OTKA-NKFIH K 105635

PROJECT CLOSING REPORT

Molecular genetics and ecology of Salmonella genomic island 1 (SGI1): secrets of mobility, spread and pathogenetic significance

2013-2017

INTRODUCTION

The rapid evolution of bacteria due to their outstanding genome plasticity represents a serious challenge for researchers as well as for actors in agriculture, human and animal healthcare. Genomic islands, plasmids and other mobile genetic elements are key factors in spreading the determinants of pathogenicity and antibiotic resistance (AR), thereby assisting rapid adaptation and evolution of prokaryotes. Combating the antibiotic resistant pathogenic bacteria began nearly after the discovery of antibiotics. Nowadays, the continuous battle against the emerging AR pathogens demands the use of new generation antibiotics in human medicine as well as in livestock-farming, which entails significant food safety hazards for humans. Hence, reducing these risks is an important EU priority, as well.

AR genes are generally vehicled by mobile genetic elements such as conjugative plasmids and genomic islands (GI), which ensure rapid adaptation of bacteria to the strong selective pressure due to excessive use of antimicrobials and the fast worldwide spread of multidrug resistance (MDR). The conjugative and mobilizable GIs have a key role in rapid dissemination of MDR in *Salmonella, Proteus* and other Gram-negative pathogenic bacteria. *Salmonellae*, particularly the serovars Typhimurium and Enteritidis, are among the most prevalent zoonotic pathogens worldwide.

One of the best examples of such GIs is the 42.4-kb chromosomal gene cluster named *Salmonella* genomic island 1 (SGI1) that was identified in MDR clones (resistant for ampicillin, chloramphenicol/florfenicol, streptomycin/spectinomycin, sulphonamides and tetracycline - ACSSuT phenotype) of *Salmonella* Typhimurium DT104 (*S.* T. DT104) in early 1990s and became prevalent globally in that decade. SGI1 and its variants [1], [2] [3], [4] [5], [6], [7], [8] are now widely distributed among numerous *Salmonella* serovars, *Proteus mirabilis* strains and several other Gram-negative pathogens, posing significant challenges for clinicians in human and animal healthcare. SGI1 proved to be an integrative mobilizable element (IME) [9] that does not encode a complete gene set of the conjugation apparatus and is efficiently and exclusively mobilized by the IncA/C-family conjugative plasmids [6], [10], [11], which also carry many different AR genes and are widely distributed among Gram-negative bacteria [12].

AIMS OF THE PROJECT

Better understanding the different aspects of spreading the mobilizable genomic islands such as SGI1 and its interaction with the IncA/C family conjugative plasmids may lead to improvements of preventive measures against the accelerating distribution of MDR pathogens.The SGI1 – IncA/C system appears to be a good model to study the role of genomic islands in dissemination of AR. Therefore, this project aimed to study

- the mechanisms of SGI1 transfer at the molecular level,
- the effects of SGI1 on the fitness of host bacteria and
- the *in vivo* horizontal transfer of the island.

WP1 focused on the molecular mechanisms of SGI1 transfer including the contribution of the IncA/C helper plasmids to its mobilization. Particular attention was payed to the identification of the mobility functions of SGI1, functions of the helper plasmids and the type of interactions between the island and the plasmid. The other two WPs mostly studied the "ecology" of SGI1. In WP2 the role of SGI1 in fitness and virulence of *S.* T. DT104 strains and in pathogenesis of *Salmonella* infections was studied, as it may have a great relevance to the worldwide spread of the strains in different hosts and to predictions on future chances for preventive intervention methods. In WP3 a specific *in vivo* experimental system was established to follow the transfer of SGI1 in the gut flora of a living animal among natural conditions.

RESULTS AND DISCUSSION

WP 1.

