**Supplementary Information**

# 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

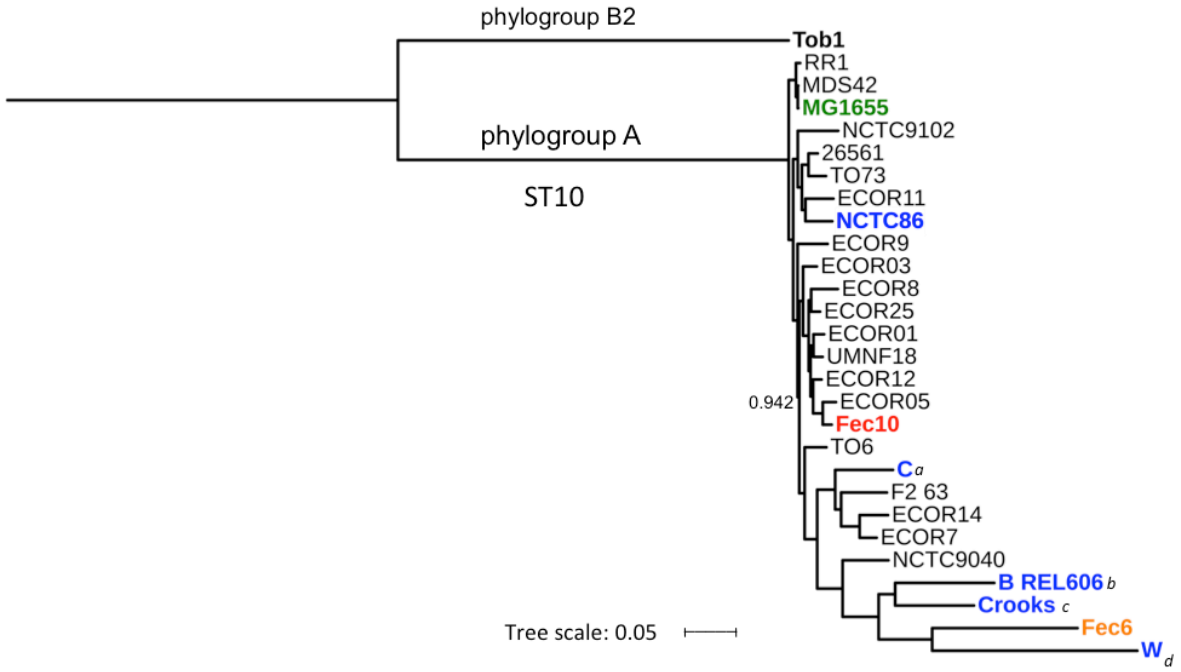
[7]Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124, Braunschweig, Germany

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

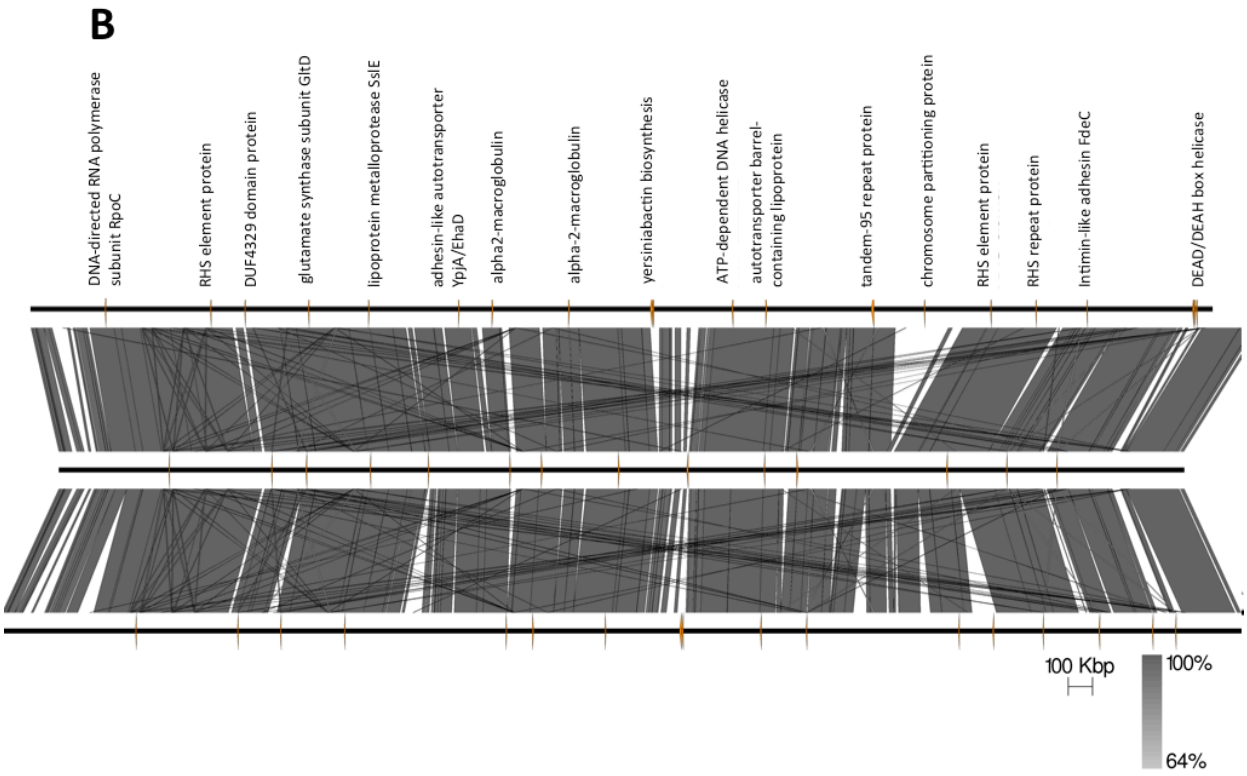
**Table of content - Supplementary Information**

## Supplementary Figures

**A**




**B**



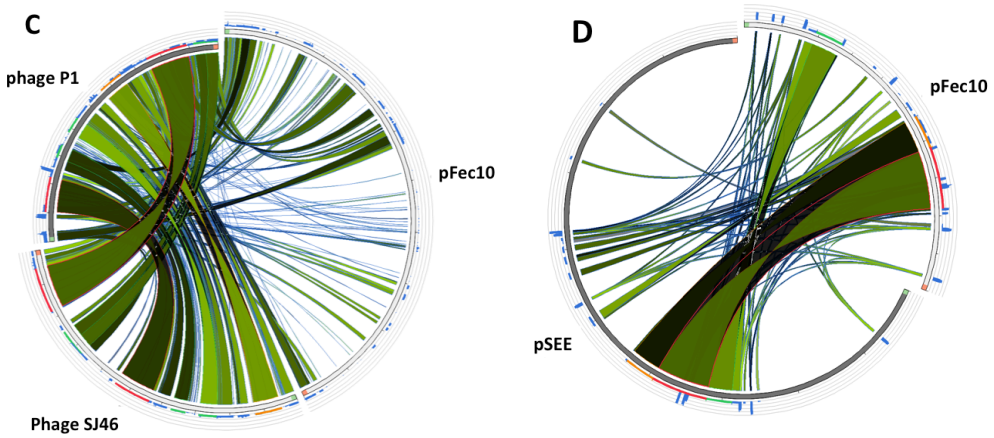
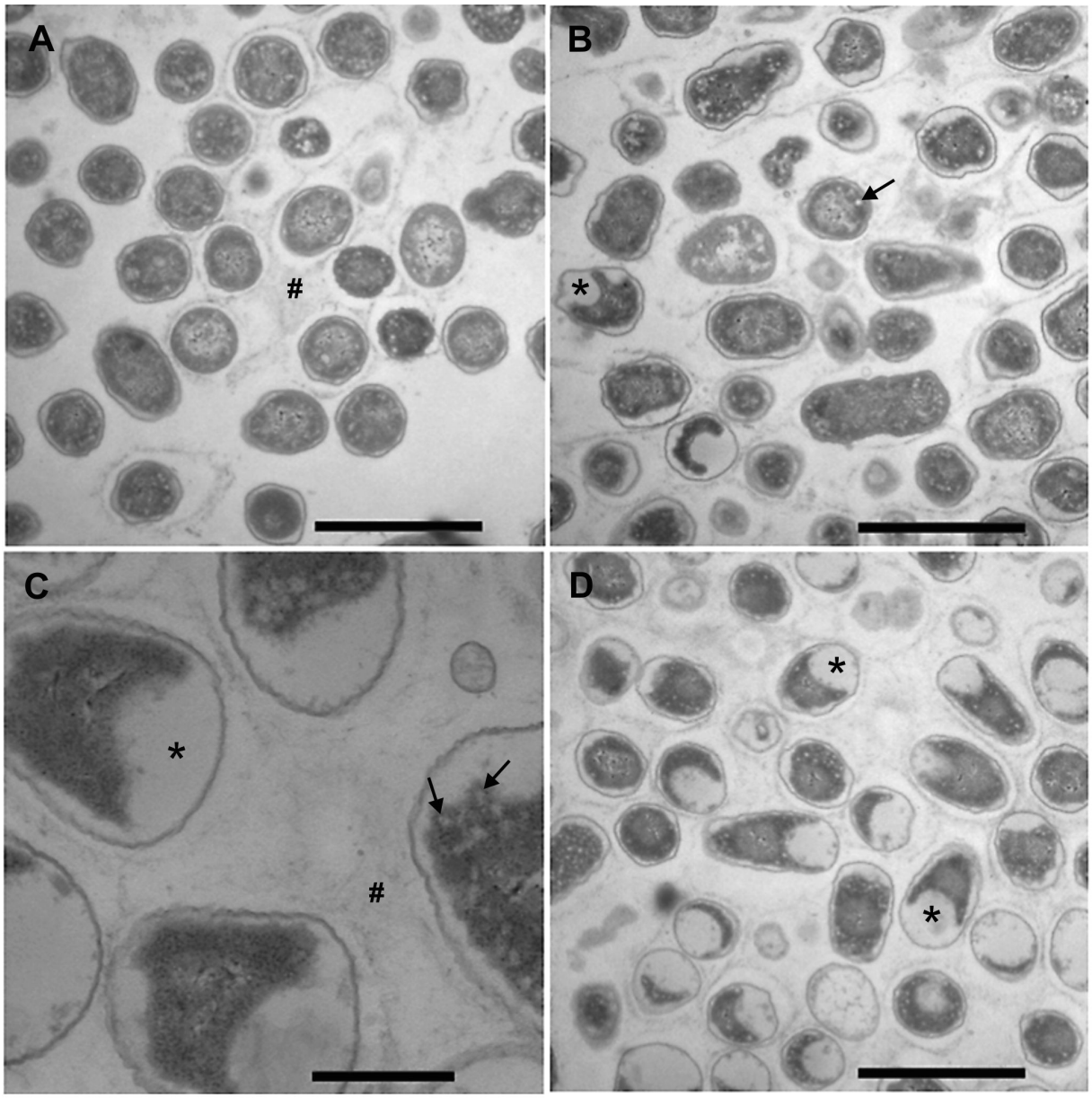



