Evolution towards small colony variants of pandemic multidrug resistant ST131 *Escherichia coli* isolates from a 10-year bone infection

Fengyang Li[1#], Petra Lüthje[2,3#], Sulman Shafeeq[1#], Marcus Ahl[4], Janja Trcek[5], Analucia Diaz Lacosta[3], Volkan Özenci[3], Måns Ullberg[3], Börje Åkerlund[4] and Ute Römling[1*]

[1]Department of Microbiology, Tumor and Cell Biology, Biomedicum, Karolinska Institutet, SE-171 77 Stockholm, Sweden
[2]Staatliches Tierärztliches Untersuchungsamt (STUA) Aulendorf - Diagnostikzentrum, Aulendorf, Germany
[3]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden
[4]Department of Medicine, Unit of Infectious Diseases, Karolinska University Hospital, Stockholm Huddinge, Sweden
[5]Department of Biology, Faculty of Natural Sciences and Mathematics, Univeristy of Maribor, SE-2000 Maribor, Slovenia

#, contributed equally
*corresponding author. E-mail: Ute.Romling@ki.se

ABSTRACT

Background

Chronic wounds are usually challenging to treat due to underlying medical conditions of the individual and as they readily become infected by microorganisms due to the failure of mechanical and physiological first line innate immune responses. We report here the characterization of host adaptation of five *E. coli* genomes including three *E. coli* ST131 genomes that occurred concomitantly with *Enterococcus faecalis* from a 10-year chronic wound infection after a foot fracture during the 2004 tsunami.

Methods

The five *E. coli* strains were characterized by various microbiological and genomic approaches. Microbiological methods were antimicrobial resistance, growth in different media and biofilm formation. Genomic methods were determination of the genome sequence by PacBio RSII and Illumina sequencing. Phylogenetic analyses and genome alterations such as single nucleotide polymorphisms, deletion and rearrangements that led to pseudogenes and chromosomal inversions were documented. Relevant selected metabolic and physiological pathways were analyzed for integrity.

Finding

Strains of two initially present sequence types, including the highly antimicrobial resistant ST405 clone, were subsequently replaced by isolates of the ubiquitous ST131 clone. The three *E. coli* ST131 strains showed a heavily host-adapted genome with a high number of pseudogenes and a large chromosomal inversion compared to ST131 reference strains. Furthermore, two of three *E. coli* ST131 isolates were small colony variants with its genetic basis in multiple genome alterations including pseudogenes and deletions in the pathway for heme biosynthesis. Pseudogene analysis indicated also the three ST131 strains to be mutator strains. Although enhanced capability of biofilm formation of the ST131 isolates was indicated by the agar plate assay, the a liquid culture biofilm assay did not display pronounced biofilm formation suggesting unconventional modes of biofilm formation.

Interpretation

ST131 clone members, which originally appeared as commensal strains can cause urinary tract and blood stream infections and are ubiquitously found in the environment including waste water and in animals. ST131 strains have presumably been already acquired from the environment on occurrence of the initial foot fracture and can persist in wounds showing an outmost genome plasticity and adaptability which might causing the chronic infection. Although co-infection with *E. faecalis* might have supported chronicity, these findings indicate that in individuals with underlying metabolic diseases wound infection by ST131 *E. coli* isolates can be a health risk.

Funding

This work was partially funded by ALF.

Keywords: bone infection, chronic infection, Escherichia coli, multiresistance, small colony variants, ST131

INTRODUCTION

Being a ubiquitous commensal of the gastrointestinal tract of almost every human isolates of the genetically diverse species *Escherichia coli* can readily develop into a pathogen by few subsequent genetic alterations, habitat switch or overgrowth. As examples, the probiotic strain Nissle 1917 is closely related to the uropathogenic isolate CT087 (Cimdins, Simm et al. 2017). Enterotoxigenic *E. coli* arise by the acquisition of virulence plasmids and few genomic alterations (von Mentzer, Connor et al. 2014). Stress conditions and impaired immune responses can lead to microbial overgrowth and translocation (Berg 1992) Use of antimicrobial agents has substantially triggered the development of distinct multidrug resistant clones in Bacteria with the species *E. coli* being no exception. The sequence type ST131 of *E. coli* is today the most widely distributed clone among fluoroquinolone and/or extended spectrum beta-lactamase (ESBL) isolates world-wide (Nicolas-Chanoine, Bertrand et al. 2014; Schembri, Zakour et al. 2015; Pitout and Finn 2020; Cummins, Snaith et al. 2021).

The development of ST131 clone members from antibiotic susceptible isolates to pandemic highly prevalent multresistant pathogens can be monitored from within the 19th century and has accelerated in the 1980ies/1990ies. Thereby, three temporarily evolved genetically clearly distinguishable clades with different antimicrobial resistance patterns, alleles (single nucleotide polymorphisms), plasmid acquisition and accessory genomes, clades A, B and C can be discriminated, with an expansion of antimicrobial resistant clade C1 and C2 strains (Pitout and Finn 2020).

The ST131 high-risk clone *E. coli* is predominantly associated with community as well as hospital acquired (nosocomial) urinary tract and blood stream infections in humans (Nicolas-Chanoine, Bertrand et al. 2014). ST131 can efficiently colonize the gastrointestinal tract and is found asymptomatically in community carriers which contributes to its successful spread. Further more, ST131 isolates can be readily transmitted from person to person. However, ST131 members can even cause disease in animals including companion animals. Transmission occurs by ST131 present in poultry and other food. In the environment ST131 members can persist in wastewater as a habitat.

Although not exceptionally antibiotic resistant as compared to other *E. coli* linages such as ST405, the success of ST131 is partly based on its ability to sequentially have acquired point mutations and horizontally transferred genes that contribute to enhanced antimicrobial resistance. Early developed resistance was against fluoroquinolones and subsequently ESBL genes with CTX-M-14 and CTX-M-15 the most prominent had been acquired on different plasmid backgrounds.

Pathogens can adopt two fundamentally different successful strategies which lead to either an acute or a chronic infection. While expression of host detrimental virulence factors leads to acute infection that is readily tackled by antimicrobial therapy, the chronic infection status is associated with biofilm formation and the formation of persister cells. Small colony variants (SCV) are a specific form of host adapted high biofilm forming isolates (Haussler, Tummler et al. 1999; Drenkard and Ausubel 2002; Kahl, Becker et al. 2016). These host adaptated metabolic variants which can arise readily during chronic infection grow slower on standard medium compared to their otherwise fast *in vitro* growing counterparts. Already at the beginning of the 20th century have small colony variants of *Staphylococcus aureus* been recognized and *Pseudomonas aeruginosa* develops SCVs during chronic lung infection in patients with an underlying defect in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), a chloride channel predominantly expressed on epithelial surfaces. Although auxotrophy for various classes of metabolites such as hemin, menachinone, thymidine, CO_2 and lipoic acid has been observed, the most common physiological denominator is a defect in the electron transport chain and decrease in the proton motif force

(PMF) that ultimately leads to increased resistance to aminoglycosides, but also increased susceptibility to antibiotics that are substrates of PMF-dependent efflux pumps (Roggenkamp, Sing et al. 1998; Lazar, Pal Singh et al. 2013; Santos and Hirshfield 2016; Xia, Tang et al. 2017; Matsumoto, Hashimoto et al. 2020). Consequently exposure to aminoglycosides and other antibiotics, but also the immune system, antimicrobial peptides, the intracellular host environment and biofilm formation have been recognized as selective pressures that leads to selection for small colony variants *in vitro* (Yegian, Gallo et al. 1959; Musher, Baughn et al. 1979; Cano, Pucciarelli et al. 2003; Pranting and Andersson 2011; Ramiro, Costa et al. 2016). Nevertheless, small colony variants can have an advantage *in vivo* including being high biofilm formers (Drenkard and Ausubel 2002; Starkey, Hickman et al. 2009; Ramiro, Costa et al. 2016). Small colony variants of *E. coli* have rarely been observed, although the phenotype has been reported to arise during urinary tract infection, bacteremia and chronic hip infection (Roggenkamp, Sing et al. 1998; Tappe, Claus et al. 2006; Park, Le Phuong et al. 2018).

Not mutually exclusive adaptation mechanisms during chronic infections include the emergence of large chromosomal inversions, arisal of mutator strains and expansion of IS elements (The, Thanh et al. 2016). As large chromosomal inversions are associated with speciation, their occurrence upon chronic infection indicates niche adaptation (Ginard, Lalucat et al. 1997; Römling, Schmidt et al. 1997). Mutator strains which lack repair of replication errors facilitate adaptation, but might on the other hand lead to detrimental accumulation of mutations (Denamur and Matic 2006; Engelhardt and Shakhnovich 2019).

In this work, we describe the features of five *E. coli* isolates that were recovered from a more than 10 year long chronic wound and bone infection. All the multiresistant isolates harboured IncFII based multireplicon plasmids with different versions of ESBL genes. Notably, three persistent isolates were host adapted members of the pandemic clone ST131 with chromosomal alterations such as pseudogenes large genome rearrangements that had developed a small colony variant phenotype.

RESULTS

Case description. An individual with a foot fracture caused during the 2004 tsunami suffered post-traumatically from a ten year long chronic bacterial infection (Figure1, Ahl et al, manuscript in preparation). Microbiological analysis recovered, besides a diverse microbiological flora at the initiation of the infection, consistently multiresistant *E. coli* and *Enterococcus faecalis* isolates. Although rare, we estimated *E. coli* to be the causative agent of the recurrent infection that was eventually cured in 2015 (Figure1, Ahl et al, Römling, manuscript in preparation).

