *clpG*$_{GI\text{-}Fec10}$ was cloned into the broad-host range vector pJN105 (Newman & Fuqua, 1999) with the L-arabinose inducible P_{BAD} promoter. Primers *clpG*$_{GI}$ _Fec10_NheI_pJN_F/R (Table S3) were used to amplify *clpG*$_{GI\text{-}Fec10}$ with a C-terminal 6xHis tag from *E. coli* Fec10 DNA. After restriction digestion, the amplified fragment was cloned between NheI/XbaI restriction sites of the multiple cloning site using standard procedures. For complementation, *clpG*$_{GI\text{-}Fec10}$, *dhc*$_{GI\text{-}Fec10}$ and *clpG*$_{GI\text{-}SG17M}$ were cloned with a C-terminal 6xHis tag between EcoRI/SalI restriction sites into the pBAD30 vector under the L-arabinose inducible P_{BAD} promoter (Guzman *et al.*, 1995). *ClpG*$_{GI\text{-}Fec10}$ was cloned via BamHI/XbaI restriction sites into the pUHE21 vector for complementation studies in *E. coli ΔclpB* and *dnaK103* chaperone mutants. Primers are listed in Table S3.

4.6 Rdar biofilm assay

10 μl overnight bacterial culture was spotted onto a 25 ml LB without NaCl agar plate supplemented with CR (40 μg ml−1) and Coomassie brilliant blue (20 μg ml−1) (CR agar plate) or 50 μg ml−1 Calcoflour white (fluorescence brightener 28) (Römling *et al.*, 1998a). When

indicated, sodium tetrathionate (1.5 mM or 15 mM) was added. Plates were incubated at 28 °C and 37 °C for 48 h and the colony morphotype was documented.

4.7 Lethal-temperature-tolerance assay

Cells grown in LB broth containing 0.1% L-arabinose and 100 µg ml^{-1} ampicillin with shaking at 200 rpm at 37 °C were harvested after 18 h. OD_{600} was adjusted to one in the same medium and 500 µl of the cell suspension was incubated at 55 °C, 60 °C and 65 °C for the indicated amount of time, while a control cell suspension was kept on ice. After preparation of 10-fold serial dilutions in LB medium, cell viability was assessed by the spot assay. 10 µl of cell suspension of each dilution was spotted onto LB agar and plates were incubated at 37 °C for 14-15 h.

Viability assays for *E. coli* K-12 Δ*clpB* and *dnaK103* cells harboring the plasmid placIq and pUHE21-derivatives after IPTG-controlled expression of *clpB*, *dnaK* and $clpG_{GI}$ were performed as described previously (Lee et al., 2018).

4.8 Protein purification

$ClpG_{GI\text{-}Fec10}$ with a C-terminal 6xHis-tag was expressed from the pJN105 plasmid in *E. coli* TOP10 (Table 2). Cultures were incubated at 37 °C, expression induced at OD_{600} = 0.5 by 0.1% L-arabinose and cells subsequently harvested 6 h post induction. Cells were broken with a French press and the cell debris removed by centrifugation at 15 000 g for 45 min. The protein was purified according to standard procedure using Ni-NTA (Quiagen) or Ni-IDA (Protino). The purified protein was subsequently subjected to Superdex S200 size exclusion chromatography in MDH assay buffer supplemented with 5% (v/v) glycerol. Afterwards, dialysis was performed in 50 mM Na_2HPO_4, 300 mM NaCl and 5% glycerol, pH 8 overnight at 4°C. The protein purity was assessed by SDS-PAGE and the protein concentration was determined by the Bradford assay using a BSA calibration curve. Protein concentrations refer to monomers. $ClpG_{GI\text{-}SG17M}$, ClpB, DnaK, DnaJ, GrpE and firefly Luciferase were purified as described before (Katikaridis et al., 2019; Lee et al., 2018).

4.9 *In vitro* disaggregation activity assays

Disaggregation assays were performed as described previously with modifications (Mogk *et al.*, 2003). The model substrates used to test the disaggregation activity were malate dehydrogenase (MDH; Roche) and firefly luciferase (Roche). In brief, 2 µM MDH and 0.2 µM luciferase were heat aggregated at 47 °C for 30 min and 45 °C for 15 min, respectively, in MDH assay buffer

(50 mM Tris pH 7.5, 20 mM $MgCl_2$, 150 mM KCl and 2 mM DTT) and subsequently incubated at room temperature for 5 min. 50 µl of aggregated MDH and luciferase (final concentration 1 µM and 0.1 µM, respectively) was mixed with 54 µl of assay buffer containing 2 µM disaggregase. In the MDH assay, 1 µM each GroEL/GroES was added to aid refolding of the enzyme. The KJE mix consisted of 1 µM DnaK, 0.2 µM DnaJ, 0.1 µM GrpE. The reaction was started by adding 6 µl ATP regeneration system (2 mM ATP, 3 mM phosphoenolpyruvate and 20 ng/µl pyruvate kinase) and samples were incubated at 30 °C for disaggregation and refolding. To determine the recovered MDH activity, consumption of NADH was measured by monitoring the decrease in absorbance at 340 nm using an UV spectrophotometer (Novaspec Plus, Amersham). For each MDH activity assay, 10 µl sample was mixed with 690 µl measurement buffer (150 mM potassium phosphate pH 7.6, 2 mM DTT, 0.5 mM oxaloacetate and 0.28 mM NADH). To assess luciferase activity, 2 µl of sample was mixed with 125 µl 2X luciferase assay buffer (25 mM glycylglycine pH 7.4, 12.5 mM $MgSO_4$, 5 mM ATP) and 125 µl 25 mM luciferin (Gold Biotech.) and activity measured by Berthold Lumat LB 9507. Similarly, 10 and 2 µl of native MDH and luciferase, respectively, were mixed with measurement buffer. The value corresponds to 100% substrate activity before heat denaturation.