**Figure S1** Phylogenetic analysis, gene synteny and sequence homology of *E. coli* Fec10 and pFec10. A) Phylogenetic relationship of the genomes of selected closely related *E. coli* strains based on single nucleotide polymorphism (SNP) variant calling on 77.44% (3.269 Mbp) of the *E. coli* MG1655 core genome. In red=*E. coli* Fec10, sequence type ST10; orange=*E. coli* Fec6, sequence type ST399; blue=ancient *E. coli* reference strains. ECOR=*E. coli* strains from the *E. coli* reference collection. All strains are ST10, besides *a*=ST1721; *b*=ST93; *c*=ST3021; *d*=phylogroup B1, ST1079. Green=reference strain *E. coli* K-12 MG1655; outgroup *E. coli* Tob1 (phylogroup B2, ST2650). Bootstrap value=1, if not indicated otherwise. Tree visualized with iTOL https://itol.embl.de/ (Letunic and Bork 2016). B) Gene synteny of the *E. coli* Fec10 chromosome compared with *E. coli* K-12 MG1655 and *E. coli* NCTC86. Figure has been created with Easyfig (Sullivan, Petty and Beatson 2011). C) Homologous regions between pFec10 and most closely related phages P1 and SJ46. pFec10 has been normalized to the origin of replication. The graphical display of the nucleotide sequences aligned with tBLASTx using standard parameters was drawn with Circoletto (Darzentas 2010). D) Homologous regions between pFec10 and plasmid pSEE of *Salmonella enterica* subsp. *enterica* serovar Senftenberg ATCC43845. The graphical display of the nucleotide sequences aligned with tBLASTx using standard parameters was drawn with Circoletto (Darzentas 2010).


A
#
B
*
C
*
#
D
*
*

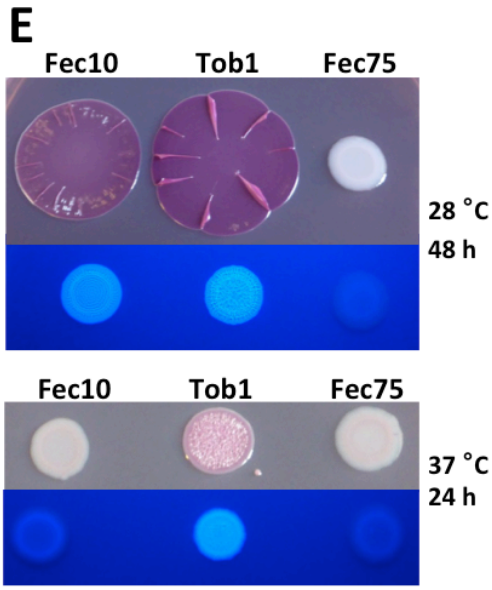



**Figure S2** Biofilm phenotype of *E. coli* Fec10. (A-D) Transmission electron microscopy of sections of Fec10 rdar morphotype colonies. (A, B) Arrangement and extracellular matrix of bacterial cells that grow at the colony rim (bars: 2 µm). (C) Detailed view of bacterial cells, which are situated less peripheral, near the colony-air interface (bar: 500 nm). (D) Bacterial cells, less peripheral, at the colony-agar interface (bar: 2 µm). Extracellular matrix (#) is present around the bacterial cells. Dilated periplasmic space (*) is mainly visible with cells of colony 'non-rim' areas. Arrows indicate electron translucent inclusions of roughly 60 nm in diameter that may represent polyhydroxybutyrate storage granules. (E) Rdar biofilm morphotype of *E. coli* Fec10 (Cimdins et al. 2017) compared to the semi-constitutive rdar morphotype of *E. coli* Tob1 and the smooth and white (saw) morphotype of *E. coli* Fec75 on Congo red and Calcoflour white agar plates at 28 °C and 37 °C.

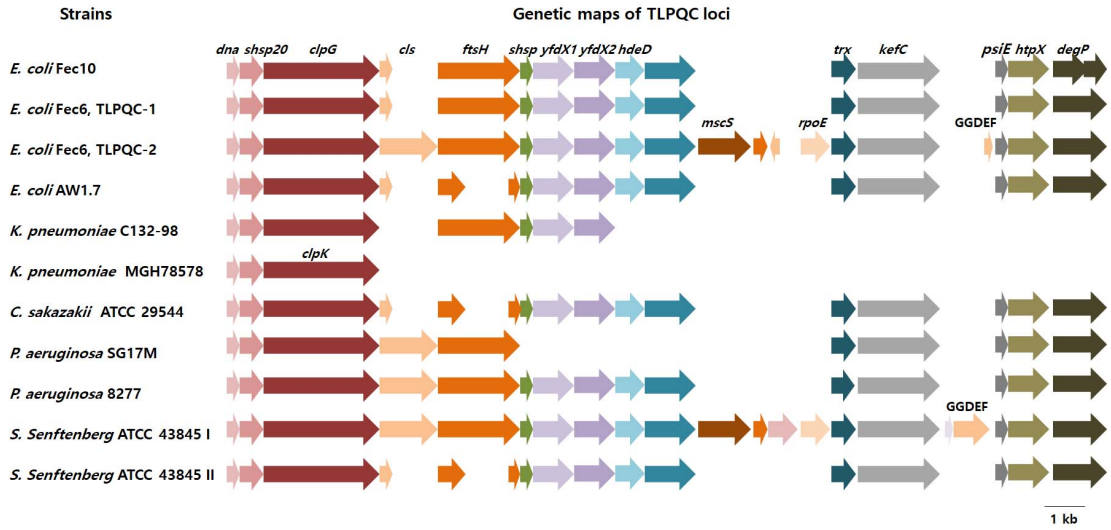



**Figure S3** Genetic map of TLPQC in *E. coli* Fec6 and Fec10 strains. TLPQC genetic maps from *E. coli* Fec10 and Fec6 strains are described in comparison with other strains from *E. coli* and other genera (Lee et al. 2016, Nguyen et al. 2017). Dna, MerR-like transcriptional regulator; sHsp20, small heat shock protein; ClpG/K, disaggregating chaperone; Cls, cardiolipid synthase; FtsH, metalloprotease; sHsp, small heat shock protein, YfdX1, 2, acid, osmotic stress responsive protein; HdeD, transmembrane protein involved in acid tolerance; MscS, small conductance mechanosensitive channel; RpoE, sigma E factor responding to heat shock and membrane stress; Trx, thioredoxin; KefC, glutathionine-dependent potassium-efflux system and methylglyoxal detoxification; GGDEF, diguanylate cyclase to synthesize cyclic di-GMP; PsiE, putative phosphate starvation-inducible protein; HtpX, inner-membrane associated peptidase; DegP, periplasmic protein with chaperone and protease activity.

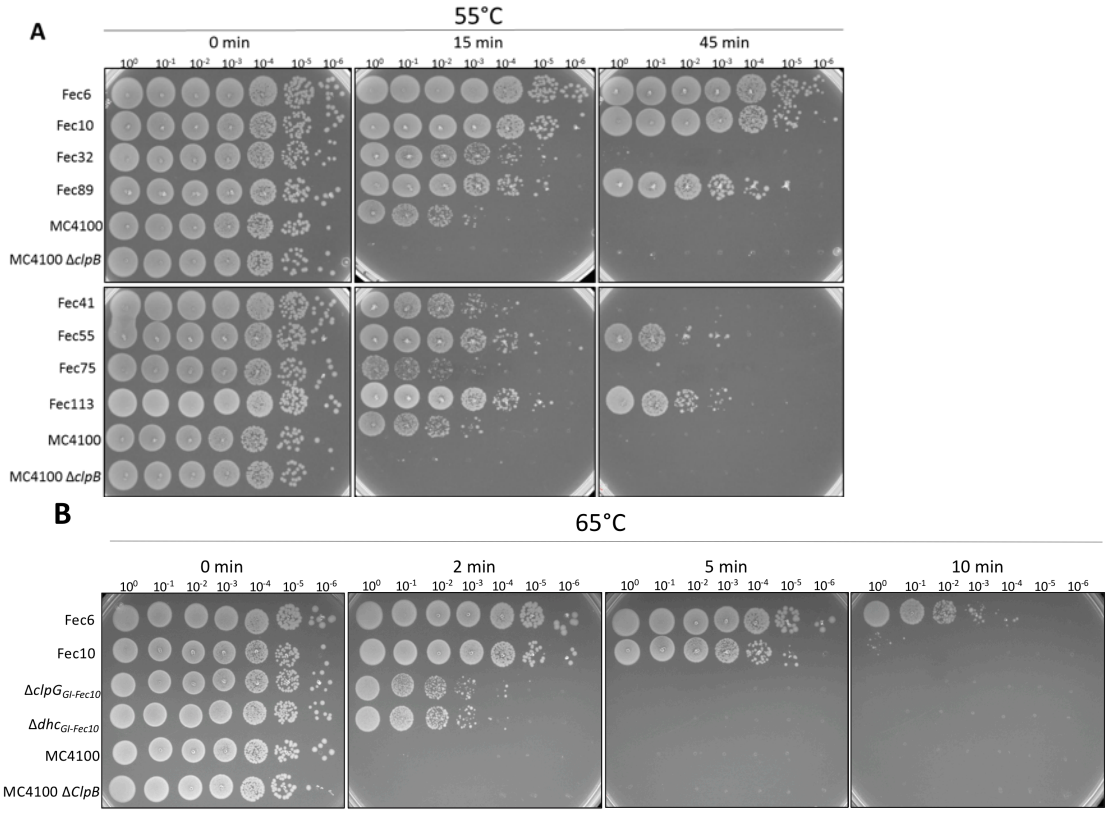



**Figure S4** Heat shock tolerance of *E. coli* fecal isolates and derivatives. (A) Screening of a panel of *E. coli* fecal isolates for heat shock tolerance at 55 °C. (B) Selected fecal isolates and respective *clpB* and $clpG_{GI}$ mutants were subjected to treatment with the extreme lethal heat of 65 °C. The elevated temperature was subjected for the indicated time and, subsequently, 10-fold serial dilutions were spotted on LB agar plates and incubated for 16 h. Laboratory strain MC4100 and mutant MC4100 Δ*clpB* were included as references.