Antibiotic profile of bone isolates. Strains of the species *E. coli* infrequently cause persistent bone or tissue infections. To understand the molecular basis of persistence, we analyzed five *E. coli* isolates recovered at the beginning of the infection in 2005 (Bone1B and Bone2, same isolation date with different colony morphologies on blood agar plates), in 2013 upon recurrence of the infection after a >seven year silence (Bone4) and Bone8 and Bone7 within a three months interval in 2015 shortly before the infection was cured (Figure 1). All strains were multiresistant and commonly showed resistance against the fluoroquinolone ciprofloxacin and the third-generation class cephalosporins cefotaxim and ceftazidime (Supplementary Table 1; Supplementary Figure 1). Furthermore, all strains except Bone2 displayed gentamicin resistance, and, besides Bone8 and Bone7, were resistant against trimetroprime/sulfamethoxazole.

Sequence type classification of isolates. Pulsed-field gel electrophoresis (Schwartz and Cantor 1984; Römling, Wingender et al. 1994) and, after subjecting the five *E. coli* isolates to PacBio and Illumina sequencing, subsequent *in silico* multilocus sequence typing (MLST) of seven genes and the core genome (Feng, Zou et al. 2021) determined that the initial isolates Bone1B and Bone2 showed different genetic backgrounds and sequence types (Supplementary Table 1). In particular, while Bone1B and Bone2 were sequence type (ST) ST940 and ST405, respectively, the latter especially antimicrobial resistant; the subsequent isolates, Bone4 in 2013 and Bone8 and Bone7 in 2015 belonged to ST131, the pandemic multiresistant *E. coli* clone predominantly causing urinary tract infections. ST131 strains have diversified into three major clades with a recent expansion of clade C with subclades C1 and C2. Bone4, Bone8 and Bone7 clustered with clade C1 strains (Figure 2A). The basic genomic characteristics of the strains are summarized in Supplementary Table 1.

Acquired genes and altered chromosomal targets confer antibiotic resistance. Distinct, but related, multireplicon IncFII plasmid harbouring the majority of antimicrobial resistance determinants are present in all strains (Supplementary Table 1; Supplementary Figure 1; Figure 3). While Bone1B and Bone2 IncFII plasmids harbor a broad-spectrum CTX-M-15 beta-lactamase; in agreement with their phylogenetic position, a broad-spectrum CTX-M-14 beta-lactamase characteristic for ST131 clade 1 strains is present on the IncFII plasmid of Bone4, Bone8 and Bone7 (Supplementary Figure 2). Additional acquired resistance determinants are encoded by alternative plasmids or on the chromosome. Of note, the IncFII plasmids of the ST131 strains have, concomitantly with diminished antimicrobial treatment, successively lost their antimicrobial resistance determinants (Supplementary Figure 2). Single nucleotide polymorphisms are not only present in the characteristic chromosomal targets mediating fluoroquinolone resistance in ST131 strains such as the two gyrase subunits GyrA and GyrB and the ParC and ParE subunits of DNA topoisomerase IV showing the characteristic amino acid substitutions, but encompass additional targets that potentially mediate pulvomycin, fosfomycin and multidrug resistance (Supplementary Figure 2). In

summary, a broad panel of acquired resistance genes and point mutations has accumulated in the strains due to the intensive antibiotic therapy during the 10-year infection.

Phenotypes of the sequential bone isolates

Growth of the strains on different agar media showed that while Bone1B and Bone2 grew well on all media, ST131 isolates Bone8 and Bone7 grew similar to Bone4 on blood and Mueller Hinton Fastidious agar, but displayed a small colony variant phenotype on Mueller Hinton agar medium (Supplementary Figure 3A). Partial growth restoration occurred upon provision of hemin (factor X), but not nicotinamide-adenine-dinucleotide (NAD; factor V), auxotrophies for both factors are commonly associated with microbial host adaptation (Supplementary Figure 3B).

Biofilm formation is known to be a major determinant for the chronic infection process (Costerton, Stewart et al. 1999; Römling and Balsalobre 2012). Among the biofilm phenotypes of *E. coli*, most ancient and conserved is the rdar morphotype composed of amyloid curli fimbriae and the exopolysaccharide cellulose, while a poly-N-acetyl-glucosamine based biofilm, also characteristically visible on agar plates, is variable among strains and subject to horizontal gene transfer (Zogaj, Nimtz et al. 2001; Wang, Preston et al. 2004; Cimdins et al. 2017). Assessment of the ability of the isolates to express rdar or PNAG-based biofilm indicated that the ST131 isolates Bone4, Bone8 and Bone7 gradually showed enhanced dye binding, while Bone1B and Bone2 isolates did not display dye binding indicative for biofilm formation (Supplementary Figure 3C).

Initial characterization of the chromosomes of ST131 bone isolates

Genome annotation revealed 5371 genes (5257 coding sequences (CDSs)) for isolate Bone4, 5267 genes (5155 CDSs) for Bone8 and 5290 genes (5177 CDSs) for Bone7 (Supplementary Table 1). A number of phages are encoded by the three genomes (Supplementary Table 2). Of note is phage mEp460 which occurs in two distinct copies in the Bone4 and Bone7 genomes, but has been deleted from the Bone8 genome. Interestingly, a high number of pseudogenes had been annotated for the genomes of the three isolates (Figure 4; Supplementary Table 3). Already the genome of Bone 4 has 1141 annotated pseudogenes. Among the nucleotide alterations of those (partly) nonfunctional genes were 1033 frameshifts, 159 incomplete genes, 40 genes with internal stop and 85 genes with multiple distortions (information as from the NCBI annotation pipeline). Subsequently, Bone8 displayed 2287 pseudogenes, whereby 2176 were frameshifted, 211 incomplete, 41 had an internal stop and 132 showed multiple distortions. 1231 pseudogenes were identified in Bone7, with 1123 genes frameshifted, 153 incomplete, while 22 genes had an internal stop and 74 showed multiple distortions. A total of 301 pseudogenes occur in all three isolates. This contrasts with the initial recovered isolates Bone1B and Bone2, which do not possess an elevated number of pseudogenes.

We investigated whether pathways are specifically subject to pseudogene evolution. Investigating of the biofilm pathways showed that the ST131 isolates possessed impaired biofilm genes clusters. While the gene clusters for biosynthesis of amyloid curli fimbriae including the rdar biofilm regulator CsgD and the exopolysaccharide cellulose seem to be intact, the gene cluster for poly-N-acetyl-glucosamine biosynthesis is fragmented in Bone8 (Supplementary Figure 4). The gene cluster for type 1 fimbriae of ST131 isolates has the characteristic two gene insertion which disrupts the recombinase FimB. Of note, Bone1B lacks the type 1 fimbrial gene cluster.

Cyclic di-GMP is a ubiquitous second messenger that positively regulates biofilm formation including expression of the biofilm activator *csgD* and coregulated extracellular matrix components curli and cellulose and type 1 fimbriae. We subsequently determined whether

cyclic di-GMP turnover proteins are among the pseudogenes. The cyclic di-GMP turnover proteins of Bone4, Bone8 and Bone7 were compared with the proteins of the ST131 clade reference strains, clade A strain SE15, clade B strain JJ1987, clade C1 strains MRSN17749 and AR_0058 and clade C2 strain JJ1886 (Supplementary Figure 5; Supplementary Table 4). Compared to the commensal isolates *E. coli* K-12 MG1655 and Fec10, the clade reference strains possessed mostly protein variants containing several amino acid substitutions in the cyclic di-GMP turnover proteins with the occasional deletion of a cyclic di-GMP turnover gene or transformation into a pseudogene. In all reference strains, the diguanylate cyclase DosC had been deleted and only the catalytic domain of the phosphodiesterase (PDE) YcgG remained. Furthermore, lack of the transmembrane domain of the GGDEF-EAL protein YegE also led to a cytoplasmic protein. In the representative clade A and C1 strains SE15 and MRSN17749 the diguanylate cyclase YedQ is absent.

In sharp contrast, the majority of the cyclic di-GMP turnover proteins in the ST131 Bone isolates were substantially altered, besides harbouring amino acid substitutions either truncated or fragmented, however, sometimes even with substantial additional amino acid sequences at the N- or C-terminus (Supplementary Table 4). The Bone isolate chromosomes possesses an identical sequence deletion leading to a truncated PDE YcgG (Supplementary Figure 6). Conserved compared to references proteins among all three ST131 isolates was the CSS-EAL phosphodiesterase YlaB (PdeB).

Molecular basis of small colony variants

Hemin auxotrophy causes small colony variant growth. Investigation of the molecular basis of hemin deficiency showed, in contrast to the early isolates Bone1B and Bone2, several genes in the pathway to hemin synthesis to be deleted and/or to be pseudogenes in Bone8 and Bone7 (Figure 5). In particular, *hemB* encoding the first enzyme in the pathway dedicated to heme biosynthesis had been inactivated, while in Bone8 also *hemL*, which synthesizes the precursor 5-aminolevulinic acid to channel into the heme biosynthesis pathway, is a pseudogene. Of note, Bone4, although not heme auxotroph, harbors already a *hemH* pseudogene.

As growth is not restored to wild type level upon heme supplementation, we reasoned alternative genes to cause a SCV phenotype are impaired in functionality. Indeed, besides the heme pathway, the biosynthesis pathway for menaquinone, a component of the electron transport chain, turnover of ppGpp and, partly, the CO_2 interconversion pathway with one of two carbonic anhydrases contains pseudogenes in Bone4, Bone8 and Bone7 (Supplementary Figure 7A). Indeed, Bone4 is slightly impaired in growth compared to reference *E. coli* isolates (data not shown).