At the beginning of the project limited information was available on the molecular background of SGI1 mobilization and on the role of transfer system of IncA/C helper plasmids for this chromosomal gene cluster It was known that SGI1 integrates site-specifically into the bacterial chromosome at the end of *thd*F gene, from where it can also be excised by the lambda-family integrase and excisionase (Int, Xis) encoded near to the 5' end of the island and that the excised circular form of SGI1 is mobilizable by the IncA/C-family plasmids [9]. Furthermore, at that time less than 10 members of this plasmid family had been sequenced and their conjugative transfer system had not been analyzed at all. Thus, our first goal was to identify and characterize the principal *cis* elements and self-encoded genes required for SGI1 transfer, to identify the unknown elements of the conjugative system of IncA/C plasmids and to understand the mechanisms of interaction between the island and the helper plasmid.

Identification and analysis of the self-encoded factors involved in conjugative transfer of SGI1.

Localization and functional studies of the transfer origin of SGI1.

First subtopic of WP 1 was the identification of the transfer origin (*oriT*) of SGI1 and the possible self-encoded proteins required for the transfer. We have previously developed an F plasmid-based single copy trapping system to capture SGI1 on a plasmid [6] in order to facilitate genetic manipulations. SGI1 has been entrapped in this plasmid, and then cloned into a p15A-based vector. Consecutive deletions have been generated in the SGI1 sequence and each derivatives were tested for conjugation in mating assays in the presence of an IncA/C helper plasmid. A collection of more than 30 SGI1-derived segments was established and tested by this method. Based on these results, the conjugative functions have been localized in a 2.2 kb fragment (16447-18680 bp of SGI1) called mob_{SGI1} containing 4 orfs (S019-S022) with unknown functions and the *oriT* sequence. Further shortening of this fragment blocked the conjugation, indicating that it carries indispensable genes as well as the *oriT*. Thus more exact identification of *oriT* was carried out in a complementing setup: the mob_{SGI1} fragment has been integrated into the chromosome of the *E. coli* host strain by a transposon-based vehicle. The mob_{SGI1} region was inserted into a

mini-Tn*10* transposon located in an R6K-based mobilizable replicon, which can express Tn*10* transposase. Using this suicide plasmid, the miniTn*10*:: mob_{SGI1}-Km^R unit was integrated into the chromosome of TG1Nal strain, which was then used as donor strains in further matings. After progressive shortening the plasmid-born copy of mob_{SGI1} region, the chromosomal insertion complemented the missing functions, thus the *oriT*, the only *cis* element of the conjugation system of SGI1 could be determined as a 124-bp segment located in the overlapping region of orfs S021 and S022 (Fig. 1). Unlike most *oriT*s identified to date, $oriT_{SGI1}$ is a GC-rich sequence, which, similarly to other *oriT*s, carries several inverse repeat (IR) motifs.