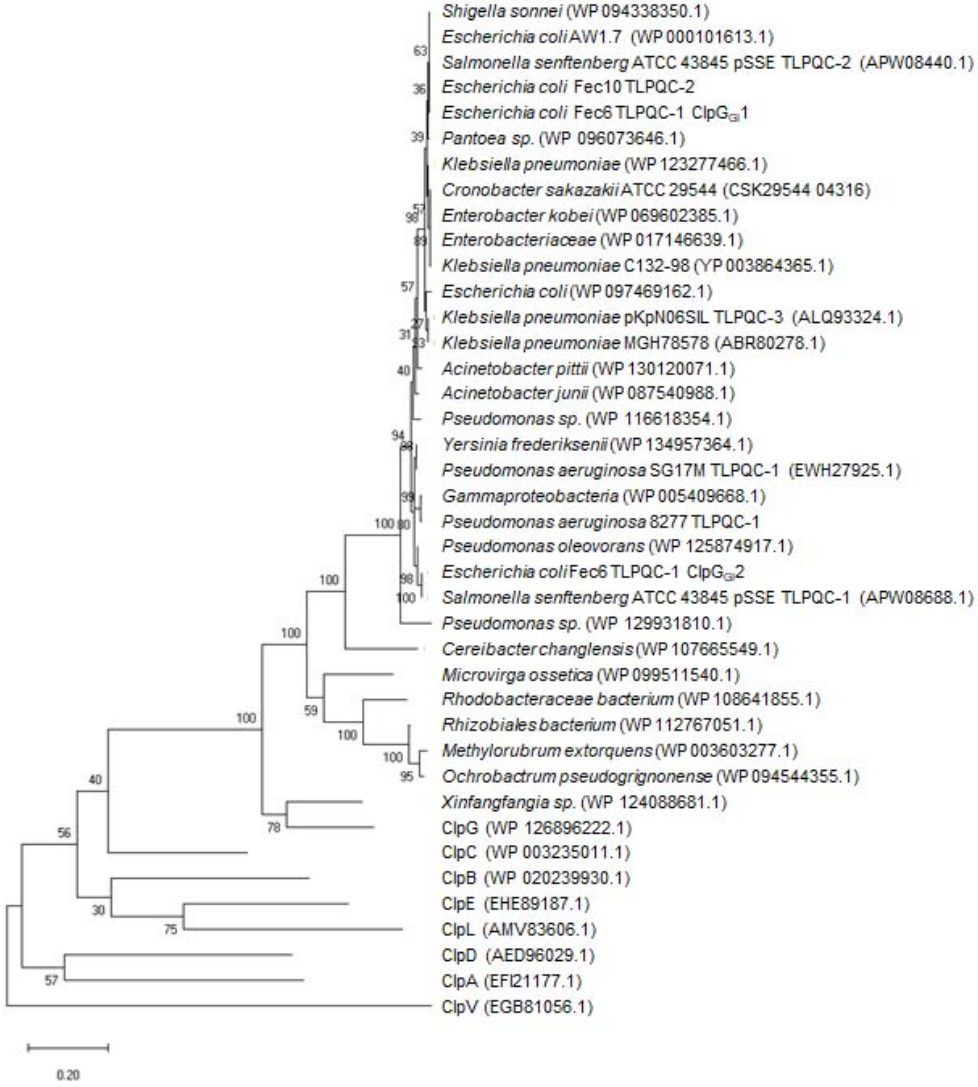
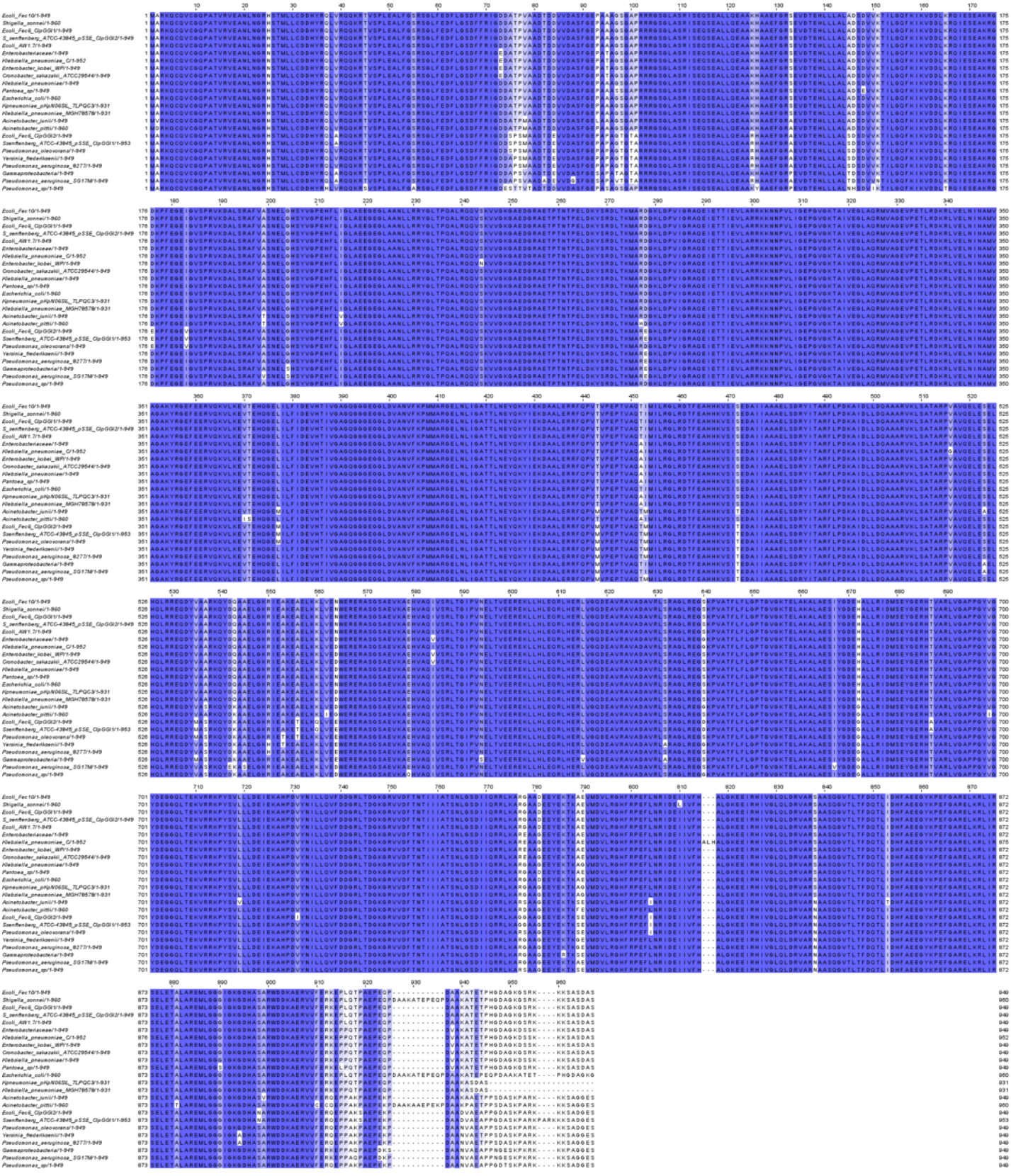



**Figure S5** Phylogenetic analysis of $ClpG_{GI}$ proteins. $ClpG_{GI}$ protein sequences from various strains of diverse genera were analyzed together with other class I Clp family members (ClpA, B, C, D, E, L and V). Protein sequences were aligned with ClustalW using standard parameters and a maximum-likelihood tree was built using MEGA X. Branch lengths correspond to substitutions per site and branch support values are indicated in %. The robustness of the phylogenetic tree topologies was evaluated by bootstrap analysis with 500 replications in MEGA X.

# A



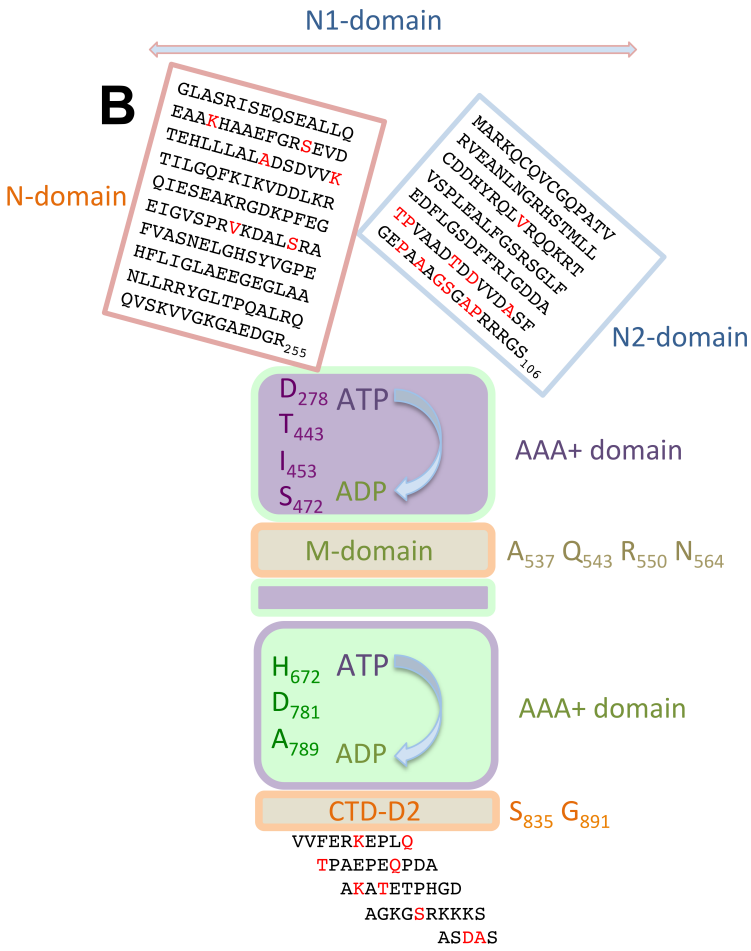



**Figure S6** Sequence variations within $ClpG_{GI}$ proteins. (A) Alignment of $ClpG_{GI}$ sequences. Residues on a dark blue background indicate 100% conservation and light blue background indicates 66% conservation. Sequences were aligned with ClustalW using default settings (Chenna et al. 2003) and displayed in JalView (Waterhouse et al. 2009). (B) Amino acid sequence variation in the N1-domain and ClpG specific C-terminal sequence between $ClpG_{GI\text{-}Fec10}$ and $ClpG_{GI\text{-}SG17M}$ indicated in red. Variant amino acids in other domains as indicated. Domain designation as in (Lee et al. 2018) with the N1 domain consisting of the ClpG-specific N2 domain and the ClpB N-terminal domain, the AAA+ ATPase domain, M (middle) domain and CTD-D2 (C-terminal domain D2 of ClpB).