As Bone4, Bone8 and Bone7 genomes have evolved rapidly in its human host, we were wondering whether the evolution of these isolates had been accelerated by mutations in genes that cause a significantly higher mutation rate transforming the isolates into mutator strains. Indeed, in all three strains, but not in the early Bone1B and Bone2 isolates, multiple, at least three previously confirmed mutator genes have been impaired (Supplementary Figure 7B). For example, *mutS* involved in DNA mismatch repair has been inactivated in Bone8, while *mutM* executing base excision repair has been inactivated in Bone4 and Bone7.

Insertions and deletions in Bone isolates

We further investigated whether the chromosomes of ST131 isolates had additional characteristics. Compared to Bone1 and Bone2, Bone4/8/7 lacked a type III and a type VI secretion system. At least nine predicted to be complete, three incomplete and one questionable phage(s) were identified integrated into the chromosome (Supplementary Table 2). Identified complete phages belonged to different classes such as Myoviridae, Peduvirus,

Lambdavirus, Siphoviridae, Marienburgvirus, Uetakevirus and others. We were wondering whether ST131 Bone isolates, as exemplified with Bone4, contained specific genome characteristics compared to representative clade 1 isolates. AR_0058 was chosen for comparison as this isolate has a completely sequenced genome. Seven regions >20 kbp were found, whereby six regions represented insertions in Bone4 and one region a deletion in Bone4 compared to the AR_0058 chromosome. Phages ΦAA91-ss, mEP460 (2x), Shigell_SfII and TL-2011b are present in Bone4, but not AR_0058. Furthermore, the chromosomal region with genes coding for the general type II secretion system, antigen 43 and the poly-sialic acid capsule K1 is absent in AR_0058. This chromosomal region is subject to variability in ST131 isolates as genes encoding antigen 43, the K1 capsule, but also aerobactin and Pap pili are absent in clade A strain SE15 (data not shown). Present in AR_0058 and absent in Bone4 is the Lederbergvirus Entero_Sf101.

Major genome rearrangements within ST131 Bone strains

As among Bone4, Bone8 and Bone7 isolates multiple small scale genome alterations such as single nucleotide polymorphisms and deletions were observed, we were wondering whether also large scale chromosomal rearrangements have occurred. Indeed, Bone8 harbors two larger deletions of >40 kbp compared to Bone4 (Supplementary Figure 8). While deletion I comprises one copy of the phage mEp460 inserted into the tRNA threonine locus, deletion II spans the core genome genes between *lacY* and *iraP* including *hemB*, the entire 3-(3-hydroxyphenyl)propionate degradation pathway and the taurine uptake system.

The genomes of representative strains of the major ST131 clades are in synteny with the genome of the reference strain *E. coli* K-12 MG1655. Comparison of the genomes of Bone4, Bone8 and Bone7 with the ST131 reference genomes showed distinct large scale genome rearrangements caused by homologous recombination between two rRNA operons (Figure 2B). Considering that the origin of replication has remained in place, the inversion of almost the entire genome observed in the Bone8 chromosome does not strongly impair the position of the terminus of replication, while the >3.5 Mbp inversion of the Bone4 chromosome causes a substantial chromosomal dissymmetry, which might contribute to impaired growth.

DISCUSSION

E. coli ST131 and *E. faecalis* have been causative agents of a chronic wound infection that was eventually cured after a more than 10-year infection. As chronic tissue infections with *E. coli* are rare, in this work, we have characterized the genomic features of these ST131 strains as compared to the *E. coli* strains initially found in the wound. Chronic infections are associated with the microbes to adapt to the host environment which can lead to substantial alterations in the genome (Römling, Schmidt et al. 1997; Kresse, Blocker et al. 2006). Indeed, the ST131 Bone4 strain isolated nine years after the onset of the chronic infection and seven years after the infection was initially silenced shows substantial genomic alterations with more than 1000 pseudogenes and a large chromosomal inversion. Bone8 isolated two years later had evolved even >2200 pseudogenes, which might represent an upper limit of host adaptation by genome alterations in a previously free-living organism. Pseudogenes developed from genes that cause a mutator phenotype upon misfunction certainly supported the accelerated genome evolution. Not all annotated pseudogenes are necessarily out-of-function, though, as exemplified with the PDE YcgG which consists only of the catalytic EAL domain deleted by its N-terminal signaling domain which can show still, although altered, functionality (Zlatkov and Uhlin 2019). Thus, pseudogenes might have acquired a novel functionality and/or expression pattern (Ojha, Dittmar et al. 2021). This genome evolution, including the arisal of pseudogenes leading to the development of small colony variants, on a short time scale resembles, for example, the evolutionary fixed evolution of the pathovar *Shigella* from commensal *E. coli*, which occurred on several occasions on an evolutionary short time scale between 270,000 and <400 years ago (The, Thanh et al. 2016).

Furthermore, large chromosomal inversions have been shown to accompany chronic infections and host adaptation, for example, in *Pseudomonas aeruginosa* during chronic lung infection of cystic fibrosis patients (Römling, Schmidt et al. 1997; Kresse, Dinesh et al. 2003). Those inversions, which can occur among rRNA operons, but recombination can also occur within alternative nearly identical longer sequences such as IS elements, can cause a significantly dissymmetric chromosome (Cui, Neoh et al. 2012). Thereby, large chromosomal inversion can modify the mode and direction of chromosomal evolution. It will certainly be of interest to assess the transcriptome and metabolome of those host-evolved Bone strains in order to get deeper insights into the altered cell physiology.

A hallmark of the ST131 Bone isolates has been their growth retardation/SCV phenotype. In most reported clinical SCVs, one single mutation, though in different pathways, is causative for this phenotype. Bone8 and Bone7, but even Bone4 have, however, accumulated a number of gene alterations that can be the genetic foundation for a SCV phenotype again demonstrating accelerated evolution.

We speculate that a ST131 isolate had been acquired upon the initial infection event from the environment and had been subsequently selected for during long term antimicrobial treatment. If so, only the ST131 clone isolate had persisted in the tissue despite of an equally or even more extended antimicrobial resistance profile and presence of virulence factors such as the type III secretion system of the genetically unrelated Bone1B and Bone2 strains. ST131 clone isolates are known for their ability to persist in the host and the environment (Nicolas-Chanoine, Bertrand et al. 2014; Schembri, Zakour et al. 2015). The infection spectrum of ST131 isolates is wide and ST131 strains have been reported to occasionally even cause keratitis and meningitis (Assimacopoulos, Johnston et al. 2012; Pouillot, Chomton et al. 2012). Of note, although the occurrence of *E. coli* is rare, an anecdotical observation reported a chronic tissue infection with *E. coli* ST131 from an environmental source with horizontal transfer of plasmid from a co-infecting *Morganella* strain (McGann, Snesrud et al. 2015). Thus, environmental acquisition of ST131 by an open wound leading to a persistent infection does occur under certain circumstances. The yet to be discovered molecular mechanisms which provide the ST131 clonal members with an exceptional capability to persist might also aid its long-term survival in host tissue.

MATERIAL AND METHODS

Bacterial strains

Investigated *E. coli* isolates were Bone1 (isolated 01/01/2005), Bone2 (isolated 01/01/2005), Bone 4 (isolated 09/06/2013), Bone8 (07/07/2015), Bone7 (24/10/2015). Strains were grown on LB without salt plates, blood agar plates, Mueller Hinton agar and Mueller Hinton fastidious agar.

Phenotypic characterisation

Resistance against antibiotics was determined according to EUCAST standards. Assessment of extracellular matrix production biofilm formation on Congo Red agar plates was performed as described besides that LB without salt agar plates were used (Hammar, Arnqvist et al. 1995). Of note, Congo Red agar plates are also used to assess membrane integrity, with microbes showing impaired cell envelope integrity to stall growth.

Isolation of high molecular weight genomic DNA

High molecular weight genomic DNA was isolated with the QIAGEN Genomic-tip 500/G using the provided buffers and following the instructions of the manufacturer. Integrity and purity of the extracted DNA was analyzed by measuring the 260/280 nm ratio by NanoDrop and running the genomic DNA on an 0.8% agarose gel in 1x TAE buffer.

Whole genome sequencing

Whole genome sequencing for all strains was performed with the PacBio RS II system (Pacific Biosciences at NGI Uppsala, Science For Life Laboratory [SciLifeLab], Uppsala, Sweden). The assembly was performed on SMRT portal version 2.3, with HGAP3 default settings. Subsequently, all strains were sequenced on the Illumina MiSeq version 3 platform with read length up to 2 x 300 bp (NGI Stockholm, SciLifeLab, Solna, Sweden). Illumina sequence reads for Bone 4, Bone 8 and Bone 7 were mapped on the PacBio RII created assembly in UGENE (Okonechnikov, Golosova et al. 2012) and manually inspected for sequencing inconsistencies. Few SNPs that showed a difference between the PacBio RII and Illumina sequencing were detected, but a large number of SNP heterogeneity was observed within the Illumina reads, possibly reflecting adaptation of the strains to laboratory conditions. Closing of the chromosome sequence of Bone4 and Bone7 which had two non-overlapping contig and two junctions with almost identical overlapping ends, respectively, was concluded by PCR using primers located at the ends of the contig. Genome sequences have been submitted to the NCBI database.

Bioinformatic analysis

The genomes were preliminary annotated with RASTtk and in MicroScope (Brettin, Davis et al. 2015; Vallenet, Calteau et al. 2020). Genome sequences were submitted to the NCBI database and annotated with the NCBI prokaryotic genome annotation pipeline (PGAP; (Tatusova, DiCuccio et al. 2016)). BLAST was used to assess homology of proteins and nucleotide sequences (Altschul, Gish et al. 1990).