4.10 ATPase assay

The ATPase assay is based on the coupled enzymatic activity of pyruvate kinase (PK) and lactate dehydrogenase (LDH). The disaggregase hydrolyzes ATP to ADP that is recycled by PK to convert phosphoenolpyruvate (PEP) to pyruvate. LDH in turn converts pyruvate into lactate thereby oxidizing NADH to NAD+. The decrease in NADH corresponds to the ATPase rate of the disaggregases and is recorded by the decrease in absorbance at 340 nm in a 96 well plate by FLUOstar-Omega (BMG-Labtech). One µM disaggregase was assessed in 100 µl MDH assay buffer containing 10 µl 10X reaction mix (50 µl 5 mM NADH, 50 µl 10 mM PEP (Sigma) and 50 µl PK/LDH (Sigma)). Optionally, casein was added to a final concentration of 0.1 µg/µl to stimulate the ATPase activity of ClpB. The reaction was started by adding 100 µl of 4 mM ATP in MDH assay buffer.

4.11 Circular Dichroism (CD) spectroscopy

Circular dichroism spectroscopy was used to assess the thermal stability of the ClpB and $ClpG_{GI}$ proteins. The disaggregases were dialyzed against 10 mM potassium phosphate buffer pH 7.4 and diluted to a final concentration of 3 µM. Optionally, $MgAc_2$ and ATPγS was added to a final concentration of 5 mM and 1 mM, respectively. CD spectra were recorded from 190 to 250

nm for the different disaggregases using a Jasco J750 spectropolarimeter. The melting curves of disaggregases were obtained by increasing the temperature at a rate of 0.5 °C per minute from 20 °C to 85 °C with the circular dichroism signal recorded at 222 nm. T_M values were calculated using Prism software.

4.12 Anisotropy measurements

Binding of ClpB, $ClpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$ to 100 nM FITC-casein was monitored by fluorescence anisotropy measurements using a BMG Biotech CLARIOstar plate reader. Samples were incubated in MDH assay buffer for 5 min at 30 °C in the presence of 2 mM ATPγS and polarization of FITC-casein was determined in black 384 well plates (excitation: 482 nm; emission: 530 nm, Target mP: 35). A sample containing FITC-casein only served as reference. K_d values were determined using nonlinear regression curve fitting (Prism software).

4.13 Transmission electron microscopy

A detailed description of this experimental procedure is found in Supporting Information.

ACKNOWLEDGEMENTS

Mikhail S. Gelfand supervised Zaira Seferbekova and Robert Afasizhev under RSF grant 18-14-00358. Inge Kristen (Helmholtz Center for Infection Research, Braunschweig, Germany) is gratefully thanked for excellent electron microscopic sample preparation. Shady Mansour Kamal received a travel grant from the Swedish Society to conduct work at the University of Heidelberg, Germany. Panagiotis Katikaridis was supported by the Heidelberg Biosciences International Graduate School (HBIGS). This work was supported by the Swedish Research Council for Natural Sciences and Engineering (project number 621-2013-4809) to U.R. and by a grant of the Deutsche Forschungsgemeinschaft (MO970/4-3) to A.M. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

AUTHOR CONTRIBUTIONS

Conception of the study: UR and AM; acquisition of data: SMK, HL, CL, ACA, FL, AJM-R, ZS, RA, HTW, LM, PK, AM, UR; analysis of data: SMK, HL, ACA, FL, AJM-R, HTW, ZS, RA, PK, UD, AM, UR; interpretation of the data: SMK, HL, ACA, AJM-R, UD, AM, UR; writing of the manuscript: UR and SMK. All authors contributed to the revision and commented on the final version of the manuscript.

ORCID

Shady Mansour Kamal, ID: 0000-0002-7565-811X

Annika Cimdins-Ahne, ID: 0000-0001-7414-9914

Changhan Lee, ID: 0000-0003-0327-4712

Heinrich Lünsdorf, ID: 0000-0003-2778-8611

Zaira Seferbekova, ID: 0000-0002-3131-3697

Haleluya Wami, ID: 0000-0002-9929-2570

Lena Meins, ID: 0000-0001-6214-477X

Panagiotis Katikaridis, ID: 0000-0001-5783-1487

Alberto Jonatan Martin-Rodriguez, ORCID ID: 0000-0003-2422-129X

Robert Afasizhev, ID: orcid.org/0000-0002-3513-8873

Ulrich Dobrindt, ID: 0000-0001-9949-1898

Axel Mogk: ID: 0000-0003-3674-5410

Ute Römling: ID: 0000-0003-3812-6621

DATA AVAILABILITY STATEMENT

Material and strains are available upon request.

FIGURE LEGENDS

FIGURE 1 Unrooted phylogenetic tree of the species *E. coli.* The tree is constructed with 521 complete *E. coli* genomes using 238 orthologous genes. The two new *E. coli* strains (Fec10 and Fec6) are shown in red, and the cluster of *E. coli* K-12 strains is shown in blue. Tree visualized with iTOL https://itol.embl.de/ (Letunic and Bork 2016).

FIGURE 2 Plasmid map of pFec10. Coding regions are indicated in black (sense strand) or grey (anti-sense strand). Origin of replication (ori) and major cargo genes mentioned in the text are highlighted. TLPQC, transmissible locus of protein quality control. IS element as identified by ISfinder (http://www-is-biotoul.fr; (Siguier *et al.*, 2006)) are highlighted in red. Numbers refer to IS elements as listed in Table S1 and Table S2. The figure was created by DNAplotter and manually modified.

FIGURE 3 Effect of tetrathionate supplementation on rdar biofilm colony morphology of strain Fec10 on CR agar plates. (A) Arrangement of the tetrathionate gene cluster with the constructed mutants indicated. *E. coli* Fec10 WT, Δ*ttrSR*, Δ*ttrBCA* and Δ*ttrSR* Δ*ttrBCA* colony biofilms after incubation on CR agar plates supplemented with sodium tetrathionate (1.5 or 15 mM) for 48 h at 28 °C (B-D) or 37 °C (E-G).