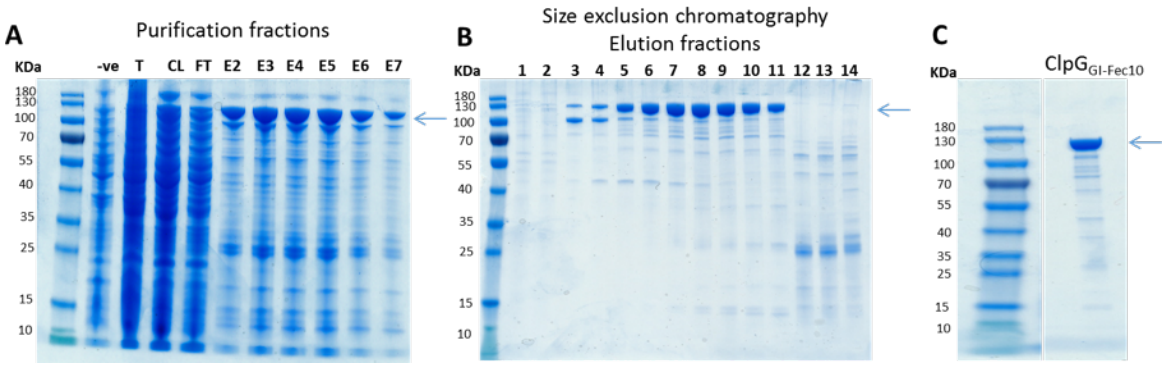



**Figure S7** Purification of $ClpG_{GI\text{-}Fec10}$. (A) SDS-PAGE gel showing the purification intermediates of $ClpG_{GI\text{-}Fec10}$, where –ve is total lysate of *E. coli* TOP10 pJN105 control, T is total lysate of *E. coli* TOP10 pJN105 *clpG*$_{GI\text{-}Fec10}$ expressing $ClpG_{GI\text{-}Fec10}$, CL is clear lysate, FT is flow through and E is elution. (B) SDS-PAGE showing elution volumes of $ClpG_{GI\text{-}Fec10}$ after Superdex 200 10/300 size exclusion chromatography. (C) SDS-PAGE gel showing $ClpG_{GI\text{-}Fec10}$ after consolidation from the size exclusion column and overnight dialysis.

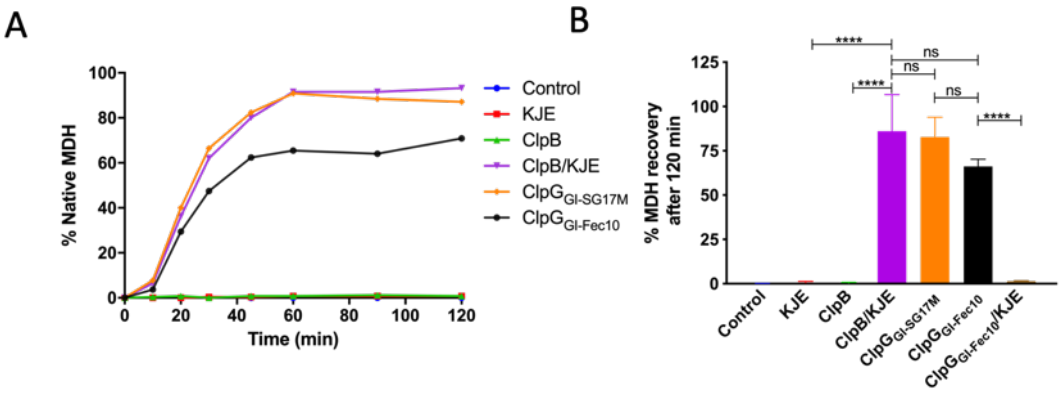



**Figure S8** $ClpG_{GI\text{-}Fec10}$ disaggregase restores activity of heat-treated MDH independent of accessory chaperons. (A) Refolding of aggregated MDH was monitored over 120 min. The activity of the native MDH was set to 100%. One representative experiment of three independent biological experiments with three technical replicates each is shown. (B) Recovery of MDH after 120 min incubation with the indicated disaggregase. The mean value was calculated from three independent experiments with three technical replicates each. Error bars indicate SD (****$P$<0.0001).

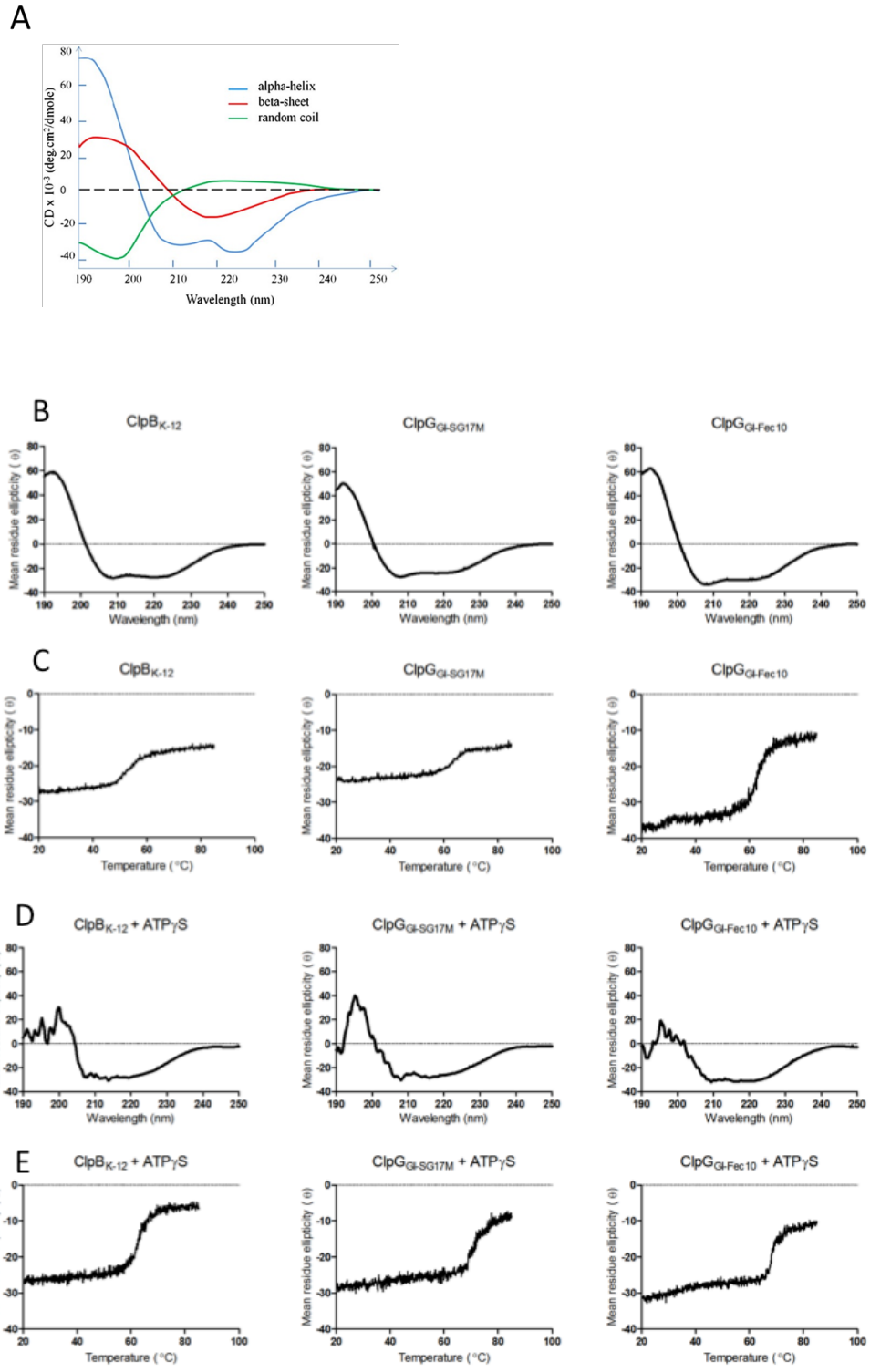



**Figure S9** Secondary structure analysis and melting temperature curves for $ClpB_{K\text{-}12}$, $ClpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$. (A) A reference signature spectrum for alpha and beta helices and random coils obtained from (Wei, Thyparambil and Latour 2014) with permission. Circular Dichroism Spectra (B, D) and melting temperature curves recorded at 222 nm (C, E) were recorded in the absence (B, C) and presence (D, E) of $Mg^{2+}$/ATPγS. The difference in curve

shape of $ClpG_{GI\text{-}Fec10}$ (C) is explained by a lower starting mean residue ellipticity value. This does not impact $T_m$ calculation.

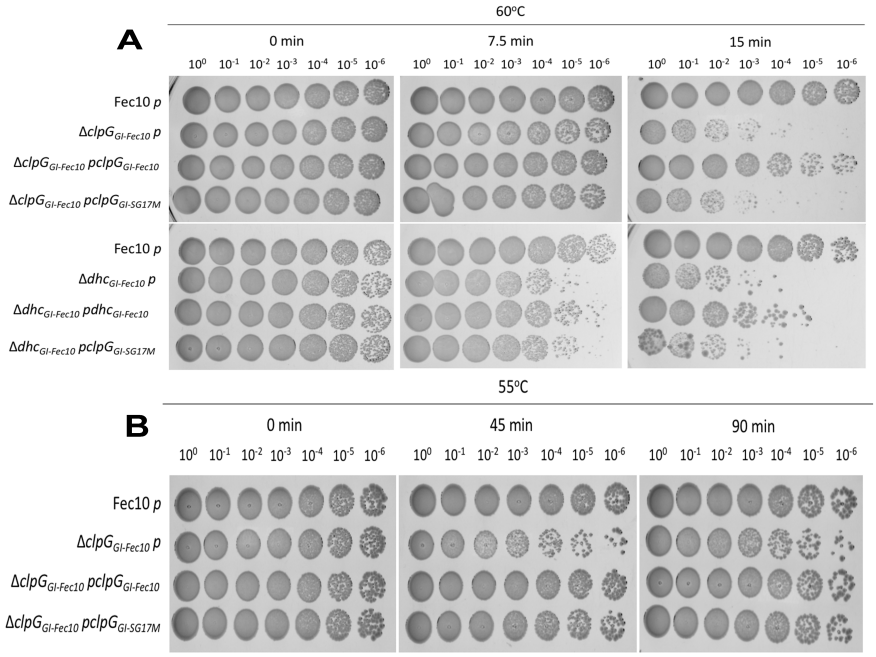



**Figure S10** Ability of overexpressed $ClpG_{GI\text{-}SG17M}$ and $ClpG_{GI\text{-}Fec10}$ to complement *clpG*$_{GI}$ and *dna-shsp20*$_{GI}$*-clpG*$_{GI}$ mutants of *E. coli* Fec10 upon exposure to lethal heat. (A) Assessment of heat shock tolerance of the *E. coli* Fec10 deletion mutant *clpG*$_{GI\text{-}Fec10}$ and *dhc*$_{GI\text{-}Fec10}$ exposed to 60 °C for 7.5 and 15 min upon complementation with *clpG*$_{GI\text{-}Fec10}$ and *clpG*$_{GI\text{-}SG17M}$. (B) Assessment of heat shock tolerance of the *E. coli* Fec10 deletion mutant *clpG*$_{GI\text{-}Fec10}$ exposed to 55 °C for 45 and 90 min upon complementation with *clpG*$_{GI\text{-}Fec10}$ and *clpG*$_{GI\text{-}SG17M}$. p=plasmid pBAD30; p*clpG*$_{GI\text{-}Fec10}$ = *clpG*$_{GI\text{-}Fec10}$ cloned in plasmid pBAD30; p*clpG*$_{GI\text{-}SG17M}$ = *clpG*$_{GI\text{-}SG17M}$ cloned in plasmid pBAD30.