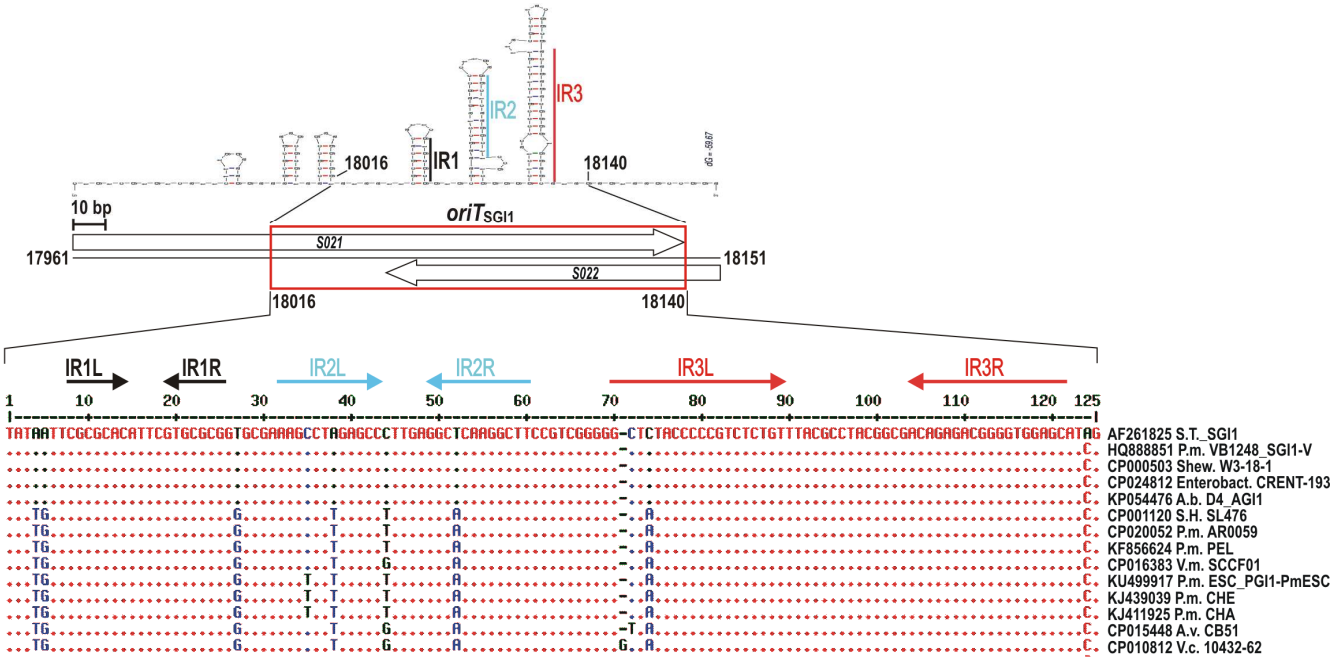

Fig. 1. Localization, structure and comparative analysis of $oriT_{SGI1}$. The location of fully functional *oriT* is indicated by red box in the schematic map. Coordinates of the map are according to the published SGI1 sequence from *S.* T. DT104 (AF261825). The potential secondary structure of the region (generated by mFold webserver [13]) and the stem-forming IRs are shown above the map. The repeated motifs are shown above the sequence by colour coded arrows. The left and right arms of the inverted repeats longer than 6 bp are designated as IR1/2/3L and R, respectively. The alignment shows the comparison of the homologous sequences to $oriT_{SGI1}$ in 54 fully sequenced SGI1-related elements found in the GenBank database. Only the 13 *oriT* sequences having divergent positions are shown, while those identical to the reference sequence are represented by the respective region of the published SGI1 sequence AF261825. The short names of bacterial strains and the SGI1 variant is indicated after the GenBank acc. number of the sequences. Abbreviations are as follows: *A.b.* – *Acinetobacter baumannii*; *A.v.* – *Aeromonas veronii*; *Enterobact.* - *Enterobacter* sp.; *P.m.* – *Proteus mirabilis*; *Shew.* – *Shewanella* sp.; *S.*T. – *Salmonella* Typhimurium; *S.*H. – *Salmonella* Heidelberg; *V.c.* – *Vibrio cholerae*; *V.m.* – *Vibrio mimicus.*

Analysis of deletion and site-directed mutations showed that 2 inverse repeats and a 7-bp AT rich tract located at the 5' end of *oriT* sequence have important role in *oriT* function. Elimination of the first repeat (IR1) has the most deleterious effect as the transfer of this *oriT* mutant was undetectable. Thus, the initiation site of transfer (nic site) is probably located in or near to the first arm of IR1. Many efforts have been made to identify nic site more exactly. Since biochemical methods (primer extension and RACE PCR) failed to give results, a biological approach (interrupted mating [14]) has also been applied. $OriT_{SGI1}$ flanked by different resistance marker genes was inserted into a large, F-factor-based plasmid lacking its own transfer capacity. Mobilization of this construct by an IncA/C helper plasmid was then tested in very short mating periods (2-15 min). Although the transfer frequencies dropped to the detection limit below 5 min mating time, partially transferred (deletion) derivatives, of which sequencing would have revealed the start point of plasmid transfer, could not be isolated to date. Thus, the identification of nic site and characterization of the initiation complex may be carried out in an in vitro system, in which purified protein components and *oriT* template DNA are mixed and analysed.