FIGURE 4 $ClpG_{GI\text{-}Fec10}$ contributes to heat tolerance in *E. coli* Fec10 and restores heat tolerance in *E. coli* MC4100 in the absence of *dnaK* and *clpB*. (A) Assessment of heat shock tolerance in the *E. coli* Fec10 $clpG_{GI\text{-}Fec10}$ and $dhc_{GI\text{-}Fec10}$ deletion mutants and respective complementation. Bacterial cells were exposed to 60 °C for 7.5 and 15 min. p=plasmid pBAD30; p$clpG_{GI\text{-}Fec10}$ = $clpG_{GI\text{-}Fec10}$ cloned in plasmid pBAD30. p$dhc_{GI\text{-}Fec10}$ = $dna\text{-}shsp20_{GI}\text{-}clpG_{GI}$ cloned in plasmid pBAD30. (B) Western blot analysis of the production level of $ClpG_{GI}$ and ClpB in *E coli* Fec10 and Fec6 in logarithmic and stationary phase of growth at 37 °C. Detection of protein production was done using antisera generated against $ClpG_{GI\text{-}SG17M}$ and $ClpB_{K\text{-}12}$. (C) Complementation of the heat shock tolerance of the *E. coli* MC4100 deletion mutant of *dnaK* (upper panel) and *clpB* (lower panel) in MC4100 with plasmids expressing *dnaK, clpB,* $clpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$. Cells were exposed to 50 °C for the indicated amount of time. Vc= vector control pUHE21; p*dnaK*=*dnaK* cloned in pUHE21; p*clpB*=*clpB* cloned in pUHE21; p$clpG_{GI\text{-}Fec10}$=$clpG_{GI\text{-}Fec10}$ cloned in pUHE21; p$clpG_{GI\text{-}SG17M}$=$clpG_{GI\text{-}SG17M}$ cloned in pUHE21. (D) SDS-PAGE gel of the expression of ClpB, $ClpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$ from the complementation experiment in panel (C). The positions of the proteins on the gel are indicated.

FIGURE 5 $ClpG_{GI\text{-}Fec10}$ disaggregates heat-folded luciferase independent of accessory chaperons and demonstrates high basal ATPase activity. (A) Refolding of aggregated luciferase was monitored over 120 min. The activity of the native luciferase was set to 100%. One representative experiment is shown. (B) Recovery of luciferase after 120 min incubation with the indicated disaggregase. The mean value was calculated from three independent experiments

766 with 9 technical replicates. Error bars indicate SD (****P<0.0001). (C) ATPase rates were
767 calculated with and without casein, the stimulator substrate for ClpB ATPase activity. The mean
768 value was calculated from 3 independent experiments with 9 technical replicates. Error bars
769 indicate SD (****P<0.0001). (D) Binding of ClpB, $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$ to
770 fluorescein-labeled FITC-casein was determined in the presence of 2 mM ATPγS by anisotropy
771 measurements.
772
773 **FIGURE 6** Determination of the melting temperature for $ClpB_{K\text{-}12}$, $ClpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$. Circular Dichroism melting temperature was recorded in the absence (black bars) and
775 presence (red bars) of Mg^{2+}/ATPγS. The change in differential absorption of circular polarized
776 light upon secondary structure alteration to random coil was recorded at 222 nm. The melting
777 temperatures of the indicated enzymes are shown above the bars.
778
779
780
781 **TABLE 1** *E. coli* strains used in this study

Strain	Genotype/Source/Characteristic	Reference
Commensal *E. coli* isolates and derivatives		
Fec6	commensal strain/human feces	(Bokranz et al., 2005)
Fec10	commensal strain/human feces	(Bokranz et al., 2005)
Fec10 Δ$clpG_{GI}$	deletion of disaggregase $ClpG_{GI}$	This study
Fec10 Δdhc_{GI}	deletion of *dna-shsp20*$_{GI}$-*clpG*$_{GI}$ operon	This study
Fec10 Δ*ttrSR*	deletion of two component system TtrSR	This study
Fec10 Δ*ttrBCA*[1]	deletion of tetrathionate reductase TtrBCA	This study
Fec10 Δ*ttrSR* Δ*ttrBCA*[1]	double deletion of tetrathionate reductase regulatory and structural genes	This study

Fec32	commensal strain/human feces	(Bokranz et al., 2005)
Fec41	commensal strain/human feces	(Bokranz et al., 2005)
Fec55	commensal strain/human feces	(Bokranz et al., 2005)
Fec75	commensal strain/human feces	(Bokranz et al., 2005)
Fec89	commensal strain/human feces	(Bokranz et al., 2005)
Fec113	commensal strain/human feces	(Bokranz et al., 2005)
TOB1	commensal strain/human feces	(Bokranz et al., 2005)
E. coli K-12 derivatives		
XL1 blue	XL1 blue placIq	(Weibezahn *et al.*, 2004)
MC4100 Δ*clpB*	MC4100 △*clpB*::Km placIq	(Weibezahn et al., 2004)
MC4100 Δ*dnaK103*	MC4100 Δ*dnaK103* placIq	(Winkler *et al.*, 2010)
TOP10	cloning and propagating plasmids F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen

782 *deletion of the *ttrBAC* operon included a hypothetical coding sequence located immediately
783 upstream *ttrB*
784
785
786
787
788
789

TABLE 2 Plasmids used in this study

Plasmid	Description	Source/Reference
pKD3	template plasmid for λ Red mediated	(Datsenko & Wanner,

	recombination, Cm^r, Amp^r	2000)
pKD46	P_{BAD} promoter, encodes λ Red recombinase, Amp^r	(Datsenko & Wanner, 2000)
pBAD30	cloning vector, pACYC origin, L-arabinose inducible *araBAD* promoter, Amp^r	(Guzman et al., 1995)
pJN105	broad-host range vector with L-arabinose inducible *araBAD* promoter; pBBR1ori, Gm^r	(Newman & Fuqua, 1999)
pBAD30 $clpG_{GI\text{-}Fec10}$	Amp^r; $clpG_{GI\text{-}Fec10}$ cloned with C-terminal 6xHis tag	This study
pBAD30 $dhc_{GI\text{-}Fec10}$	Amp^r; $dhc_{GI\text{-}Fec10}$ cloned with C-terminal 6xHis tag	This study
pJN105 $clpG_{GI\text{-}Fec10}$	Gm^r; $clpG_{GI\text{-}Fec10}$ cloned with C-terminal 6xHis tag	This study
pBAD30 $clpG_{GI\text{-}SG17M}$	Amp^r; $clpG_{GI\text{-}SG17M}$ cloned with C-terminal 6xHis tag	This study
placIq	$Spec^r$, harbors lacIq to allow for IPTG controlled gene expression, p15A ori	Lee et al., 2018
pUHE21	Amp^r, IPTG-inducible Ptac promoter, pBR322 ori	Lee et al., 2018
pUHE21 *clpB*	Amp^r, *clpB* cloned into BamHI/XbaI restriction sites	Lee et al., 2018
pUHE21 *dnaK*	Amp^r, *dnaK* cloned into BamHI/HindIII restriction sites	Lee et al., 2018
pUHE21 $clpG_{GI\text{-}Fec10}$	Amp^r, $clpG_{GI\text{-}Fec10}$ cloned into BamHI/XbaI restriction sites	This study
pUHE21 $clpG_{GI\text{-}SG17M}$	Amp^r, $clpG_{GI\text{-}SG17M}$ cloned into BamHI/XbaI restriction sites	Lee et al., 2018
pET24a $clpG_{GI\text{-}SG17M}$	Kan^r, $clpG_{GI\text{-}SG17M}$ with C-terminal 6x His tag	Lee et al., 2018

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers & D.J. Lipman (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Arndt, D., J.R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang & D.S. Wishart (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16-21.