## Supplementary Tables

**Table S1** Coding sequences on plasmid pFec10

| Locus tag | Type | Start | Stop | Protein ID | Protein function | general functionality |
|---|---|---|---|---|---|---|
| BFS32_023620 | CDS | 155792 | 792 | NVB22357.1 | RepFIB family plasmid replication initiator protein | phage proteins |
| BFS32_023625 | CDS | 1192 | 2397 | NVB22358.1 | ParA family protein | |
| BFS32_023630 | CDS | 2394 | 3359 | NVB22359.1 | ParB/RepB/Spo0J family partition protein | |
| BFS32_023635 PSEUDOGENE | CDS | 3615 | 4808 | | phage related protein | |
| | REPEAT | 4802 | 5576 | | repeat region | IS1A element 1 |
| BFS32_023640 | CDS | 5200 | 4823 | | IS1 family transposas protein InsB | |
| | CDS | 5520 | 5245 | | IS1 family transposase protein InsA | |
| BFS32_023645 | CDS | 5705 | 6034 | NVB22360.1 | holin | phage proteins |
| BFS32_023650 | CDS | 6031 | 6474 | NVB22361.1 | lysis protein LydB | |
| BFS32_023655 | CDS | 6461 | 7063 | NVB22362.1 | OdaE | |
| BFS32_023660 PSEUDOGENE | CDS | 7065 | 8087 | | internal head protein | |
| BFS32_023665 PSEUDOGENE | CDS | 8084 | 9481 | | helicase | |
| BFS32_023670 | CDS | 9515 | 9955 | NVB22363.1 | peptide-binding protein | |
| BFS32_023675 | CDS | 9952 | 10200 | NVB22364.1 | modulator protein | |
| BFS32_023680 | CDS | 10283 | 11263 | NVB22365.1 | DUF2971 domain-containing protein | |
| | REPEAT | 11258 | 12026 | | repeat region | IS1A element 2 |
| BFS32_023685 | CDS | 11314 | 11589 | | IS1 protein InsA | |
| | CDS | 11634 | 12011 | | IS1 protein InsB | |
| BFS32_023690 PSEUDOGENE | CDS | 12026 | 12340 | | sigma-70 family RNA polymerase sigma factor | cargo |
| BFS32_023695 | CDS | 13198 | 12392 | NVB22366.1 | encapsulating protein for peroxidase | cargo: detoxification |
| BFS32_023700 | CDS | 14255 | 13200 | NVB22367.1 | DyPc-type peroxidase | |
| BFS32_023705 | CDS | 15182 | 14349 | NVB22368.1 | NAD-dependent formate dehydrogenase | |
| | REPEAT | 15209 | 15980 | | repeat region | IS1S element 3 |
| BFS32_023710 | CDS | 15265 | 15540 | | IS1 protein InsA | |
| | | 15567 | 15962 | | IS1 protein InsB | |
| BFS32_023715 PSEUDOGENE | CDS | 15974 | 16357 | | methionine ABC transporter substrate-binding protein | cargo |
| **BFS32_023720** | **CDS** | **19517** | **16452** | **NVB22369.1** | **Tetrathionate reductase subunit TtrA** | |
| **BFS32_023725** | **CDS** | **20532** | **19510** | **NVB22370.1** | **Tetrathionate reductase subunit TtrC** | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **BFS32_023730** | **CDS** | **21285** | **20533** | **NVB22371.1** | **Tetrathionate reductase subunit TtrB** | **cargo: tetrathionate reduction** |
| **BFS32_023735** | **CDS** | **21495** | **23228** | **NVB22372.1** | **PhnD/SsuA/transferrin family substrate-binding protein TtrS** | |
| **BFS32_023740** | **CDS** | **23203** | **23787** | **NVB22373.1** | **two-component system response regulator TtrR** | |
| BFS32_023745 | CDS | 23880 | 24083 | NVB22374.1 | Fumarate hydratase FumD | cargo |
| | REPEAT | 24139 | 25010 | | repeat region | IS5 element 4 |
| BFS32_023750 | CDS | 24995 | 24192 | | Transposase InsH for insertion sequence element IS5 | |
| | CDS | 25172 | 25038 | | Transposase InsH for insertion sequence element IS5 | |
| BFS32_023755 | CDS | 25590 | 25213 | NVB22375.1 | Mat domain protein | phage protein |
| | REPEAT | 25335 | 26164 | | repeat region | repeat |
| BFS32_023760 | CDS | 25839 | 25624 | NVB22376.1 | Mat domain protein | phage proteins |
| BFS32_023765 | CDS | 26267 | 26028 | | Recombination enhancement function protein | |
| | REPEAT | 26430 | 26541 | | repeat region | transposases/repeat region |
| BFS32_023770 PSEUDOGENE | CDS | 26486 | 26560 | | hypothetical protein | |
| BFS32_023775 PSEUDOGENE | CDS | 26880 | 26533 | | transposase | |
| | REPEAT | 27069 | 27186 | | repeat region | |
| | REPEAT | 27519 | 27837 | | repeat region | |
| BFS32_023780 PSEUDOGENE | CDS | 27188 | 27502 | | transposase | |
| BFS32_023785 PSEUDOGENE | CDS | 27519 | 27827 | | DDE-type integrase/transposase/recombinase | |
| BFS32_023790 | CDS | 29810 | 28257 | NVB22378.1 | L-lactate permease LctP | cargo: lactate utilization |
| BFS32_023795 | CDS | 30246 | 31943 | NVB22379.1 | D-lactate dehydrogenase Dld | |
| BFS32_023800 | CDS | 32018 | 32737 | NVB22380.1 | predicted L-lactate dehydrogenase YkgE, (Fe-S)-binding protein | |
| BFS32_023805 | CDS | 32748 | 34175 | NVB22381.1 | predicted L-lactate dehydroogenase YkgF, (Fe-S) cluster-binding protein | |
| BFS32_023810 | CDS | 34168 | 34863 | NVB22382.1 | predicted L-lactate dehydrogenase YkgG, lactate utilization protein C | |
| BFS32_023815 PSEUDOGENE | CDS | 35293 | 34913 | | gluconate permease | cargo |
| BFS32_023820 | CDS | 35504 | 35815 | NVB22383.1 | type II toxin-antitoxin system HigB family toxin | |
| BFS32_023825 | CDS | 35812 | 36231 | NVB22384.1 | helix-turn-helix domain-containing protein | |

| | | | | | | |
|---|---|---|---|---|---|---|
| BFS32_023830 | CDS | 37470 | 36268 | NVB22385.1 | phage tail protein | phage proteins |
| BFS32_023835 | CDS | 37792 | 37463 | NVB22386.1 | baseplate protein | |
| BFS32_023840 | CDS | 38448 | 37789 | NVB22387.1 | maturation control protein Ref | |
| BFS32_023845 | CDS | 38872 | 39465 | NVB22388.1 | recombinase | |
| BFS32_023850 | CDS | 39781 | 40134 | NVB22389.1 | As(III)-sensing metalloregulatory transcriptional repressor ArsR | arsenite resistence |
| BFS32_023855 | CDS | 40182 | 40544 | NVB22390.1 | arsenite efflux transporter metallochaperone ArsD | |
| BFS32_023860 PSEUDOGENE | CDS | 40562 | 41185 | | arsenite efflux transporter ATPase subunit ArsA | |
| | REPEAT | 41185 | 41954 | | repeat region | IS1A element 5 |
| BFS32_023865 PSEUDOGENE | CDS | 41897 | 41200 | | IS1 family transposase | |
| BFS32_023870 PSEUDOGENE | CDS | 41953 | 42171 | | VRR-NUC domain-containing protein | phage proteins |
| BFS32_023875 | CDS | 42238 | 42393 | NVB22391.1 | norphogenetic protein | |
| BFS32_023880 | CDS | 42895 | 43521 | NVB22392.1 | norphogenetic protein | |
| BFS32_023885 | CDS | 43518 | 44195 | NVB22393.1 | metallophosphoesterase | |
| BFS32_023890 | CDS | 44192 | 44893 | NVB22394.1 | hypothetical protein | |
| BFS32_023895 | CDS | 45403 | 45909 | NVB22395.1 | 3'-phosphatase | |
| BFS32_023900 | CDS | 49290 | 46174 | NVB22396.1 | HsdR family type I site-specific deoxyribonuclease | cargo: restriction-modification |
| BFS32_023905 | CDS | 50647 | 49412 | NVB22397.1 | restriction endonuclease subunit S | |
| BFS32_023910 | CDS | 52200 | 50644 | NVB22398.1 | type I restriction-modification system subunit M | |
| BFS32_023915 | CDS | 52383 | 52604 | NVB22399.1 | type II toxin-antitoxin system Phd/YefM family antitoxin | |
| BFS32_023920 | CDS | 52604 | 52984 | NVB22400.1 | type II toxin-antitoxin system death-on-curing family toxin | |
| BFS32_023925 | CDS | 52989 | 53168 | NVB22401.1 | PdcA protein | phage proteins |
| BFS32_023930 | CDS | 53196 | 54239 | NVB22402.1 | DUF968 domain-containing protein | |
| BFS32_023935 | CDS | 54642 | 54959 | NVB22403.1 | Terminase A protein | |
| | REPEAT | 55057 | 56372 | | repeat region | IS1414 element 6 |
| BFS32_023940 | CDS | 56299 | 55091 | | IS256 family transposase | |
| BFS32_023945 | CDS | 57911 | 56683 | NVB22405.1 | IS3 family transposase, programmed frameshift | IS2 element 7 |
| BFS32_023950 | CDS | 57975 | 59102 | NVB22406.1 | terminase | |
| BFS32_023955 | CDS | 59324 | 59443 | NVB22407.1 | ash family protein | |
| BFS32_023960 | CDS | 59462 | 59683 | NVB22408.1 | host cell division inhibitor Icd-like protein | |