Determination of the MLST types was done with http://bacdb.org/BacWGSTdb/Tools_results_multiple.php (Feng, Zou et al. 2021) assessing 7-gene MLST (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*). Virulence genes were determined by VirulenceFinder 2.0; search with standard parameters, % identity >90% and minimum length 60% (https://cge.cbs.dtu.dk/services/VirulenceFinder/). Resistance genes were determined by ResFinder 4.1 (https://cge.cbs.dtu.dk/services/ResFinder) and the Comprehensive Antibiotic Resistant Database (CARD; https://card.mcmaster.ca/) to assess

acquired and intrinsic resistance genes including *parC* and *gyrA* alleles. Origin of replication has been identified by PlasmidFinder 2.0.1 with standard parameters of 95% identity and 60% minimum coverage (https://cge.cbs.dtu.dk/services/PlasmidFinder/) (Carattoli, Zankari et al. 2014). FimH alleles were determined by FimFinder 1.0 (https://cge.cbs.dtu.dk/services-/FimTyper) and serotype was determined by SerotypeFinder all at the Centre of Genomic Epidemiology website (http://www.genomicepidemiology.org/).

Phylogenetic analysis of ST131 genomes

To perform genome-based phylogenetic analysis by core genomes, FASTA genomic sequence files of the representative ST131 from different clades were downloaded from the NCBI and EnteroBase websites considering the completeness of the individual genomes. To ensure equal analysis treatment of the different genomes, reannotation of the genomes was performed using the annotation package Prokka 1.14.5 (Seemann 2014). Subsequently, the core genomes from the target data set were calculated by Roary software 3.11.2 with a threshold of 90% identity (Page, Cummins et al. 2015). Alignments from the conserved genes were concatenated to calculate a maximum-likelihood genome-based phylogenetic tree using PhyML 3.0 of the SeaView software 5.0.4 (Gouy, Tannier et al. 2021). If not otherwise indicated, the respective reference genomes were used.

Visualisation tools

Annotations were visualized and compared with EasyFig (Sullivan, Petty et al. 2011). Circoletto was used to compare plasmid sequences (Darzentas 2010). MEGA 7.0 was applied to visualize the phylogenetic tree of ST131 genomes (Kumar, Stecher et al. 2018).

.

Figure legends

FIG 1 Time line of the chronic infection after foot fracture caused during the 2004 tsunami. Red arrow, date of injury; orange arrow, relapse and measures. Red star, recovery of *E. coli*; yellow star, recovery of *E. faecalis*; black star, investigated isolates. Designation of *E. coli* strains by number is indicated; red=ST904; blue=ST405 and green=ST131.

FIG 2 Classification of the ST131 bone isolates. A) Maximum likelihood phylogenetic tree of ST131 *E. coli* clades B and C showing the position of Bone4, Bone8 and Bone7 isolates within ST131 strains of clade A, B and C. B) Comparison of Bone4 and Bone8 (Bone7 is equivalent) chromosomes with the chromosomes of representative isolates of major ST131 clades. Ribosomal RNA operons are indicated by red arrows in the outermost ring. All reference strains of clade A, B and C have a gene synteny and the seven rRNA operons arranged around the origin of replication (*ori*) comparable to *E. coli* K-12 MG1655, although the t-RNA content differs at the respective position (data not shown). Five rRNA are in transcribed in clockwise and two rRNA operons are transcribed in the counter-clockwise direction. Bone4 and Bone8 possess a large chromosomal inversion around the ori which produces an asymmetric chromosome with the terminus of replication (*ter*) been shifted as indicated by the G+C skew of the leading strand. The G+C skew and G+C content of the genome is indicated in the inner circles. Reference strains for the ST131 clades *E. coli* SE15, clade A (not displayed); *E. coli* JJ1897, clade B; *E. coli* AR_0058 and MRSN17749, clade C1 and *E. coli* JJ1886, clade C2.

FIG 3 Circular map of the 156.2 kbp large multireplicon IncFII plasmid pBone4. The sequence has been normalized to the IncFII origin of replication. An IncFIA, IncFIB and a Col156 origin of replication is also present as indicated by green bars in the third circle. Of note, antimicrobial resistance determinants cluster around the IncFII origin of replication. The 184 open reading frames, color coded according to functionality, are shown in the outermost circle. The G+C skew and G+C content is indicated as displayed. The plasmid map has been created with CG view (Grant and Stothard 2008) with programme embedded annotation of the open reading frames with Prokka. Note that the annotation using Prokka differs from the prokaryotic annotation pipeline of NCBI.

FIG 4 Venn diagram of the number of unique and common pseudogenes in Bone4, Bone8 and Bone7

FIG 5 Impairment of functionality of genes in the heme biosynthesis and heme related pathway potentially causative for growth as small colony variants. A) Heatmap of functionality of genes of the heme biosynthesis and heme related pathway. Green, gene functional; red, pseudogene; light red, gene deleted. B) Heme biosynthesis pathway. Proteins indicated in light violet, impaired or absent; violet genes, functional genes.

Supplementary FIG 1 Heatmaps of plasmid and chromosomal antibiotic resistance determinants. A) Plasmid encoded resistance determinants. Resistance determinants: beta-lactamases: class A TEM-1B, class A CTX-M-14, class A CTX-M-15 and class D blaOXA-1. aminoglycoside resistance by *aac(3)-IIa* and *aac(3)-IId* encoding aminoglycoside-N(3)-acetyltransferases, *aph3-Ib* and *aph6-Id* encoding aminoglycoside 3'-phosphotransferase and *aadA5* encoding aminoglycoside nucleotidyltransferase; fluoroquinolone resistance by *aac(6')-Ib-cr* encoding aminoglycoside-(6)-N-acetyltransferase: chromate resistance by *chrA* encoding an efflux protein*;* macrolide resistance by *mph(A)* encoding macrolide 2´-phosphotransferase; *MFS* and *MFS tet(A)* encoding multidrug transporters*;* chloramphenicol resistance by *catB3* encoding type B-3 chloramphenicol O-acetyltransferase*;* quaternary ammonium compound-resistance *qacE; tunicamycin* resistance protein; trimethoprim resistance by *dfrA12* and *dfrA17* encoding dihydrofolate reductases; sulfonamide resistance *by sul1* and *sul2 e*ncoding dihydropteroate synthases. Of note, Bone1B and Bone2, but not the ST131 clade 1 strains Bone4, Bone8 and Bone7 harbor the class A CTX-M-15 beta-lactamase. Red, presence of the resistance cassette on the major IncFII plasmids; dark red, resistance cassette encoded by another plasmid; green, absence of the resistance cassette. B) Chromosomally encoded resistance determinants. Amino acid substitutions in chromosomal genes that mediate resistance are indicated in red. Amino acid substitutions in black have not been demonstrated to mediate resistance. A streptothricin resistance cassette is encoded on the Bone1B genome.

Supplementary FIG 2 Comparison of selected IncFII plasmids of Bone isolates with most homologous plasmids from the database. A) Comparison of plasmid pBone4IncFII with pU12A_A with localization of major resistance determinants. Resistance determinants as indicated in Supplementary Figure 1. *tonB*, encoding periplasmic iron compounds and cobalamin uptake protein; *tonB_R*, outer membrane receptor for iron compounds and cobalamin; *senB*, toxin. B) Comparison of plasmid pBone1BIncFII (above) with p4_1_1 with localization of major resistance determinants. Resistance determinants as indicated in Supplementary Figure 1. C) Comparison of IncFII plasmids pBone4 with pBone8 and pBone7 plasmids. Plasmids were normalized to the IncFII origin of replication. Sequence orientation is clockwise (green , 5'start of sequence as in Figure 3; red, end of plasmid sequence). pBone8 and pBone7 sequences are to a major part in synteny with pBone4 (red indicates homology >75%, but lack approximately 25 kbp genetic information that encode also several antimicrobial resistance determinants. Sequence comparison and figure was created with Circoletto (Darzentas 2010).

Supplementary FIG 3 Characterisation of the biofilm formation capacity of Bone isolates. A) Growth of investigated Bone isolates on different agar media, blood, Mueller-Hinton, Mueller-Hinton fastidious and CLED agar plates. Bone8 and Bone7 isolates showed a small colony variant phenotype on Mueller-Hinton agar plates. B) Partial growth restoration of the small variant phenotype with hemin (factor X) and nicotinamide-adenine-dinucleotide (NAD; factor V) on Mueller Hinton broth after 24 h at 35 C. C) Biofilm formation on LB without salt agar plates. Number of *E. coli* Bone isolates as in Figure 1. Reference strains, 1C, *E. coli* TOB1 wild type control expressing curli fimbriae and cellulose; 2C, *E. coli* TOB2 (TOB1 Δ*csgD*), 3C, *E. coli* TOB2 (curli only expressing, TOB1 Δ*bcsA*), 4C, cellulose only expressing *E. coli* ECOR31 Δ*csgBA*. Growth for 72 h at 28 C and 37 C.

Supplementary FIG 4 Presence of genetic information for major biofilm determinants. A) curli fimbriae and poly-N-acetyl-glucosamine, B) the poly-N-acetyl-glucosamine gene cluster is at a different genomic location in strain Bone2; C) cellulose biosynthesis and D) type 1 fimbriae in Bone isolates, *E. coli* K-12 MG1655 reference strain and ST131 clade C2

strain JJ1886. Strain Bone8 contains a fragmented pseudogene rich poly-N-acetyl-glucosamine gene cluster. Strain Bone1B does not contain the type 1 fimbriae gene cluster. The comparison of the genetic information has been displayed with EasyFig (Sullivan, Petty et al. 2011).

Supplementary FIG 5 Heatmap showing the presence of the cyclic di-GMP turnover proteins in ST131 Bone isolates and ST131 reference strains. Green, presence of an intact protein potentially with single amino acid substitutions (see Supplementary Table 3); light green, shorter protein potentially functional; dark green, longer protein potentially functional; red, protein/gene deleted; dark red, truncated protein potentially functional; light red, internal stop potentially leading to a functionally altered protein. In green, GGDEF diguanylate cyclases synthesize cyclic di-GMP (dark green, GGDEF protein with catalytically inactive EAL domain); in green, EAL phosphodiesterases hydrolize cyclic di-GMP (dark green, EAL protein with catalytically inactive GGDEF domain). Evolved GGDEF (red) and EAL (red) domains without catalytic activity exist.