Aziz, R.K., D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke & O. Zagnitko (2008) The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**: 75.

Bachmann, B.J. (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* **36**: 525-557.

Beghain, J., A. Bridier-Nahmias, H. Le Nagard, E. Denamur & O. Clermont (2018) ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb Genom* **4**:e000192.

Bettelheim, K.A. & S.M. Lennox-King (1976) The acquisition of *Escherichia coli* by new-born babies. *Infection* **4**: 174-179.

Blount, Z.D. (2015) The unexhausted potential of *E. coli. Elife* **4**:e05826.

Bojer, M.S., A.M. Hammerum, S.L. Jorgensen, F. Hansen, S.S. Olsen, K.A. Krogfelt & C. Struve (2012) Concurrent emergence of multidrug resistance and heat resistance by CTX-M-15-encoding conjugative plasmids in *Klebsiella pneumoniae*. *APMIS* **120**: 699-705.

Bojer, M.S., K.A. Krogfelt & C. Struve (2011) The newly discovered ClpK protein strongly promotes survival of *Klebsiella pneumoniae* biofilm subjected to heat shock. *J Med Microbiol* **60**: 1559-1561.

Bojer, M.S., C. Struve, H. Ingmer, D.S. Hansen & K.A. Krogfelt (2010) Heat resistance mediated by a new plasmid encoded Clp ATPase, ClpK, as a possible novel mechanism for nosocomial persistence of *Klebsiella pneumoniae. PloS One* **5**: e15467.

Bokranz, W., X. Wang, H. Tschäpe & U. Römling (2005) Expression of cellulose and curli fimbriae by *Escherichia coli* isolated from the gastrointestinal tract. *J Med Microbiol* **54**: 1171-1182.

Boll, E.J., R. Marti, H. Hasman, S. Overballe-Petersen, M. Stegger, K. Ng, S. Knochel, K.A. Krogfelt, J. Hummerjohann & C. Struve (2017) Turn up the heat-food and clinical *Escherichia coli* isolates feature two transferrable loci of heat resistance. *Front Microbiol* **8**: 579.

Brettin, T., J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, G.J. Olsen, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, J.A. Thomason, 3rd, R. Stevens, V. Vonstein, A.R. Wattam & F. Xia (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **5**: 8365.

Browning, D.F., T.J. Wells, F.L. Franca, F.C. Morris, Y.R. Sevastsyanovich, J.A. Bryant, M.D. Johnson, P.A. Lund, A.F. Cunningham, J.L. Hobman, R.C. May, M.A. Webber & I.R. Henderson (2013) Laboratory adapted *Escherichia coli* K-12 becomes a pathogen of C*aenorhabditis elegans* upon restoration of O antigen biosynthesis. *Mol Microbiol* **87**: 939-950.

Carattoli, A., E. Zankari, A. Garcia-Fernandez, M. Voldby Larsen, O. Lund, L. Villa, F. Moller Aarestrup & H. Hasman (2014) In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* **58**: 3895-3903.

Carver, T., N. Thomson, A. Bleasby, M. Berriman & J. Parkhill (2009) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**: 119-120.

Chai, Y., R. Kolter & R. Losick (2009) A widely conserved gene cluster required for lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation. *J Bacteriol* **191**: 2423-2430.

Cimdins, A., P. Lüthje, F. Li, I. Ahmad, A. Brauner & U. Römling (2017a) Draft genome sequences of semiconstitutive red, dry, and rough biofilm-forming commensal and uropathogenic *Escherichia coli* isolates. *Genome Announc* **5**:e01249-16.

Cimdins, A., R. Simm, F. Li, P. Lüthje, K. Thorell, A. Sjöling, A. Brauner & U. Römling (2017b) Alterations of c-di-GMP turnover proteins modulate semi-constitutive rdar biofilm formation in commensal and uropathogenic *Escherichia coli. Microbiologyopen* **6**:e00508.

Clermont, O., S. Bonacorsi & E. Bingen (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**: 4555-4558.

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

Day, M.J., K.L. Hopkins, D.W. Wareham, M.A. Toleman, N. Elviss, L. Randall, C. Teale, P. Cleary, C. Wiuff, M. Doumith, M.J. Ellington, N. Woodford & D.M. Livermore (2019) Extended-spectrum beta-lactamase-producing *Escherichia coli* in human-derived

and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. *Lancet Infect Dis* **19**: 1325-1335.

Dlusskaya, E.A., L.M. McMullen & M.G. Gänzle (2011) Characterization of an extremely heat-resistant *Escherichia coli* obtained from a beef processing facility. *J Appl Microbiol* **110**: 840-849.

Dublan Mde, L., J.C. Ortiz-Marquez, L. Lett & L. Curatti (2014) Plant-adapted *Escherichia coli* show increased lettuce colonizing ability, resistance to oxidative stress and chemotactic response. *PloS One* **9**: e110416.

Escobar-Paramo, P., A. Le Menac'h, T. Le Gall, C. Amorin, S. Gouriou, B. Picard, D. Skurnik & E. Denamur (2006) Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ Microbiol* **8**: 1975-1984.

Freitag, C., G.B. Michael, J. Li, K. Kadlec, Y. Wang, M. Hassel & S. Schwarz (2018) Occurrence and characterisation of ESBL-encoding plasmids among *Escherichia coli* isolates from fresh vegetables. *Vet Microbiol* **219**: 63-69.

Guzman, L.M., D. Belin, M.J. Carson & J. Beckwith (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.