| | | | | | | |
|---|---|---|---|---|---|---|
| BFS32_023965 | CDS | 59680 | 60798 | NVB22409.1 | phage antirepressor Ant | phage proteins |
| BFS32_023970 | CDS | 61682 | 60831 | NVB22410.1 | hypothetical protein | |
| BFS32_023975 | CDS | 62002 | 61793 | NVB22411.1 | C1 repressor inactivator | |
| BFS32_023980 | CDS | 62607 | 62828 | NVB22412.1 | creatininase | |
| BFS32_023985 | CDS | 62836 | 63867 | NVB22413.1 | tyrosine-type recombinase/integrase | |
| BFS32_023990 | CDS | 63918 | 64229 | NVB22414.1 | lysogeny establishment protein | |
| BFS32_023995 | CDS | 64477 | 65037 | NVB22415.1 | Ref family protein | |
| | REPEAT | 64901 | 66924 | | repeat region | ISKpn26 element 8 |
| BFS32_024000 PSEUDOGENE | CDS | 65226 | 65729 | | maturation control protein | |
| BFS32_024005 | CDS | 65797 | 66777 | NVB22416.1 | IS5-like element ISKpn26 family transposase | |
| BFS32_024010 | CDS | 66942 | 67523 | NVB22417.1 | NAD(P)H:quinone oxidoreductase WrbA | cargo |
| | REPEAT | 66942 | 69191 | | repeat region | repeat |
| BFS32_024015 | CDS | 67544 | 67771 | NVB22418.1 | hypothetical protein | cargo |
| BFS32_024020 | CDS | 69050 | 67809 | NVB22419.1 | bifunctional glucose-1-phosphatase/inositol phosphatase Agp | |
| | REPEAT | 69192 | 70708 | | repeat region | IS30 element 9 |
| BFS32_024025 PSEUDOGENE | CDS | 69208 | 69465 | | transposase | |
| BFS32_024030 | CDS | 70629 | 69478 | NVB22420.1 | IS30-like element IS30 family transposase | |
| BFS32_024035 PSEUDOGENE | CDS | 70999 | 70697 | | transposase | transposases |
| BFS32_02404 | CDS | 72365 | 71343 | NVB22421.1 | IS110 family transposase | |
| BFS32_024045 | CDS | 75427 | 72389 | NVB22422.1 | Tn3 family transposase | |
| BFS32_024050 | CDS | 75595 | 76236 | NVB22423.1 | recombinase family protein | |
| BFS32_024055 | CDS | 77416 | 76493 | NVB22424.1 | cation transporter | cargo |
| BFS32_024060 | CDS | 78188 | 77616 | NVB22425.1 | cytochrome b/b6 domain-containing protein | |
| BFS32_024065 | CDS | 79902 | 78664 | NVB22426.1 | IS110 family transposase | ISEsa2 element 10 |
| BFS32_024070 | CDS | 81196 | 80201 | NVB22427.1 | IS110 family transposase | |
| BFS32_024075 PSEUDOGENE | CDS | 81427 | 81816 | | transposase | transposase |
| BFS32_024080 | CDS | 83192 | 82041 | NVB22428.1 | trypsin-like peptidase domain-containing protein DegP-GI | |
| BFS32_024085 | CDS | 84182 | 83217 | NVB22429.1 | M48 family metalloprotease HtpX-GI | |
| BFS32_024090 | CDS | 84657 | 84160 | NVB22430.1 | heat resistance protein PsiE-GI | |
| BFS32_024095 | CDS | 86369 | 84654 | NVB22431.1 | heat resistance system K+/H+ antiporter KefC-GI | |
| BFS32_024100 | CDS | 86813 | 86373 | NVB22432.1 | heat resistance system thioredoxin Trx-GI | |
| BFS32_024105 | CDS | 87948 | 86803 | NVB22433.1 | hypothetical protein | |

| | | | | | | |
|---|---|---|---|---|---|---|
| BFS32_024110 | CDS | 88639 | 88028 | NVB22434.1 | heat resistance membrane protein HdeD-GI | **cargo: TLPQC locus** |
| BFS32_024115 | CDS | 89623 | 88736 | NVB22435.1 | heat resistance protein YfdX2 | |
| BFS32_024120 | CDS | 90640 | 89726 | NVB22436.1 | heat resistance protein YfdX1 | |
| BFS32_024125 | CDS | 91121 | 90663 | NVB22437.1 | small heat shock protein sHSP20-GI | |
| BFS32_024130 | CDS | 91349 | 91209 | NVB22438.1 | hypothetical protein | |
| BFS32_024135 PSEUDOGENE | CDS | 92035 | 91300 | | ATP-dependent metallopeptidase FtsH/Yme1/Tma family protein_FtsH-GI | |
| BFS32_024140 | CDS | 92287 | 92096 | NVB22439.1 | hypothetical protein | |
| **BFS32_024145** | **CDS** | **95136** | **92287** | **NVB22440.1** | **heat shock survival AAA family ATPase ClpG-GI** | |
| **BFS32_024150** | **CDS** | **95811** | **95242** | **NVB22441.1** | **small heat shock protein sHSP20-GI** | |
| **BFS32_024155** | **CDS** | **96127** | **95846** | **NVB22442.1** | **helix-turn-helix domain-containing protein Dna** | |
| BFS32_024160 PSEUDOGENE | CDS | 96181 | 96092 | _ | DNA-binding protein | |
| BFS32_024165 | CDS | 96584 | 96321 | NVB22443.1 | hypothetical protein | |
| | REPEAT | 96712 | 96829 | | repeat region | repeat |
| BFS32_024170 PSEUDOGENE | CDS | 96744 | 96980 | | transposase | IS26 element 11 |
| BFS32_024175 | CDS | 97736 | 97032 | NVB22444.1 | IS6-like element IS26 family transposase | |
| BFS32_024180 PSEUDOGENE | CDS | 97799 | 98343 | | IS5 family transposase | ISKpn26 element 12 |
| | REPEAT | 97800 | 98494 | | repeat region | |
| BFS32_024185 | CDS | 98620 | 98907 | NVB22445.1 | hypothetical protein | cargo |
| BFS32_024190 | CDS | 98907 | 99164 | NVB22446.1 | DNA polymerase III subunit theta | |
| BFS32_024195 | CDS | 99205 | 102144 | NVB22447.1 | AAA family ATPase | |
| BFS32_024200 | CDS | 102141 | 103565 | NVB22448.1 | tetratricopeptide repeat protein | |
| | REPEAT | 103561 | 103748 | | repeat region | transposases |
| BFS32_024205 PSEUDOGENE | CDS | 103747 | 103571 | | IS3 family transposase | |
| BFS32_0242010 PSEUDOGENE | CDS | 104058 | 103801 | | DDE-type integrase/transposase/recombinase | |
| | REPEAT | 104040 | 106476 | | repeat region | ISEc78 element 13 |
| BFS32_024215 | CDS | 104123 | 104506 | NVB22449.1 | IS66 family insertion sequence hypothetical protein | |
| BFS32_024220 | CDS | 104503 | 104850 | NVB22450.1 | IS66 family insertion sequence element accessory protein TnpB | |

| | | | | | | |
|---|---|---|---|---|---|---|
| BFS32_024225 | CDS | 104900 | 106435 | NVB22451.1 | IS66 family transposase | |
| BFS32_024230 PSEUDOGENE | CDS | 106950 | 106477 | | IS6-like element IS26 family transposase | IS15 element 14 |
| BFS32_024235 | CDS | 107185 | 107583 | NVB22452.1 | ester cyclase | |
| BFS32_024240 | CDS | 107603 | 108556 | NVB22453.1 | helix-turn-helix domain-containing protein | cargo |
| BFS32_024245 | CDS | 108627 | 109403 | NVB22454.1 | MBL fold metallo-hydrolase | |
| | REPEAT | 109471 | 110473 | | repeat region | |
| BFS32_024250 PSEUDOGENE | CDS | 109543 | 109701 | | IS5/IS1182 family transposase | IS1A element 15 |
| BFS32_024255 | CDS | 110095 | 109718 | | IS1 protein InsB | |
| | CDS | 110415 | 110140 | | IS1 protein InsA | |
| | REPEAT | 110569 | 112176 | | repeat region | repeat |
| BFS32_024260 | CDS | 111519 | 110569 | NVB22455.1 | magnesium/cobalt transporter CorA | cargo |
| BFS32_024265 | CDS | 111889 | 112203 | NVB22456.1 | hypothetical protein | |
| | REPEAT | 112176 | 113925 | | repeat region | |
| BFS32_024270 | CDS | 112232 | 112507 | | IS1 protein InsA | IS1A element 16 |
| | CDS | 112552 | 112929 | | IS1 protein InsB | |
| BFS32_024275 PSEUDOGENE | CDS | 112950 | 113774 | | IS5-like element ISKpn26 family transposase | ISKpn26 element 17 |
| BFS32_024280 | CDS | 113984 | 115981 | NVB22457.1 | choline BCCT transporter BetT | cargo |
| BFS32_024285 | CDS | 116045 | 117322 | NVB22458.1 | DUF2254 domain-containing protein | |
| BFS32_024290 PSEUDOGENE | CDS | 117579 | 117764 | | Tn3 family transposase | transposases |
| BFS32_024295 PSEUDOGENE | CDS | 118045 | 117764 | | IS6-like element IS26 family transposase | |
| BFS32_024300 PSEUDOGENE | CDS | 118109 | 118363 | | helix-turn-helix domain-containing protein | cargo |
| BFS32_024305 | CDS | 118372 | 118776 | NVB22459.1 | arsenical pump-driving ATPase ArsA | |
| BFS32_024310 | CDS | 119232 | 118807 | NVB22460.1 | glutaredoxin-dependent arsenate reductase ArsC | arsenite resistence |
| BFS32_024315 | CDS | 120534 | 119245 | NVB22461.1 | arsenite efflux transporter membrane subunit ArsF | |
| BFS32_024320 PSEUDOGENE | CDS | 121719 | 120583 | | arsenite efflux transporter ATPase subunit ArsA | |
| | REPEAT | 121719 | 122488 | | repeat region | IS1A |
| BFS32_024325 PSEUDOGENE | CDS | 121775 | 122472 | | IS1 family transposase | element 18 |
| BFS32_024330 | CDS | 122905 | 122720 | NVB22462.1 | hypothetical protein | phage protein X (fragment) |
| | REPEAT | 123004 | 123220 | | repeat region | repeat |
| | REPEAT | 123329 | 124100 | | repeat region | IS1X2 element 19 |
| BFS32_024335 PSEUDOGENE | CDS | 124043 | 123346 | | IS1 family transposase | |

| | | | | | | |
|---|---|---|---|---|---|---|
| BFS32_024340 | CDS | 125118 | 124606 | NVB22463.1 | hypothetical protein | |
| BFS32_024345 | CDS | 125651 | 125121 | NVB22464.1 | hypothetical protein | |
| BFS32_024350 | CDS | 126679 | 127164 | NVB22465.1 | plasmid transfer protein | cargo |
| BFS32_024355 | CDS | 127277 | 127882 | NVB22466.1 | hypothetical protein | |
| BFS32_024360 PSEUDOGENE | CDS | 127893 | 129359 | | TraI domain-containing protein | |
| | REPEAT | 129587 | 130786 | | repeat region | ISKpn26 |
| BFS32_024365 | CDS | 129655 | 130635 | NVB22467.1 | IS5-like element ISKpn26 family transposase | element 20 |
| BFS32_024370 PSEUDOGENE | CDS | 131074 | 130784 | | Tn3 family transposase | transposase |
| BFS32_024375 | CDS | 133585 | 131453 | NVB22468.1 | HlyD family efflux transporter periplasmic adaptor subunit | |
| BFS32_024380 | CDS | 134936 | 133587 | NVB22469.1 | HlyD family efflux transporter periplasmic adaptor subunit | cargo: multi-drug efflux |
| BFS32_024385 | CDS | 135814 | 134933 | NVB22470.1 | efflux RND transporter periplasmic adaptor subunit | |
| BFS32_024390 | CDS | 137748 | 135829 | NVB22471.1 | TolC family protein | |
| BFS32_024395 | CDS | 149410 | 137825 | NVB22472.1 | tandem-95 repeat protein | phage protein |
| | REPEAT | 149859 | 152308 | | repeat region | |
| BFS32_024400 | CDS | 149932 | 150912 | NVB22473.1 | IS5-like element ISKpn26 family transposase | IS30 element 21 |
| BFS32_024405 PSEUDOGENE | CDS | 151967 | 151061 | | IS30-like element IS30 family transposase | |
| BFS32_024410 PSEUDOGENE | CDS | 152292 | 152035 | | transposase | transposase |
| BFS32_024415 PSEUDOGENE | CDS | 152915 | 152361 | | recombination-associated protein RdgC | |
| BFS32_024420 | CDS | 153192 | 152908 | NVB22474.1 | hypothetical protein | |
| BFS32_024425 | CDS | 153654 | 154442 | NVB22475.1 | hypothetical protein | phage proteins |
| BFS32_024430 | CDS | 154482 | 154904 | NVB22476.1 | endopeptidase PpfA | |
| BFS32_024435 | CDS | 155082 | 155474 | NVB22477.1 | hypothetical protein | |

Blue underlined=phage SJ46 derived genes; grey=IS elements as identified with ISfinder; light grey=other transposases/recombinases; dark orange=cargo genes mentioned in the text; light orange=other cargo genes; green=analyzed cargo genes; plum=phage P1 genes; dark plum=phage P1 genes mentioned in text.