Supplementary FIG 6 Deletion of 668 nucleotides in all ST131 Bone strains and ST131 reference strains as compared to the *E. coli* K-12 MG1655 created a truncated YcgG cyclic di-GMP phosphodiesterase. The comparison of the genetic information has been displayed with EasyFig (Sullivan, Petty et al. 2011).

Supplementary FIG 7 Impairment of genes demonstrated to be causative for small colony variant growth (A) and to trigger a mutator phenotype (B). Green, gene functional; red, pseudogene. Impairment of genes in the pathway for menaquinone synthesis *menA* - *menH*, *lipA* involved in the synthesis of lipoic acid, *fabI* encoding an NADH dependent enoyl-acyl carrier protein reductase, for de novo thymidine synthesis *thyA* encoding thymidylate synthase, turnover of the alarmone second messenger ppGpp *spoT* and *relA* encoding a bifunctional synthase/hydrolase and synthase, respectively, *rpoB* encoding the beta subunit of the RNA polymerase and *can*, *cah* and *cynT* encoding different classes of carbonic anhydrases involved in the interconversion between CO_2 and its water soluble form have been reported to cause a small colony variant phenotype. Genes to cause a higher mutation rate leading to a so-called mutator phenotype are *mutS*, *mutL* and *mutH* encoding mismatch DNA repair complex proteins, *mutM*, *mutY* and *mutT* composing the base excision pathway, *recA* - *recD* mediating homologous recombination and repair, *uvrD* encoding a DNA helicase II, *dnaQ* encoding the epsilon subunit of DNA polymerase III which has 3'>5' proofreading function, *dam* encoding the orphan methylase of the GATC palindrome, *oxyR* encoding the regulator of antioxidant genes and *sodA* encoding a manganese dependent superoxide dismutase.

Supplementary FIG 8 Two deletions of >40 kbp occurred in ST131 strain Bone8 compared to strain Bone4. Deletion I encompases one of the phage mEp460 copies inserted into tRNA threonine. Deletion II comprises the core genome region between *lacY* and *iraP*. Figure was constructed with EasyFig (Sullivan, Petty et al. 2011).

Supplementary TABLE 1 Genomic and plasmid characteristics of Bone isolates

Supplementary TABLE 2 Phages integrated into the chromosome of Bone strains

Supplementary TABLE 3 Pseudogenes in Bone4, Bone8 and Bone7 isolates

Supplementary TABLE 4 Inventory of cyclic di-GMP turnover proteins in ST131 Bone isolates and ST131 reference strains.

References

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-410.

Assimacopoulos, A., B. Johnston, et al. (2012). "Post-prostate biopsy infection with Escherichia coli ST131 leading to epididymo-orchitis and meningitis caused by Gram-negative bacilli." J Clin Microbiol **50**(12): 4157-4159.

Berg, R. D. (1992). "Bacterial translocation from the gastrointestinal tract." J Med **23**(3-4): 217-244.

Brettin, T., J. J. Davis, et al. (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.

Cano, D. A., M. G. Pucciarelli, et al. (2003). "Selection of small-colony variants of Salmonella enterica serovar typhimurium in nonphagocytic eucaryotic cells." Infect Immun **71**(7): 3690-3698.

Carattoli, A., E. Zankari, et al. (2014). "In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing." Antimicrob Agents Chemother **58**(7): 3895-3903.

Cimdins, A., R. Simm, et al. (2017). "Alterations of c-di-GMP turnover proteins modulate semi-constitutive rdar biofilm formation in commensal and uropathogenic Escherichia coli." Microbiologyopen **6**(5).

Costerton, J. W., P. S. Stewart, et al. (1999). "Bacterial biofilms: a common cause of persistent infections." Science **284**(5418): 1318-1322.

Cui, L., H. M. Neoh, et al. (2012). "Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria." Proc Natl Acad Sci U S A **109**(25): E1647-1656.

Cummins, E. A., A. E. Snaith, et al. (2021). "The role of potentiating mutations in the evolution of pandemic Escherichia coli clones." Eur J Clin Microbiol Infect Dis.

Darzentas, N. (2010). "Circoletto: visualizing sequence similarity with Circos." Bioinformatics **26**(20): 2620-2621.

Denamur, E. and I. Matic (2006). "Evolution of mutation rates in bacteria." Mol Microbiol **60**(4): 820-827.

Drenkard, E. and F. M. Ausubel (2002). "Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation." Nature **416**(6882): 740-743.

Engelhardt, D. and E. I. Shakhnovich (2019). "Mutation rate variability as a driving force in adaptive evolution." Phys Rev E **99**(2-1): 022424.

Feng, Y., S. Zou, et al. (2021). "BacWGSTdb 2.0: a one-stop repository for bacterial whole-genome sequence typing and source tracking." Nucleic Acids Res **49**(D1): D644-D650.

Ginard, M., J. Lalucat, et al. (1997). "Genome organization of Pseudomonas stutzeri and resulting taxonomic and evolutionary considerations." Int J Syst Bacteriol **47**(1): 132-143.

Gouy, M., E. Tannier, et al. (2021). "Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation." Methods Mol Biol **2231**: 241-260.

Grant, J. R. and P. Stothard (2008). "The CGView Server: a comparative genomics tool for circular genomes." Nucleic Acids Res **36**(Web Server issue): W181-184.

Hammar, M., A. Arnqvist, et al. (1995). "Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12." Mol Microbiol **18**(4): 661-670.

Haussler, S., B. Tummler, et al. (1999). "Small-colony variants of Pseudomonas aeruginosa in cystic fibrosis." Clin Infect Dis **29**(3): 621-625.

Kahl, B. C., K. Becker, et al. (2016). "Clinical Significance and Pathogenesis of Staphylococcal Small Colony Variants in Persistent Infections." Clin Microbiol Rev **29**(2): 401-427.

Kresse, A. U., H. Blocker, et al. (2006). "ISPa20 advances the individual evolution of Pseudomonas aeruginosa clone C subclone C13 strains isolated from cystic fibrosis patients by insertional mutagenesis and genomic rearrangements." Arch Microbiol: 1-10.

Kresse, A. U., S. D. Dinesh, et al. (2003). "Impact of large chromosomal inversions on the adaptation and evolution of Pseudomonas aeruginosa chronically colonizing cystic fibrosis lungs." Mol Microbiol **47**(1): 145-158.

Kumar, S., G. Stecher, et al. (2018). "MEGA X: molecular evolutionary genetics analysis across computing platforms." Molecular biology and evolution **35**(6): 1547-1549.

Lazar, V., G. Pal Singh, et al. (2013). "Bacterial evolution of antibiotic hypersensitivity." Mol Syst Biol **9**: 700.

Matsumoto, T., M. Hashimoto, et al. (2020). "Molecular characterization of a carbon dioxide-dependent Escherichia coli small-colony variant isolated from blood cultures." Int J Med Microbiol **310**(5): 151431.

McGann, P., E. Snesrud, et al. (2015). "War wound treatment complications due to transfer of an IncN plasmid harboring bla(OXA-181) from Morganella morganii to CTX-M-27-producing sequence type 131 Escherichia coli." Antimicrob Agents Chemother **59**(6): 3556-3562.

Musher, D. M., R. E. Baughn, et al. (1979). "Selection of small-colony variants of Enterobacteriaceae by in vitro exposure to aminoglycosides: pathogenicity for experimental animals." J Infect Dis **140**(2): 209-214.

Nicolas-Chanoine, M. H., X. Bertrand, et al. (2014). "Escherichia coli ST131, an intriguing clonal group." Clin Microbiol Rev **27**(3): 543-574.

Ojha, R., A. A. Dittmar, et al. (2021). "Shigella flexneri Diguanylate Cyclases Regulate Virulence." J Bacteriol: JB0024221.

Okonechnikov, K., O. Golosova, et al. (2012). "Unipro UGENE: a unified bioinformatics toolkit." Bioinformatics **28**(8): 1166-1167.

Page, A. J., C. A. Cummins, et al. (2015). "Roary: rapid large-scale prokaryote pan genome analysis." Bioinformatics **31**(22): 3691-3693.

Park, Y. J., N. Le Phuong, et al. (2018). "Urinary tract infection caused by a small colony variant form of capnophilic Escherichia coli leading to misidentification and non-reactions in antimicrobial susceptibility tests." Antimicrob Resist Infect Control **7**: 139.

Pitout, J. D. D. and T. J. Finn (2020). "The evolutionary puzzle of Escherichia coli ST131." Infect Genet Evol **81**: 104265.

Pouillot, F., M. Chomton, et al. (2012). "Efficacy of bacteriophage therapy in experimental sepsis and meningitis caused by a clone O25b:H4-ST131 Escherichia coli strain producing CTX-M-15." Antimicrob Agents Chemother **56**(7): 3568-3575.

Pranting, M. and D. I. Andersson (2011). "Escape from growth restriction in small colony variants of Salmonella typhimurium by gene amplification and mutation." Mol Microbiol **79**(2): 305-315.

Ramiro, R. S., H. Costa, et al. (2016). "Macrophage adaptation leads to parallel evolution of genetically diverse Escherichia coli small-colony variants with increased fitness in vivo and antibiotic collateral sensitivity." Evol Appl **9**(8): 994-1004.

Roggenkamp, A., A. Sing, et al. (1998). "Chronic prosthetic hip infection caused by a small-colony variant of Escherichia coli." J Clin Microbiol **36**(9): 2530-2534.