Han, N., Y. Qiang & W. Zhang (2016) ANItools web: a web tool for fast genome comparison within multiple bacterial strains. *Database (Oxford)* **2016**.

Hendriksen, R.S., V. Bortolaia, H. Tate, G.H. Tyson, F.M. Aarestrup & P.F. McDermott (2019) Using genomics to track global antimicrobial resistance. *Front Public Health* **7**: 242.

Hensel, M., A.P. Hinsley, T. Nikolaus, G. Sawers & B.C. Berks (1999) The genetic basis of tetrathionate respiration in *Salmonella typhimurium*. *Mol Microbiol* **32**: 275-287.

Hunt, M., N.D. Silva, T.D. Otto, J. Parkhill, J.A. Keane & S.R. Harris (2015) Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* **16**: 294.

Jorgensen, S.B., A.V. Soraas, L.S. Arnesen, T.M. Leegaard, A. Sundsfjord & P.A. Jenum (2017) A comparison of extended spectrum beta-lactamase producing *Escherichia coli* from clinical, recreational water and wastewater samples associated in time and location. *PloS One* **12**: e0186576.

Kapralek, F. (1972) The physiological role of tetrathionate respiration in growing citrobacter. *J Gen Microbiol* **71**: 133-139.

Katikaridis, P., Mein, L., Mansour Kamal, S., Römling, U., Mogk, A. (2019) ClpG provides increased heat resistance by acting as superior disaggregase. *Biomolecules* **9**:815.

Khetrapal, V., K.S. Mehershahi & S.L. Chen (2017) Complete genome sequence of the original *Escherichia coli* isolate, strain NCTC86. *Genome Announc* **5**:e00243-17.

Koraimann, G. (2018) Spread and persistence of virulence and antibiotic resistance genes: A ride on the F plasmid conjugation module. *EcoSal Plus* **8**. doi: 10.1128/ecosalplus.-ESP-0003-2018.

Krumsiek, J., R. Arnold & T. Rattei (2007) Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**: 1026-1028.

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

Larsen, M.V., S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R.L. Marvig, L. Jelsbak, T. Sicheritz-Ponten, D.W. Ussery, F.M. Aarestrup & O. Lund (2012) Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* **50**: 1355-1361.

Lee, C., K.B. Franke, S.M. Kamal, H. Kim, H. Lünsdorf, J. Jager, M. Nimtz, J. Trcek, L. Jansch, B. Bukau, A. Mogk & U. Römling (2018) Stand-alone ClpG disaggregase confers superior heat tolerance to bacteria. *Proc Natl Acad Sci U. S. A.* **115**: E273-E282.

Lee, C., E. Wigren, H. Lünsdorf & U. Römling (2016) Protein homeostasis-more than resisting a hot bath. *Curr Opin Microbiol* **30**: 147-154.

Lee, C., E. Wigren, J. Trcek, V. Peters, J. Kim, M.S. Hasni, M. Nimtz, Y. Lindqvist, C. Park, U. Curth, H. Lünsdorf & U. Römling (2015) A novel protein quality control mechanism contributes to heat shock resistance of worldwide-distributed *Pseudomonas aeruginosa* clone C strains. *Environ Microbiol*.

Lehnherr, H., E. Maguin, S. Jafri & M.B. Yarmolinsky (1993) Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *J Mol Biol* **233**: 414-428.

Lescat, M., O. Clermont, P.L. Woerther, J. Glodt, S. Dion, D. Skurnik, F. Djossou, C. Dupont, G. Perroz, B. Picard, F. Catzeflis, A. Andremont & E. Denamur (2013) Commensal *Escherichia coli* strains in Guiana reveal a high genetic diversity with host-dependant population structure. *Environ Microbiol Rep* **5**: 49-57.

Li, H. & M. Gänzle (2016) Some like it hot: Heat resistance of *Escherichia coli* in food. *Front Microbiol* **7**: 1763.

Li, H., R. Mercer, J. Behr, S. Heinzlmeir, L.M. McMullen, R.F. Vogel & M.G. Gänzle (2020) Heat and pressure resistance in *Escherichia coli* relates to protein folding and aggregation. *Front Microbiol* **11**: 111.

Lin, L., X. Wang, L. Cao & M. Xu (2019) Lignin catabolic pathways reveal unique characteristics of dye-decolorizing peroxidases in *Pseudomonas putida*. *Environ Microbiol* **21**: 1847-1863.

Liu, M., Y. Zhang, M. Inouye & N.A. Woychik (2008) Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. *Proc Natl Acad Sci U. S. A.* **105**: 5885-5890.

Lobocka, M.B., D.J. Rose, G. Plunkett, 3rd, M. Rusin, A. Samojedny, H. Lehnherr, M.B. Yarmolinsky & F.R. Blattner (2004) Genome of bacteriophage P1. *J Bacteriol* **186**: 7032-7068.

Manges, A.R., H.M. Geum, A. Guo, T.J. Edens, C.D. Fibke & J.D.D. Pitout (2019) Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin Microbiol Rev* **32**:e00135-18.

Manges, A.R., J. Harel, L. Masson, T.J. Edens, A. Portt, R.J. Reid-Smith, G.G. Zhanel, A.M. Kropinski & P. Boerlin (2015) Multilocus sequence typing and virulence gene profiles associated with *Escherichia coli* from human and animal sources. *Foodborne Pathog Dis* **12**: 302-310.

Martin-Rodriguez, A.J., M. Rhen, K. Melican & A. Richter-Dahlfors (2020) Nitrate metabolism modulates biosynthesis of biofilm components in uropathogenic *Escherichia coli* and acts as a fitness factor during experimental urinary tract infection. *Front Microbiol* **11**: 26.

Massot, M., A.S. Daubie, O. Clermont, F. Jaureguy, C. Couffignal, G. Dahbi, A. Mora, J. Blanco, C. Branger, F. Mentre, A. Eddi, B. Picard, E. Denamur & G. The Coliville (2016) Phylogenetic, virulence and antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from community subjects in the Paris area in 2010 and evolution over 30 years. *Microbiology* **162**: 642-650.

Meric, G., E.K. Kemsley, D. Falush, E.J. Saggers & S. Lucchini (2013) Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environ Microbiol* **15**: 487-501.