OriT$_{SGI1}$ was knocked out from SGI1-C variant using the one-step gene inactivation method [15], which proved that this region is indispensable for SGI1 transfer, on the other hand, an otherwise nonmobilizable plasmid vector containing *oriT*$_{SGI1}$ could be well transferred in the presence of SGI1-C$^{\Delta oriT}$ and a helper plasmid, indicating that S021 and S022 orfs (both eliminated by Δ*oriT* mutation) are not required for conjugation. Comparative sequence analysis of *oriT*$_{SGI1}$ and the whole mob$_{SGI1}$ region showed that both are well conserved among SGI1 variants and more distantly related GI elements, supporting that we have identified the hitherto unknown mobilization unit of SGI1.

Identification of SGI1 genes required for conjugal transfer of SGI1.

The mob$_{SGI1}$ region responsible for SGI1 transfer harbours 4 orfs, S019-S022. S019 and S020 appears to be parts of a bicistronic operon, S022 locates upstream of S020 and is encoded in the same strand, while is encoded in the complementary strand and partially overlaps S022. The phenotype of SGI1-C$^{\Delta oriT}$ mutant proved that S021 and S022 orfs, if they are expressed at all, encode proteins, which are not required for *trans* mobilization of *oriT*$_{SGI1}$. Contrarily, KO mutants of S020 and S019 failed to conjugate, but both could be rescued by providing the entire mob$_{SGI1}$ segment or even the respective genes placed under the control of the *tac* promoter in a plasmid. These results proved that S020 and S019 encodes for proteins that are both required for SGI1 transfer. Since the orfs are separated only by a single codon, the biologically active protein might be translated as a fusion by ribosomal read-through. To test this hypothesis, we performed a trans-complementation assay where the chromosomal S020 KO mutant SGI1 was complemented with S019 KO plasmid construct expressing only the S020 protein and vice versa. Both combinations could rescue the SGI1 KO mutants, which verified that orfs S019 and S020 are translated into independent polypeptides.

The genetic context suggested that S019 and S020 constitute a bicistronic operon. To localize the promoter region of this putative operon, the upstream sequence extending from the start codon of *S020* to the end of *S023* or *S022* was fused to a promoterless *lacZ* gene in β-galactosidase tester plasmids and their promoter activity was measured in β-galactosidase assay [16]. The results showed that a very weak promoter drives the expression of the operon, which was neither detectable by primer extension assay. Thus their expression was analysed and the promoters were localized using a complementation assay, in which the transfer rate of the chromosomal S019 and S020 KO mutant SGI1 was examined in the presence of the IncA/C helper plasmid and progressively reduced fragments of the mob$_{SGI1}$ region provided on a complementing plasmid. By this method a promoter region with at least two potential promoters could be identified upstream of S020 (localized in intergenic region between orfs S020 and S021), which proved to drive the expression of both S020 and S019. Interestingly, S019 can also be expressed independently of S020 from a weak promoter located in S020 near the 3' end of the gene. The fact that both KO mutants could be complemented (although to a lower level than the wt transfer rate) by constructs where the orfs were cloned without their promoter regions (the coding sequences preceded by transcriptional terminators) supported again that these orfs are expressed at a very low levels. Database searches with the predicted protein sequences showed that orf *S020* encodes a protein belonging to the Tyr-recombinase/integrase superfamily of DNA breaking-rejoining enzymes, while S019 protein contains a phage integrase N-terminal SAM-4 like domain, which might be involved in protein-protein interactions (manuscript in preparation).

For further molecular analyses of the functions of these proteins and their interactions with the *oriT*$_{SGI1}$ sequence, an in vitro system was established. We have developed affinity-chromatography-based purification protocols for S019, S020 and the relaxase of the helper plasmid, the three potential components of the initiation complex (relaxosome) of SGI1. The interactions of the proteins with *oriT*$_{SGI1}$ are examined in electrophoretic mobility shift assays (EMSA) using the purified proteins alone or mixed in different combinations and concentrations. The other approach applied is the nicking assay where the DNA nicking activity of the pure protein preparates or protein mixes are examined on plasmids harbouring *oriT*$_{SGI1}$. Although both methods are in the optimization phase at present, they hopefully will help to uncover the molecular aspects of the transfer initiation of SGI1.

Studies on the IncA/C plasmids, the mobilization helpers of SGI1.