Römling, U. and C. Balsalobre (2012). "Biofilm infections, their resilience to therapy and innovative treatment strategies." J Intern Med **272**(6): 541-561.

Römling, U., K. D. Schmidt, et al. (1997). "Large chromosomal inversions occur in Pseudomonas aeruginosa clone C strains isolated from cystic fibrosis patients." FEMS Microbiol Lett **150**(1): 149-156.

Römling, U., J. Wingender, et al. (1994). "A major Pseudomonas aeruginosa clone common to patients and aquatic habitats." Appl Environ Microbiol **60**(6): 1734-1738.

Santos, V. and I. Hirshfield (2016). "The Physiological and Molecular Characterization of a Small Colony Variant of Escherichia coli and Its Phenotypic Rescue." PLoS One **11**(6): e0157578.

Schembri, M. A., N. L. Zakour, et al. (2015). "Molecular Characterization of the Multidrug Resistant Escherichia coli ST131 Clone." Pathogens **4**(3): 422-430.

Schwartz, D. C. and C. R. Cantor (1984). "Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis." Cell **37**(1): 67-75.

Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." Bioinformatics **30**(14): 2068-2069.

Starkey, M., J. H. Hickman, et al. (2009). "Pseudomonas aeruginosa rugose small colony variants have adaptations likely to promote persistence in the cystic fibrosis lung." J Bacteriol.

Sullivan, M. J., N. K. Petty, et al. (2011). "Easyfig: a genome comparison visualizer." Bioinformatics **27**(7): 1009-1010.

Tappe, D., H. Claus, et al. (2006). "First case of febrile bacteremia due to a wild type and small-colony variant of Escherichia coli." Eur J Clin Microbiol Infect Dis **25**(1): 31-34.

Tatusova, T., M. DiCuccio, et al. (2016). "NCBI prokaryotic genome annotation pipeline." Nucleic Acids Res **44**(14): 6614-6624.

The, H. C., D. P. Thanh, et al. (2016). "The genomic signatures of Shigella evolution, adaptation and geographical spread." Nat Rev Microbiol **14**(4): 235-250.

Vallenet, D., A. Calteau, et al. (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**(D1): D579-D589.

Wang, X., J. F. Preston, 3rd, et al. (2004). "The pgaABCD locus of Escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation." J Bacteriol **186**(9): 2724-2734.

von Mentzer, A., T. R. Connor, et al. (2014). "Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution." Nat Genet **46**(12): 1321-1326.

Xia, H., Q. Tang, et al. (2017). "A yigP mutant strain is a small colony variant of E. coli and shows pleiotropic antibiotic resistance." Can J Microbiol **63**(12): 961-969.

Yegian, D., G. Gallo, et al. (1959). "Kanamycin resistant staphylococcus mutant requiring heme for growth." J Bacteriol **78**(1): 10-12.

Zlatkov, N. and B. E. Uhlin (2019). "Absence of Global Stress Regulation in Escherichia coli Promotes Pathoadaptation and Novel c-di-GMP-dependent Metabolic Capability." Sci Rep **9**(1): 2600.

Zogaj, X., M. Nimtz, et al. (2001). "The multicellular morphotypes of Salmonella typhimurium and Escherichia coli produce cellulose as the second component of the extracellular matrix." Mol Microbiol **39**(6): 1452-1463.

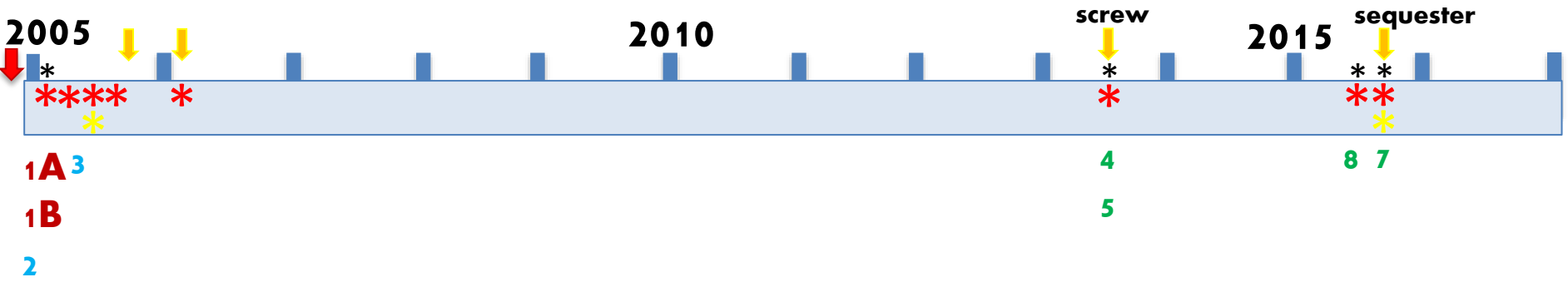


Figure 1

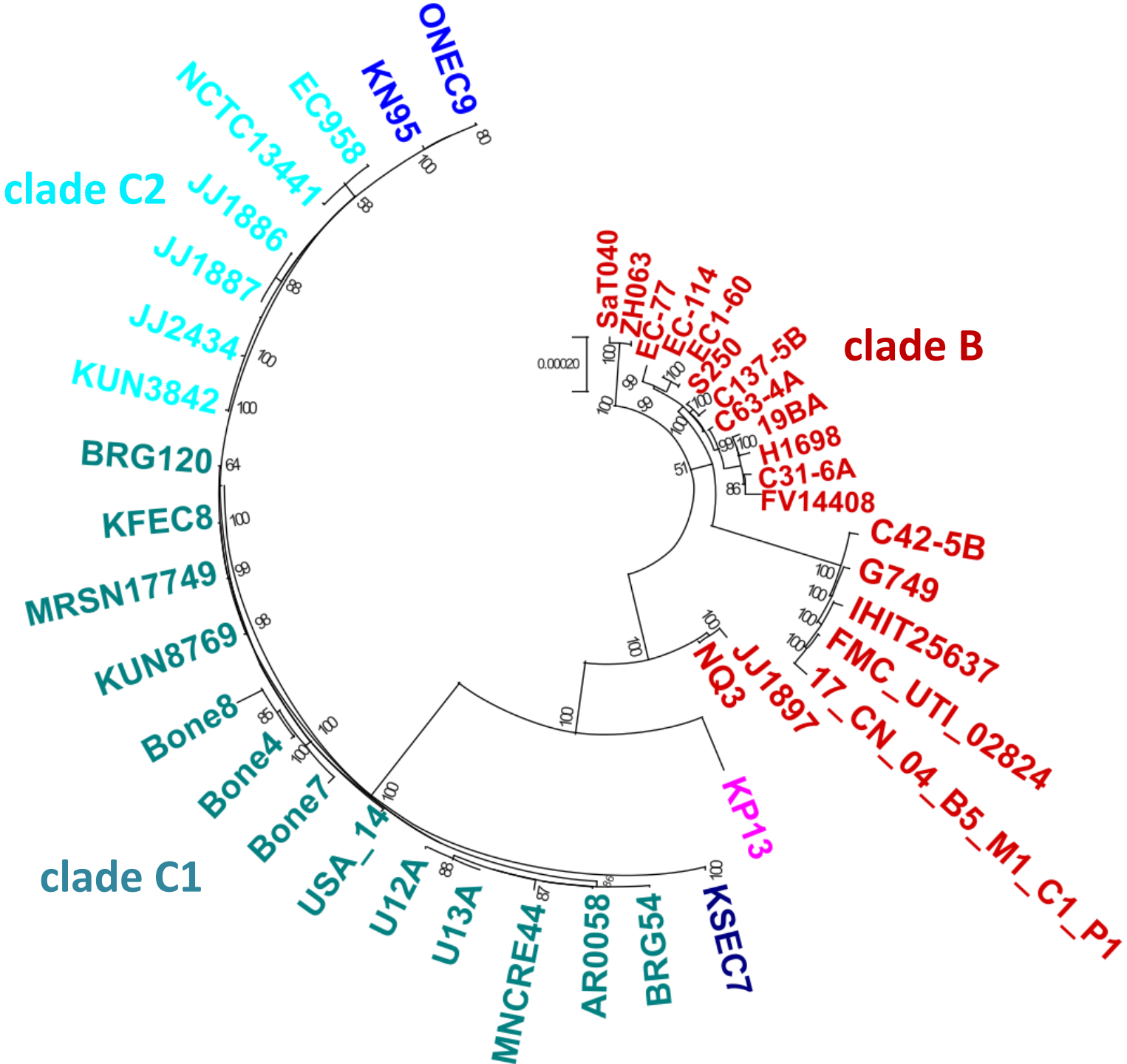

clade C2
clade C1
clade B
ONEC9
KN95
EC958
NCTC13441
JJ1886
JJ1887
JJ2434
KUN3842
BRG120
KFEC8
MRSN17749
KUN8769
Bone8
Bone4
Bone7
USA_14
U12A
U13A
MNCRE44
AR0058
BRG54
KSEC7
KP13
NQ3
JJ1897
17_CN_04_B5_M1_C1_P1
FMC_UTI_02824
IHIT25637
G749
C42-5B
FV14408
C31-6A
H1698
19BA
C63-4A
C137-5B
S250
EC1-60
EC-114
EC-77
ZH063
SaT040
0.00020