963 Mogk, A., E. Deuerling, S. Vorderwulbecke, E. Vierling & B. Bukau (2003) Small heat shock
964 proteins, ClpB and the DnaK system form a functional triade in reversing protein
965 aggregation. *Mol Microbiol* **50**: 585-595.
966 Mogk, A., T. Tomoyasu, P. Goloubinoff, S. Rudiger, D. Roder, H. Langen & B. Bukau (1999)
967 Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of
968 aggregation by DnaK and ClpB. *EMBO J* **18**: 6934-6949.
969 Newman, J.R. & C. Fuqua (1999) Broad-host-range expression vectors that carry the l-
970 arabinose-inducible *Escherichia coli araBAD* promoter and the *araC* regulator. *Gene*
971 **227**: 197-203.
972 Nguyen, S.V., G.P. Harhay, J.L. Bono, T.P. Smith & D.M. Harhay (2017) Genome sequence of
973 the thermotolerant foodborne pathogen *Salmonella enterica* serovar Senftenberg
974 ATCC 43845 and phylogenetic analysis of loci encoding increased protein quality
975 control mechanisms. *mSystems* **2**:e00190-16.
976 Palleros, D.R., K.L. Reid, J.S. McCarty, G.C. Walker & A.L. Fink (1992) DnaK, hsp73, and their
977 molten globules. Two different ways heat shock proteins respond to heat. *J Biol*
978 *Chem* **267**: 5279-5285.
979 Palumbo, S.A. & J.A. Alford (1970) Inhibitory action of tetrathionate enrichment broth.
980 *Appl Microbiol* **20**: 970-976.
981 Peng, S., Stephan, R., Hummerjohann, J., Bianco, J., Zweifel, C. (2012) In vitro
982 characterization of Shiga toxin-producing and generic *Escherichia coli* in respect of
983 cheese-production relevant stresses. *J Food Saf Food Qual* **63**: 136-141.
984 Pilla, G. & C.M. Tang (2018) Going around in circles: virulence plasmids in enteric
985 pathogens. *Nat Rev Microbiol* **16**: 484-495.
986 Reid, C.J., E.R. Wyrsch, P. Roy Chowdhury, T. Zingali, M. Liu, A.E. Darling, T.A. Chapman &
987 S.P. Djordjevic (2017) Porcine commensal *Escherichia coli*: a reservoir for class 1
988 integrons associated with IS26. *Microb Genom* **3**:e000143.
989 Richter, T.K.S., T.H. Hazen, D. Lam, C.L. Coles, J.C. Seidman, Y. You, E.K. Silbergeld, C.M.
990 Fraser & D.A. Rasko (2018) Temporal variability of *Escherichia coli* diversity in the
991 gastrointestinal tracts of Tanzanian children with and without exposure to
992 antibiotics. *mSphere* **3**:e00558-18.
993 Römling, U., Z. Bian, M. Hammar, W.D. Sierralta & S. Normark (1998a) Curli fibers are
994 highly conserved between *Salmonella typhimurium* and *Escherichia coli* with
995 respect to operon structure and regulation. *J Bacteriol* **180**: 722-731.

Römling, U., W.D. Sierralta, K. Eriksson & S. Normark (1998b) Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the *agfD* promoter. *Mol Microbiol* **28**: 249-264.

Rossi, E., A. Cimdins, P. Lüthje, A. Brauner, A. Sjöling, P. Landini & U. Römling (2018) "It's a gut feeling" - *Escherichia coli* biofilm formation in the gastrointestinal tract environment. *Crit Rev Microbiol* **44**: 1-30.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

Shepard, S.M., J.L. Danzeisen, R.E. Isaacson, T. Seemann, M. Achtman & T.J. Johnson (2012) Genome sequences and phylogenetic analysis of K88- and F18-positive porcine enterotoxigenic *Escherichia coli. J Bacteriol* **194**: 395-405.

Shintani, M., Y. Takahashi, H. Yamane & H. Nojiri (2010) The behavior and significance of degradative plasmids belonging to Inc groups in *Pseudomonas* within natural environments and microcosms. *Microbes Environ* **25**: 253-265.

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon & M. Chandler (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32-36.

Singh, R., J.C. Grigg, W. Qin, J.F. Kadla, M.E. Murphy & L.D. Eltis (2013) Improved manganese-oxidizing activity of DypB, a peroxidase from a lignolytic bacterium. *ACS Chem Biol* **8**: 700-706.

Stecher, B., R. Denzler, L. Maier, F. Bernet, M.J. Sanders, D.J. Pickard, M. Barthel, A.M. Westendorf, K.A. Krogfelt, A.W. Walker, M. Ackermann, U. Dobrindt, N.R. Thomson & W.D. Hardt (2012) Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc Natl Acad Sci U. S. A.* **109**: 1269-1274.

Sternberg, N. & R. Hoess (1983) The molecular genetics of bacteriophage P1. *Annu Rev Genet* **17**: 123-154.

Tenaillon, O., D. Skurnik, B. Picard & E. Denamur (2010) The population genetics of commensal *Escherichia coli. Nat Rev Microbiol* **8**: 207-217.

Tracey, J.C., M. Coronado, T.W. Giessen, M.C.Y. Lau, P.A. Silver & B.B. Ward (2019) The discovery of twenty-eight new encapsulin sequences, including three in Anammox bacteria. *Sci Rep* **9**: 20122.

Vallenet, D., A. Calteau, M. Dubois, P. Amours, A. Bazin, M. Beuvin, L. Burlot, X. Bussell, S. Fouteau, G. Gautreau, A. Lajus, J. Langlois, R. Planel, D. Roche, J. Rollin, Z. Rouy, V.

Sabatet & C. Medigue (2020) MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res* **48**: D579-D589.

Venturini, C., T. Zingali, E.R. Wyrsch, B. Bowring, J. Iredell, S.R. Partridge & S.P. Djordjevic (2019) Diversity of P1 phage-like elements in multidrug resistant *Escherichia coli. Sci Rep* **9**: 18861.

von Mentzer, A., T.R. Connor, L.H. Wieler, T. Semmler, A. Iguchi, N.R. Thomson, D.A. Rasko, E. Joffre, J. Corander, D. Pickard, G. Wiklund, A.M. Svennerholm, A. Sjoling & G. Dougan (2014) Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nat Genet* **46**: 1321-1326.

Wang, Z., Y. Fang, S. Zhi, D.J. Simpson, A. Gill, L.M. McMullen, N.F. Neumann & M.G. Gänzle (2020) The locus of heat resistance confers resistance to chlorine and other oxidizing chemicals in *Escherichia coli. Appl Environ Microbiol* **86**:e02123-19.