**Table S2** Identified IS elements on pFec10

| IS-element | | % identity | alignment length | bp start | bp stop | bit score |
|---|---|---|---|---|---|---|
| 1 | IS1A | 99.4 | 768 | 4808 | 5575 | 1483 |
| 2 | IS1A | 99.0 | 768 | 11259 | 12026 | 1459 |
| 3 | IS1S | 98.6 | 768 | 15210 | 15977 | 1435 |
| 4 | IS5 | 99.4 | 1195 | 24046 | 25240 | 2313 |
| 5 | IS1A | 97.1 | 768 | 41185 | 41952 | 1348 |
| 6 | IS1414 | 98.4 | 1315 | 55057 | 56371 | 2432 |
| 7 | IS2 | 97.7 | 1331 | 56672 | 58002 | 2393 |
| 8 | ISKpn26 | 99.4 | 1196 | 65729 | 66924 | 2315 |
| 9 | IS30 | 99.8 | 1221 | 69471 | 70691 | 2397 |
| 10 | ISEsa2 | 99.8 | 1493 | 78636 | 80128 | 2936 |
| 11 | IS26 | 99.8 | 820 | 96980 | 97799 | 1610 |
| 12 | ISKpn26 | 98.6 | 692 | 97800 | 98490 | 1285 |
| 13 | ISEc78 | 99.2 | 2437 | 104040 | 106476 | 4680 |
| 14 | IS15 | 99.8 | 538 | 106476 | 107013 | 1059 |
| 15 | IS1A | 99.4 | 768 | 109703 | 110470 | 1483 |
| 16 | IS1A | 99.9 | 768 | 112177 | 112944 | 1515 |
| 17 | ISKpn26 | 99.7 | 977 | 112945 | 113921 | 1913 |
| 18 | IS1A | 97.1 | 768 | 121720 | 122487 | 1348 |
| 19 | IS1X2 | 97.0 | 768 | 123331 | 124098 | 1340 |
| 20 | ISKpn26 | 99.6 | 1196 | 129587 | 130782 | 2331 |
| 21 | IS30 | 99.7 | 970 | 151060 | 152029 | 1891 |

**Table S3** Primers used in this study

| Primer name | Sequence† (5'-3') | Purpose | Source |
|---|---|---|---|
| del_*clpG*$_{GI}$_Fec10_F | TCGTCGCCTCAACCTTTTCAACCTTCTAAGGAGCAT | construction of Δ*clpG*$_{GI}$ in Fec10. | This study |
| | CAGC**GTGTAGGCTGGAGCTGCTTC** | forward primer composed of 40 bp upstream of *clpG*$_{GI}$ start codon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold) | |
| del_*dhc*$_{GI}$_Fec10_F | TTGGCCGGAACCGCAAAGCGACGCGATTGGAGAGGG | construction of Δ*dhc*$_{GI}$ in Fec10. | This study |
| | TGAG**GTGTAGGCTGGAGCTGCTTC** | forward primer composed of 40 bp upstream of the *dhc*$_{GI}$ start codon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold). | |
| del_ *clpG*$_{GI}$/*dhc*$_{GI}$ _Fec10_R | GCCAGGCCCTTGGGGTCGGACACGGCAGCCTCGCCC | construction of Δ*clpG*$_{GI}$ and Δ*dhc*$_{GI}$ in Fec10. | This study |
| | GCCA**CATATGAATATCCTCCTTAGT** | reverse primer composed of 40 bp downstream of the *clpG*$_{GI}$ or *dhc*$_{GI}$ stop codon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold). | |
| Check_del_ *clpG*$_{GI}$ _Fec10_F | ACGCTCGATCCCGATCAACGGCT | forward control primer for Δ*clpG*$_{GI}$ in Fec10 | This study |
| Check_del_ *dhc*$_{GI}$ _Fec10_F | CTGTCCATTGCCTGAGCAGTT | forward control primer for Δ*dhc*$_{GI}$ in Fec10 | This study |
| Check_del_ *clpG*$_{GI}$/*dhc*$_{GI}$ _Fec10_R | AGTTGCCAGAGGATCAGGTGAA | reverse control primer for Δ*clpG*$_{GI}$ and Δ*dhc*$_{GI}$ in Fec10 | This study |
| Lambda_ttr_sensor-F | ACCCGGCTTGTTGTTGATCTAAAACAACATAA | construction of Δsensor in Fec10. | This study |
| | TTGACAGGGACAATTGA**GTGTAGGCTGGAG CTGCTTC** | forward primer composed of 49 bp upstream of *ttrSR* start codon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold) | |
| Lambda_ttr_sensor-R | CTCTGAGGATTTAGACAAAATTCAATGAAAA TCAAATAAATGTTATCATT**CATATGAATATC CTCCTTAG** | construction of Δsensor in Fec10. reverse primer composed of 50 bp downstream of *ttrSR* operon stop codon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold) | This study |
| Lambda_ttr-F | TCAATTGTCCCTGTCAATTATGTTGTTTTAGA TCAACAACAAGCCGGGT**GTGTAGGCTGGAG CTGCTTC** | construction of Δttr in Fec10 forward primer composed of 49 bp upstream of *small ORF-ttrBCA* operon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold) | This study |
| Lambda_ttr-R | TGTCCTGTAGGCCGGGTAAAACGCGTTCGCG ATGCCATGCGGCACCAGAG**CATATGAATATC CTCCTTAG** | construction of Δttr in Fec10 reverse primer composed of 49 bp downstram of *small ORF-ttrBCA* operon and complementary sequence to amplify the chloramphenicol cassette from pKD3 (in bold) | This study |
| Check_*ttrSR*-F | TGATTTACAGTGTAAAAATCGTGACG | forward control primer for Δ*ttrSR* in Fec10 | This study |
| Check_*ttrSR*-R | CTCGTCGTCTTTCGTCTGAT | reverse control primer for Δ*ttrSR* in Fec10 | This study |
| Check_*ttrBCA*-F | ACTTAGCATCCACAGCGACG | forward control primer for Δ*ttrBCA* in Fec10 | This study |
| Check_*ttrBCA*-R | ACTGGATCCGTACTAAATATAAAGGTG | reverse control primer for Δ*ttrBCA* in Fec10 | This study |
| *clpG*$_{GI}$ _Fec10_NheI_pJN_F | CGCGCTAGCACCTTCTAAGGAGCATCAGCATG | forward cloning primer for *clpG*$_{GI\text{-}Fec10}$ with 6xHis-tag at the C-terminus in pJN105 for purification | This study |
| *clpG*$_{GI}$ _Fec10_XbaI_pJN_His_R | GCGTCTAGATCAGTGATGATGATGATGATGAGATGC GTCGCTCGCCGACTT | reverse cloning primer for *clpG*$_{GI\text{-}Fec10}$ with 6xHis-tag at the C-terminus in pJN105 for purification | This study |
| *clpG*$_{GI}$_Fec10_EcoRI_pBAD30_F | GCGCGGAATTCACCTTCTAAGGAGCATCAGCATG | forward cloning primer for *clpG*$_{GI\text{-}Fec10}$ with 6xHis-tag at the C-terminus in pBAD30 for complementation | This study |
| *dhc*$_{GI}$_Fec10_EcoRI_pBAD30_F | GCGCGGAATTCGGTCTAATATTACAAAGTATTGCAA GG | forward cloning primer for *dhc*$_{GI}$ with 6xHis-tag at the C-terminus in pBAD30 for complementation | This study |
| *clpG*$_{GI}$/*dhc*$_{GI}$_Fec10_SalI_pBAD30_His_R | GCGGTCGACTCAGTGATGATGATG ATGATGAGATGCGTCGCTCGCCGACTT | reverse cloning primer for *clpG*$_{GI\text{-}Fec10}$ and *dhc*$_{GI}$ with 6xHis-tag at the C-terminus in pBAD30 for complementation | This study |
| *dh*_Fec10_W1 | AGCAGGCAAAGGGCTTGAGT | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |
| *dh*_Fec10_W2 | AACAGGAGAAGAAGGAAGGTG | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |
| *clpG*$_{GI}$ _Fec10_W1 | GATCAAGGTCGATGACCTCAAG | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |
| *clpG*$_{GI}$ _Fec10_W2 | TGTGGGCAAGACCGCCAT | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |

| | | | |
|---|---|---|---|
| *clpG*$_{GI}$ _Fec10_W3 | ACAAGGTCAGCATCTCCGAGG | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |
| *clpG*$_{GI}$ _Fec10_W4 | AAGGCAGCAAACCGGTGGCTA | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |
| *clpG*$_{GI}$ _Fec10_W5 | GTGACGCTCACTTTCGATCAGA | confirmatory primer for the ORFs of *dhc*$_{GI}$ | This study |
| pJN105_pBAD F | CCATAGCATTTTTATCCATAAG | forward confirmatory primer for cloning into pJN105 vector | This study |
| pJN105_MCS R | AAACGACGGCCAGTGAGC | reverse confirmatory primer for cloning into pJN105 vector | This study |
| pBAD30_MCS_F | GTCTATAATCACGGCAGAAAAGTCCAC | forward confirmatory primer for cloning into pBAD30 vector | This study |
| pBAD30_MCS_R | CTGTTTTATCAGACCGCTTCTGC | reverse confirmatory primer for cloning into pBAD30 vector | This study |
| FEC10_PLASMID CHECK_F | GCTACAGCCTCAATGCGGCTTTG | confirmatory primer for pFec10 | This study |
| FEC10_PLASMID CHECK_R | GTTCAGGGCTGATGTCTGTTAG | confirmatory primer for pFec10 | This study |
| FEC10_PLASMID_WALK1_F | CAATACCTTTGATGGTGGCG | confirmatory primer for pFec10 | This study |
| FEC10_PLASMID_WALK2_F | GATGTCACGCTGAAAATGCC | confirmatory primer for pFec10 | This study |
| ClpGgi_Fec10_BamHI | GGCCATGGATCCATGGCCAGAAAACAATGCCAGGTCTGCGG | forward cloning primer for *clpG*$_{GI}$ in pUHE21 for complementation | This study |
| ClpGgi_Fec10_XbaI | GGCCATTCTAGATCAAGATGCGTCGCTCGCCGACTTCTTCT | reverse cloning primer for *clpG*$_{GI}$ in pUHE21 for complementation | This study |
| paClpGgi_EcoRI_pBAD30_F | GCGCGGAATTCCTTTACAGGAGCATCAGCATG | forward cloning primer for *clpG*$_{GI\text{-}SG17M}$ with 6xHis-tag at the C-terminus in pBAD30 for complementation | This study |
| PaClpGgi_SalI_pBAD30_His_R | GCGGTCGACTCAGTGATGATGATGATGATGAGATTCGCCGCCCGCTGACTTC | reverse cloning primer for *clpG*$_{GI\text{-}SG17M}$ with 6xHis-tag at the C-terminus in pBAD30 for complementation | This study |
| sHsp20_F | GAAACGCCGATCGGTGAT | forward primer to test presence of *shsp20*$_{GI}$ | This study |
| sHsp20_R | TTCATATCGATACCGAAACGC | reverse primer to test presence of *shsp20*$_{GI}$ | This study |