Figure 2A

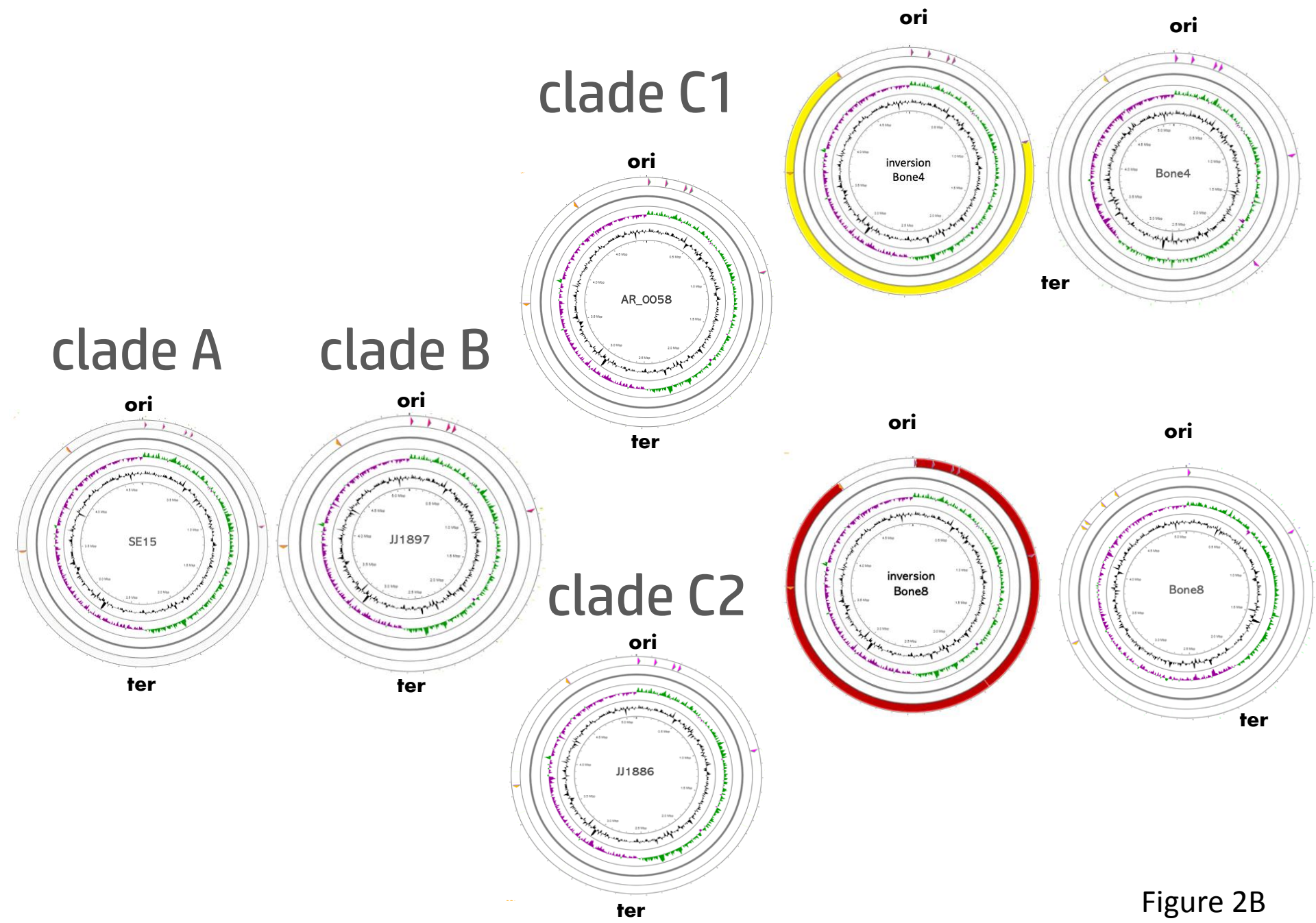

clade A
ori
SE15
ter
clade B
ori
JJ1897
ter
clade C1
ori
AR_0058
ter
ori
inversion
Bone4
ori
Bone4
ter
clade C2
ori
JJ1886
ter
ori
inversion
Bone8
ori
Bone8
ter

Figure 2B

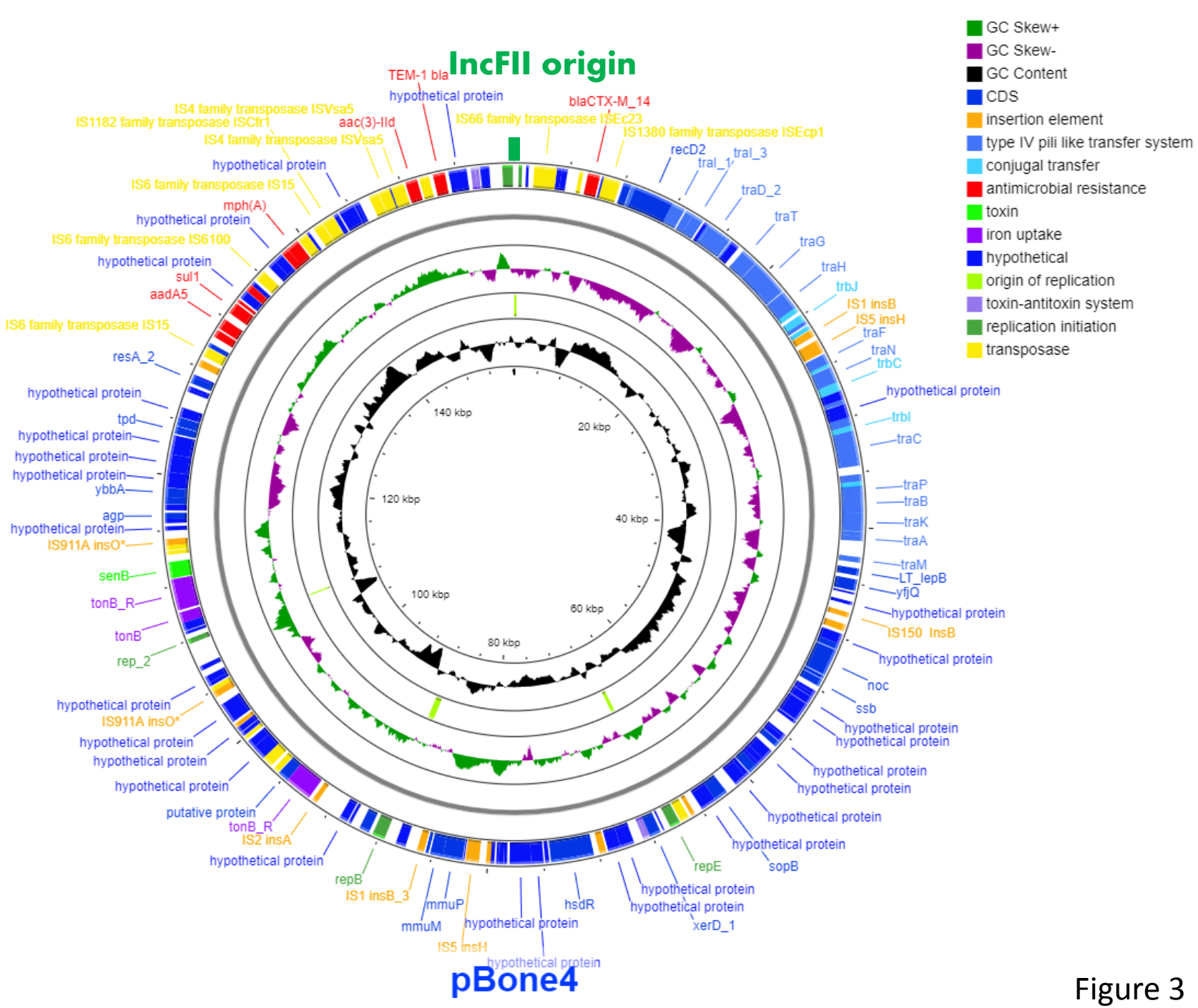

IncFII origin
pBone4
GC Skew+
GC Skew-
GC Content
CDS
insertion element
type IV pili like transfer system
conjugal transfer
antimicrobial resistance
toxin
iron uptake
hypothetical
origin of replication
toxin-antitoxin system
replication initiation
transposase
20 kbp
40 kbp
60 kbp
80 kbp
100 kbp
120 kbp
140 kbp
TEM-1 bla
hypothetical protein
blaCTX-M_14
IS66 family transposase ISEc23
IS1380 family transposase ISEcp1
recD2
traI_3
traI_1
traD_2
traT
traG
traH
trbJ
IS1 insB
IS5 insH
traF
traN
trbC
hypothetical protein
trbI
traC
traP
traB
traK
traA
traM
LT_lepB
yfjQ
hypothetical protein
IS150 InsB
hypothetical protein
noc
ssb
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
sopB
repE
hypothetical protein
hypothetical protein
xerD_1
hsdR
hypothetical protein
hypothetical protein
mmuP
mmuM
IS5 insH
repB
IS1 insB_3
hypothetical protein
IS2 insA
tonB_R
putative protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
IS911A insO*
hypothetical protein
rep_2
tonB
tonB_R
senB
IS911A insO*
hypothetical protein
agp
hypothetical protein
ybbA
hypothetical protein
hypothetical protein
tpd
hypothetical protein
resA_2
IS6 family transposase IS15
aadA5
sul1
hypothetical protein
IS6 family transposase IS6100
hypothetical protein
mph(A)
hypothetical protein
IS6 family transposase IS15
hypothetical protein
IS4 family transposase ISVsa5
aac(3)-IId
IS4 family transposase ISVsa5
IS1182 family transposase ISCfr1

Figure 3

Pseudogenes

STO Bone4 (*n*=1141)
STO Bone7 (*n*=1231)
STO Bone8 (*n*=2287)