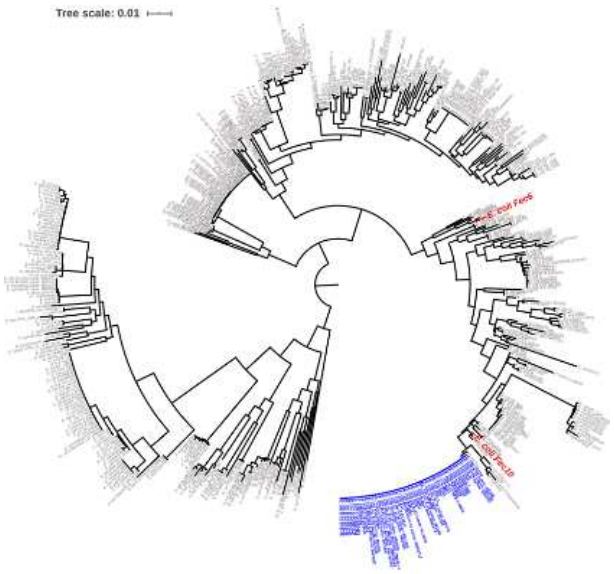
Wang, Z., P. Li, L. Luo, D.J. Simpson & M.G. Gänzle, (2018) Daqu fermentation selects for heat-resistant Enterobacteriaceae and Bacilli. *Appl Environ Microbiol* **84**:e01483-18.

Weibezahn, J., P. Tessarz, C. Schlieker, R. Zahn, Z. Maglica, S. Lee, H. Zentgraf, E.U. Weber-Ban, D.A. Dougan, F.T. Tsai, A. Mogk & B. Bukau, (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **119**: 653-665.

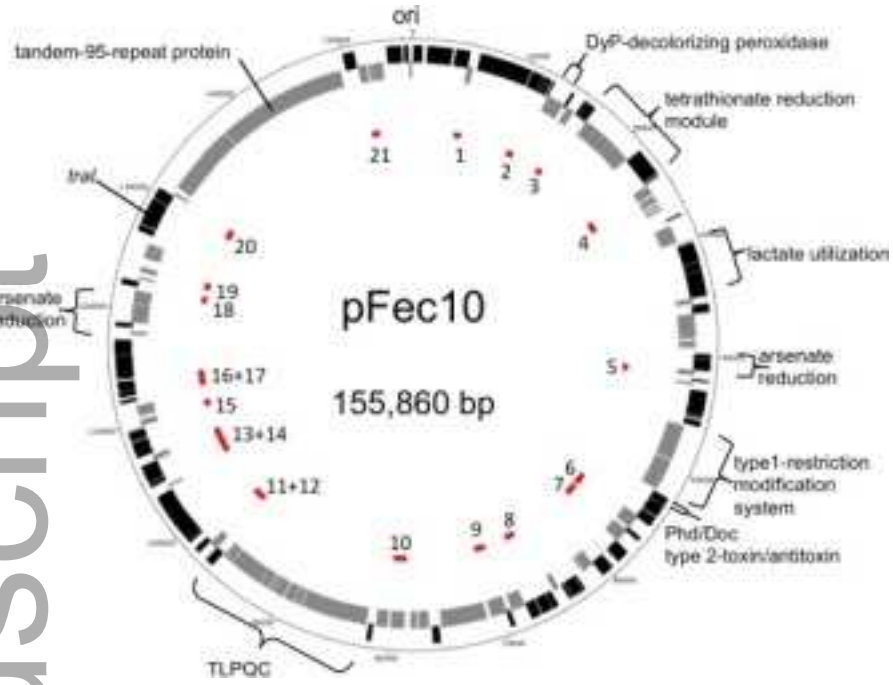
Winkler, J., A. Seybert, L. König, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A.S. Frangakis, A. Mogk & B. Bukau, (2010) Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J* **29**: 910-923.

Winter, S.E. & A.J. Bäumler, (2011) A breathtaking feat: to compete with the gut microbiota, *Salmonella* drives its host to provide a respiratory electron acceptor. *Gut Microbes* **2**: 58-60.

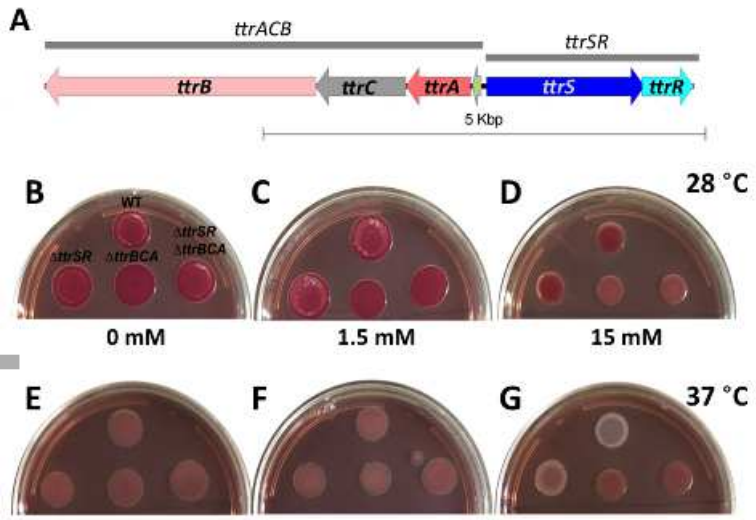
Zolkiewski, M., (1999) ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from *Escherichia coli*. *J Biol Chem* **274**: 28083-28086.