†restriction site is underlined

## Supplementary experimental procedures

### Genome characterization and bioinformatic analyses

ResFinder 3.2 ( https://cge.cbs.dtu.dk/services/ResFinder/; (Zankari et al. 2012)) was used to identify antibiotic resistance genes. VirulenceFinder (Joensen et al. 2014) was applied to check for encoded virulence factors. Alternatively assessment of strain relationships was done with core genome multilocus sequence typing (cgMLST) using Ridom SeqSphere+ software (Junemann et al. 2013). Target genes were defined using *E. coli* K-12 MG1655 as seed. Missing values were an own category. SNP calling was performed with CSIPhylogeny1.4 (https://cge.cbs.dtu.dk/services/CSIPhylogeny/) and phylogeny was inferred by calculation of approximately-maximum-likelihood using an implemented modified version of FastTree2 (Kaas et al. 2014). Genome sequences were retrieved from NCBI, accession numbers were: *E. coli* K-12 MG1655 NC_000913.3; Fec10 MDLJ00000000.1, *E. coli* NCTC86 NZ_LT601384.1; *E. coli* B str. REL606 NC_012967.1; *E. coli* C NZ_CP020543.1; *E. coli* ATCC 8739 (Crooks) NC_010468.1; *E. coli* W (ATCC 9637) NC_017635.1; ECOR01 LYBJ00000000; ECOR03 LYBH00000000; ECOR05 LYBF00000000; ECOR07 LYBD00000000; ECOR08 LYBC00000000; ECOR09 LYBB00000000; ECOR11 LYAZ00000000; ECOR12 LYAY00000000; ECOR14 LYAW00000000; ECOR25 LYAL00000000; *E. coli* 26561 NZ_CP027118.1; *E. coli* RR1 NZ_CP011113.2; *E. coli* TO6 NZ_LS992166.1; *E. coli* TO73 NZ_LS992171.1; *E. coli* MDS42 NC_020518.1; *E. coli* NCTC9102 NZ_LR134227.1; *E. coli* NCTC 9040 NZ_LR134247.1; *E. coli* UMNF18 NZ_AGTD01000001; *E. coli* F2_63 FZIJ00000000.1; *E. coli* Tob1 MIIH00000000.1.

To construct the unrooted phylogenetic tree, orthologous groups, amino acid sequences of all protein-coding genes were determined using ProteinOrtho V5.13 (Lechner et al. 2011) with parameters cov=67 (min. coverage of best BLAST alignments in per cent) and identity=50 (min. per cent identity of best BLAST alignments). The orthologous proteins from 521 complete and annotated genomes of *E. coli* available in GenBank in April 2019 (Coordinators 2018) and *E. coli* Fec10 and Fec6 were subsequently aligned. 238 common single-copied groups of orthologs were used for phylogenetic tree construction. Nucleotide sequences of genes in each orthologous group were aligned using MAFFT (Katoh and Standley 2013) in the linsi mode. Subsequently, RAxML (Stamatakis 2014) with the GTR+Gamma model with 100 bootstrap replicates for the concatenated gene alignment was used. Finally, the tree was visualised using the online tool iTOL (Letunic and Bork 2016).

Conserved gene clusters (synteny groups) were searched for with bi-directional best hits or a BLASTP alignment threshold with an identity >35% on >80% protein length with a gap parameter of five against sequenced *E. coli* genomes including *E. coli* K-12 MG1655 (Vallenet et al. 2020). Variable and strain specific coding sequences were compared to *E. coli* K-12 strains (K-12, DH10B, BW2952, J53 and DH5α) defined as >80% amino acid identity over >80% coverage. Core genome and variable CDSs compared to 1080 strains of *E. coli* population was determined by PPanGGOLiN 0.1.4 using standard parameters (Vallenet et al. 2020). To estimate the fraction of vertically inherited genomes of closely related strains were aligned using PG-explorer (http://mouse.belozersky.msu.ru/tools/npge.html) (Nagaev 2015). For each pair of strains aligned regions were cut into windows of 1 bb to calculate the number of mismatches. The obtained histogram was decomposed into a weighted sum of the Poisson and Erlang distributions, the former reflecting vertical inheritance, and the latter, regions acquired

by lateral transfer via homologous recombination with distantly related strains. The weight of the Poisson distribution in this decomposition equals the proportion of vertically inherited regions in the common fraction of two genomes.

**Transmission electron microscopy**

*E. coli* Fec10 cells, grown as *in situ* colonies on agar-plates, were fixed in the gas-phase from a filter paper inlay, soaked with 2% glutaraldehyde - 20 mM Hepes buffer (pH 7.2) – sticking to the top dish - at ambient temperature for 48 h. Colonies were layered with 2 % (w/v) low melting agarose at 30 $^{o}$C and were washed twice at room temperature for 5 min with 100 mM cacodylate buffer, pH 7.2. Cells were postfixed in 1% (w/v) osmiumtetroxide – 80 mM cacodylate, pH 7.2 fixative overnight at 30 $^{o}$C, followed by washing for 5 min at ambient temperature in 100 mM cacodylate buffer, pH 7.2. Dehydration started with an aqueous ethanol series (10 % (v/v), 20 min on ice; 30 % (v/v), 20 min on ice; 50 % (v/v), 20 min on ice; 1% (w/v) uranyl acetate, pH 4.5 in 60 % (v/v), 60 min at ambient temperature; 90 % (v/v), 20 min at ambient temperature; 100 % ethanol, 30 min at ambient temperature; 100 % acetone, 30 min at ambient temperature). Infiltration with a Spurr‘s epoxy resin (hard)/acetone mixture (1 part + 1 part) over night at ambient temperature, followed by resin (hard)/acetone mixture (2 parts + 1 part) over night at ambient temperature, and finally twice pure resin overnight and 7 h at ambient temperature. Sections of 3 x 2 mm of a colony‘s periphery were flat-embedded in a silicon mould by polymerisation at 75 $^{o}$C for 16 h. 90 nm ultrathin sections were cut with a diamond knife (RMC, W. Reichert, Labtec, Wolfratshausen, Germany) and a Reichert ultramicrotome (Reichert Ultracut S, Reichert-Jung, Vienna, Austria) and were picked up with 300 mesh Cu-grids. Mounted sections were post-stained with aqueous 4% (w/v) uranyl-acetate, pH 4.0 followed by 0.2xPb-Citrate (Ultrostain 2, Leica, Wetzlar, Germany), were blotted and air-dried.

Samples were analyzed with an energy-filtered transmission electron microscope (EF-TEM) (Libra 120 plus, Zeiss, Oberkochen, Germany). Electron micrographs were recorded with a bottom-mount, cooled 2048x2048 CCD frame transfer camera (SharpEye; Tröndle, Wiesenmoor, Germany) at close to low dose conditions in the elastic bright field mode as Zero-loss images (slit width: 10 eV; 0.5 mrad illumination aperture, 60 µm objective aperture, beam current: 1 µA), close to the Scherzer focus.

**REFERENCES**


Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins & J. D. Thompson (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res,* 31**,** 3497-500.

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

Coordinators, N. R. (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res,* 46**,** D8-D13.

Darzentas, N. (2010) Circoletto: visualizing sequence similarity with Circos. *Bioinformatics,* 26**,** 2620-1.

Joensen, K. G., F. Scheutz, O. Lund, H. Hasman, R. S. Kaas, E. M. Nielsen & F. M. Aarestrup (2014) Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol,* 52**,** 1501-10.

Junemann, S., F. J. Sedlazeck, K. Prior, A. Albersmeier, U. John, J. Kalinowski, A. Mellmann, A. Goesmann, A. von Haeseler, J. Stoye & D. Harmsen (2013) Updating benchtop sequencing performance comparison. *Nat Biotechnol,* 31**,** 294-6.

Kaas, R. S., P. Leekitcharoenphon, F. M. Aarestrup & O. Lund (2014) Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One,* 9**,** e104984.

Katoh, K. & D. M. Standley (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol,* 30**,** 772-80.

Lechner, M., S. Findeiss, L. Steiner, M. Marz, P. F. Stadler & S. J. Prohaska (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics,* 12**,** 124.

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**,** 147-54.

Letunic, I. & P. Bork (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 44**,** W242-5.

Nagaev, B., Nikolaev, M., Alexeevski, A. 2015. NPG-explorer: a new tool for nucleotide pangenome construction and closely related prokaryotic genomes analysis; http://mccmb.belozersky.msu.ru/2015/proceedings/abstracts/197.pdf. In *Proceedings 7th International Moscow Conference on Computational Molecular Biology MCCMB-2015* Moscow.

Nguyen, S. V., G. P. Harhay, J. L. Bono, T. P. Smith & D. M. Harhay (2017) Genome sequence of the thermotolerant foodborne pathogen *Salmonella enterica* Serovar Senftenberg ATCC 43845 and phylogenetic analysis of loci encoding increased protein quality control mechanisms. *mSystems,* 2:e00190-16.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

Sullivan, M. J., N. K. Petty & S. A. Beatson (2011) Easyfig: a genome comparison visualizer. *Bioinformatics,* 27**,** 1009-10.

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.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp & G. J. Barton (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* 25**,** 1189-91.

Wei, Y., A. A. Thyparambil & R. A. Latour (2014) Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim Biophys Acta,* 1844**,** 2331-7.

Zankari, E., H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup & M. V. Larsen (2012) Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother,* 67**,** 2640-4.