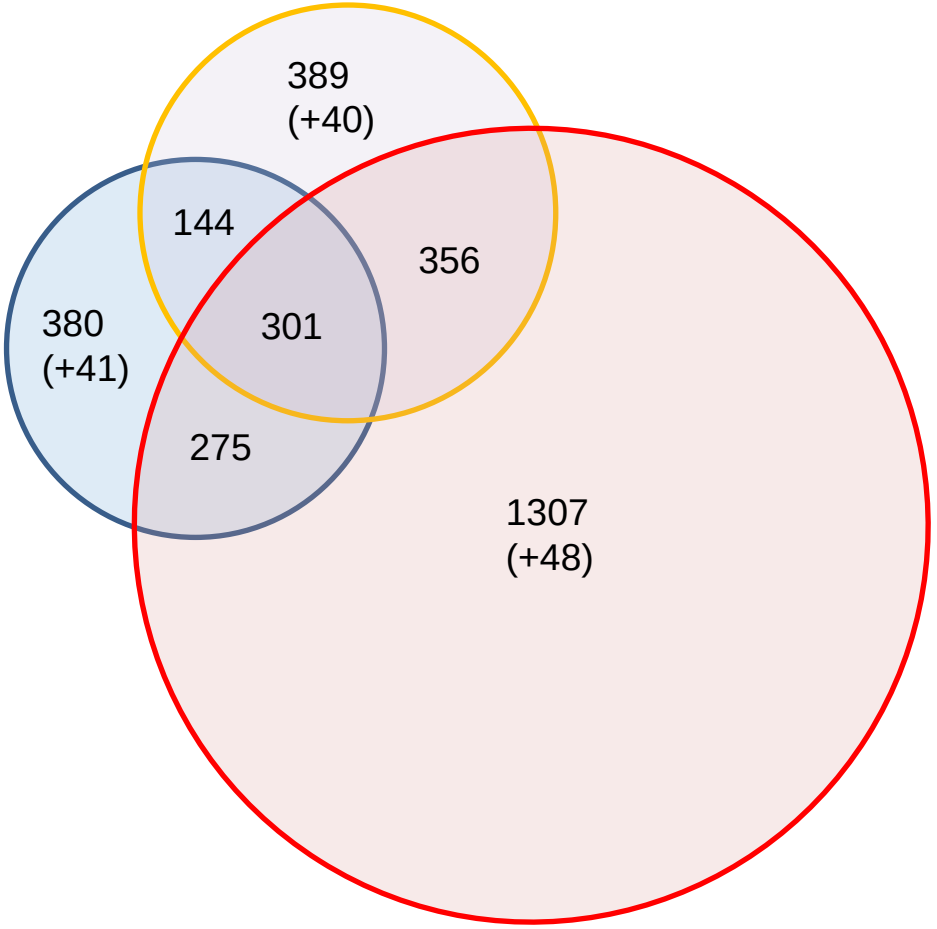
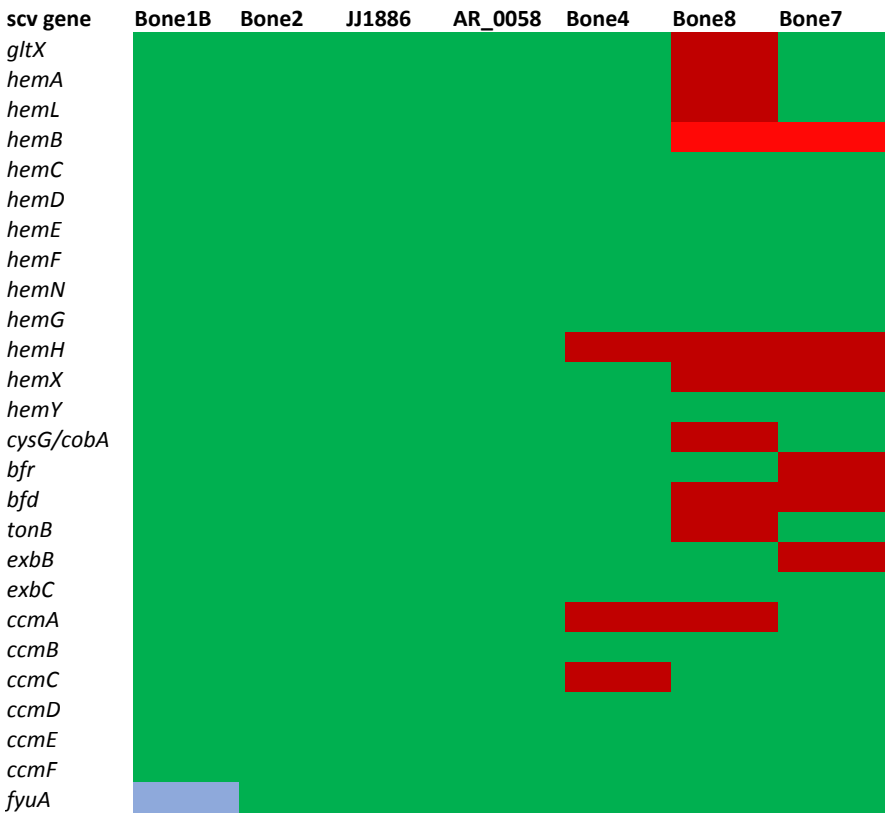


Figure 4

A


scv gene
Bone1B
Bone2
JJ1886
AR_0058
Bone4
Bone8
Bone7
gltX
hemA
hemL
hemB
hemC
hemD
hemE
hemF
hemN
hemG
hemH
hemX
hemY
cysG/cobA
bfr
bfd
tonB
exbB
exbC
ccmA
ccmB
ccmC
ccmD
ccmE
ccmF
fyuA

B

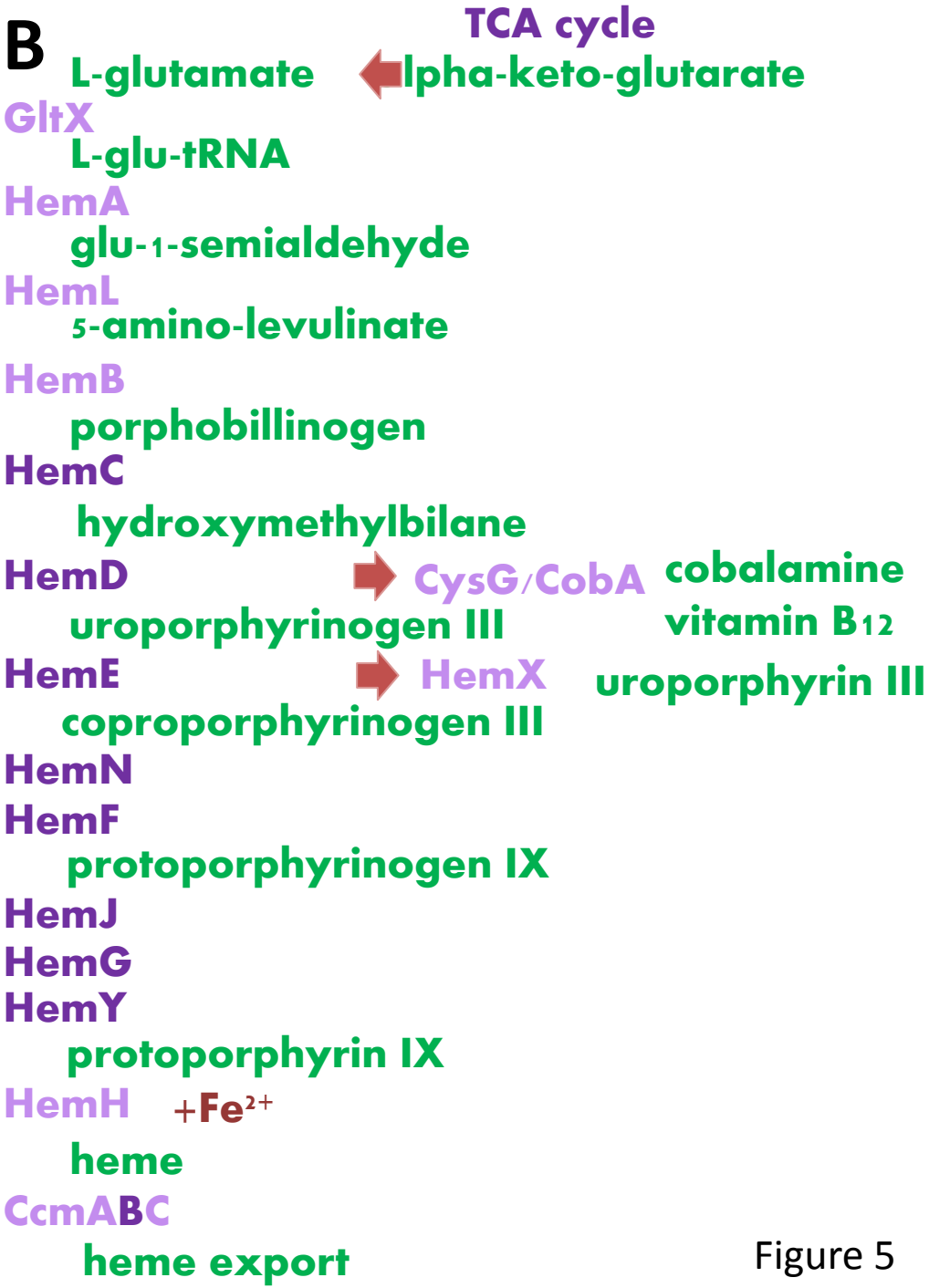

TCA cycle
L-glutamate
lpha-keto-glutarate
GltX
L-glu-tRNA
HemA
glu-1-semialdehyde
HemL
5-amino-levulinate
HemB
porphobillinogen
HemC
hydroxymethylbilane
HemD
CysG/CobA
cobalamine
vitamin B12
uroporphyrinogen III
HemE
HemX
uroporphyrin III
coproporphyrinogen III
HemN
HemF
protoporphyrinogen IX
HemJ
HemG
HemY
protoporphyrin IX
HemH
+Fe2+
heme
CcmABC
heme export

Figure 5