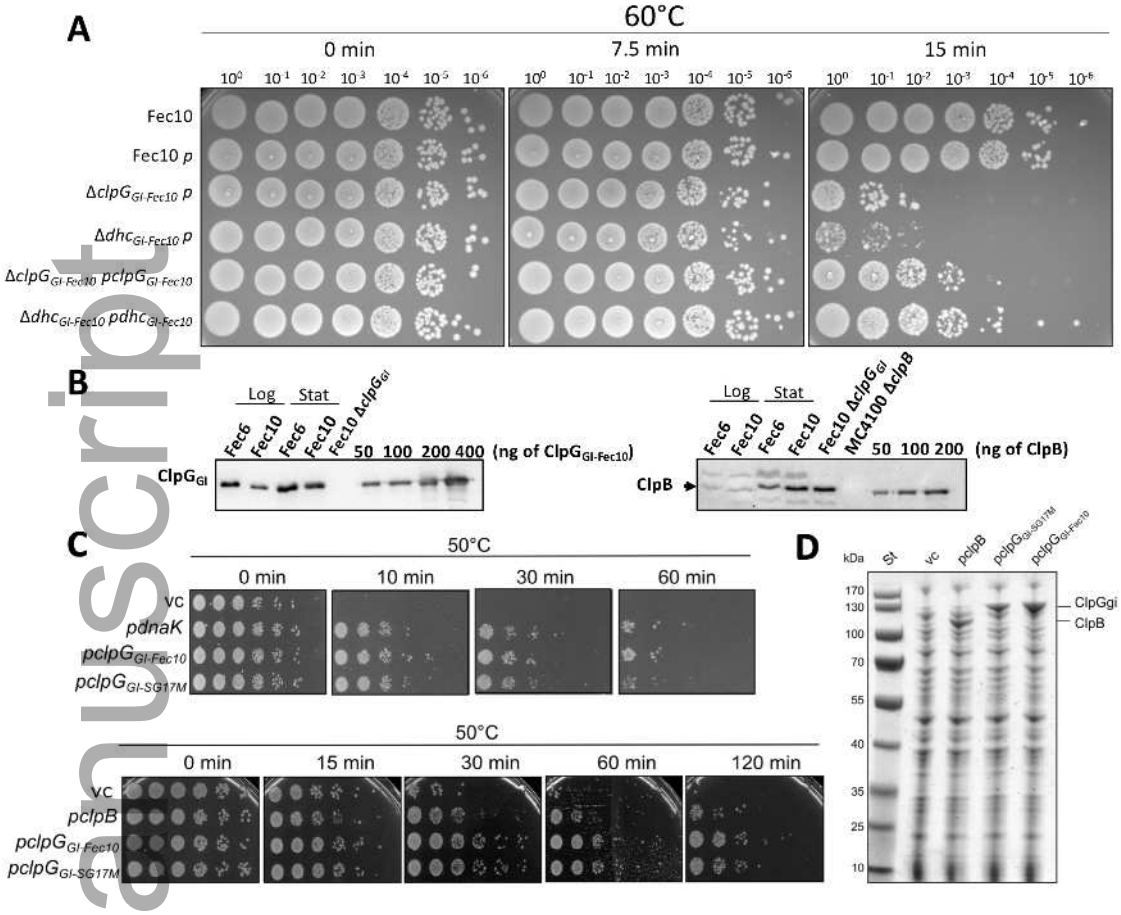
mmi_14614_f1.tif



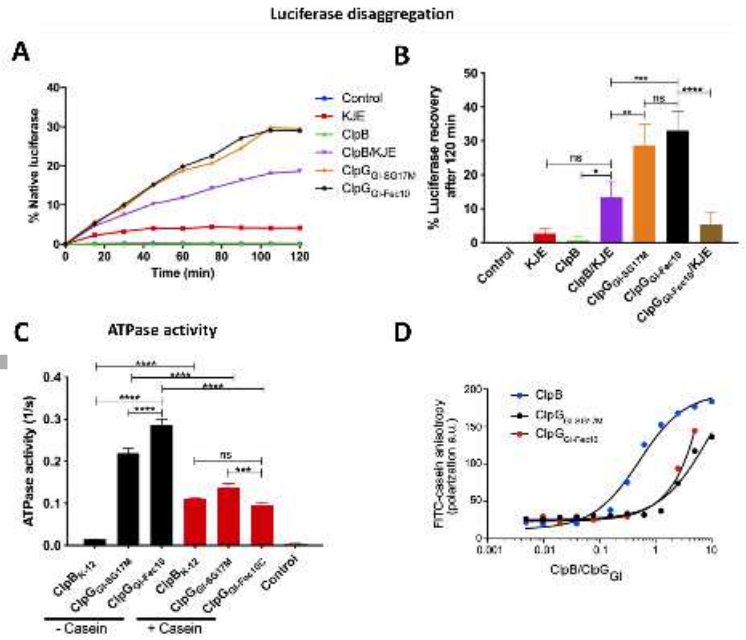
mmi_14614_f2.tif



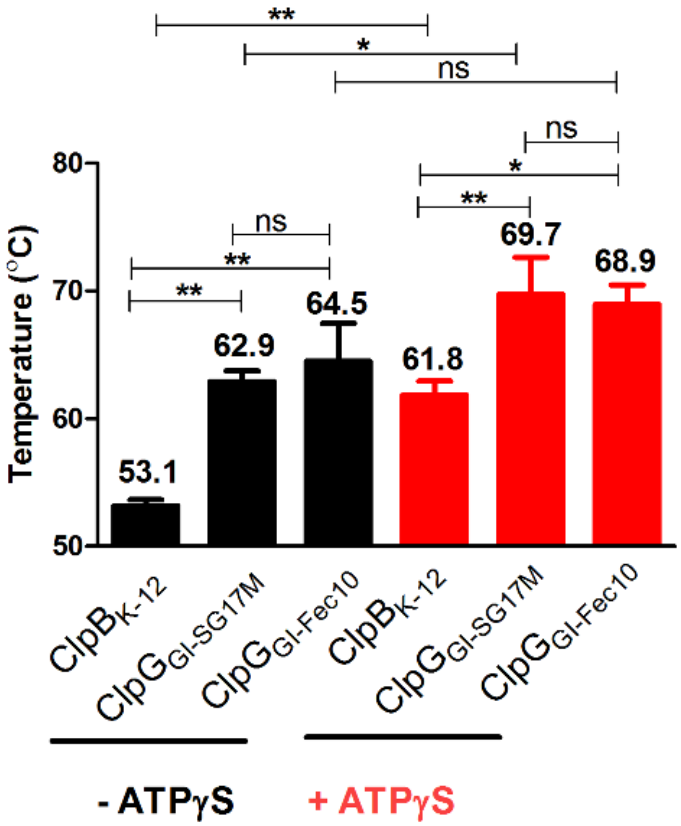
mmi_14614_f3.tif



mmi_14614_f4.tif



mmi_14614_f5.tif



mmi_14614_f6.tif