Identification and analysis of oriT of the helper plasmid.

Although the IncA/C relaxase was shown to be required for the efficient transfer of both the plasmid and the island, *oriT*$_{SGI1}$-like sequence could not be found on the plasmid, suggesting that *oriT* and initiation complex of the plasmid differs from that of SGI1. Biochemical characterization of the initiation complex of IncA/C plasmids has not yet been reported and the exact position of *oriT*$_{A/C}$ sequence has not been determined, either. For better understanding this transfer system, we decided to identify the *oriT* region of the helper plasmid. The IncA/C plasmid, R55, was shotgun cloned in a non-conjugative vector and the 9 different libraries were transformed into E. coli cells harbouring intact R55. The transformants were then used as donors in mating experiments to isolate mobilizable clones containing the potential *oriT*. Unexpectedly, two different regions were obtained. By progressively shortening of these segments, the minimal mobilizable fragments were determined (Fig. 2).

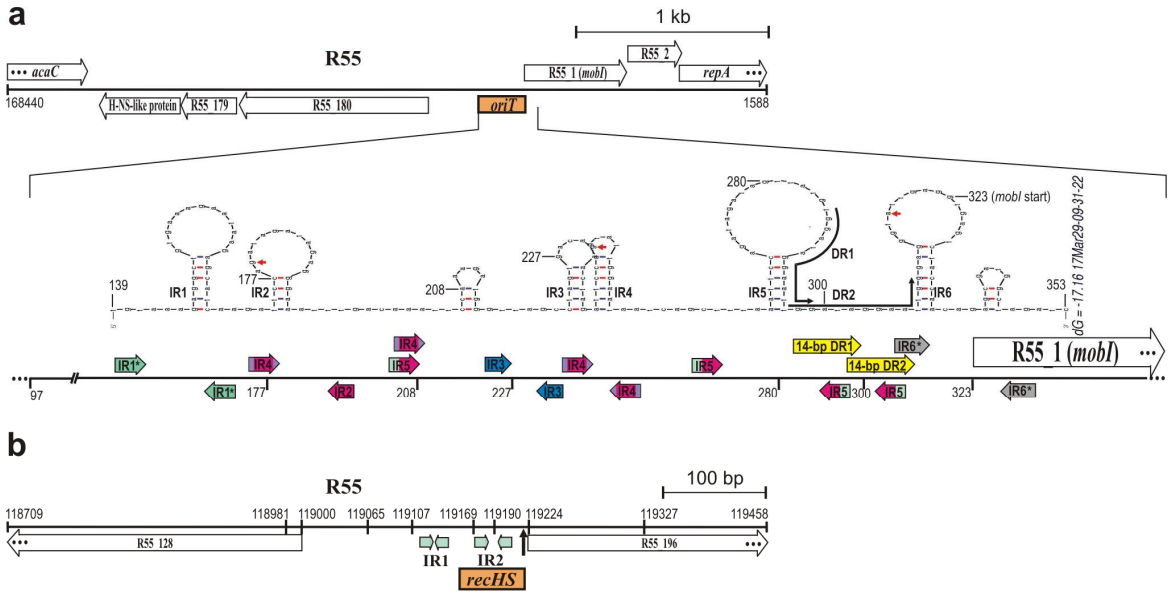

Fig. 2. Mobilizable fragments on R55. The schematic maps representing the regions of R55 covered by the subclones listed below are drawn to scale, and the coordinates are shown according to the published R55 sequence. Open arrows indicate the annotated ORFs. (**a**) Mob 1 region including *oriT*. Colour-coded arrows represent inversely (IR) and directly (DR) repeated sequence motifs of at least 4 bp in length (IR1: 6 bp with 1 mismatch, IR2: 4 bp, IR3: 5 bp, IR4: 6 bp, IR5: 6 bp and IR6: 7 bp with 1 mismatch). Asterisks in IR1 and IR6 refer to the imperfect repeat. The red region in IR4 and IR5 represents the 4-bp IR2 motif as a part of these repeats. The potential secondary structure of the region is shown. Coordinates are indicated according to the published R55 sequence (GenBank acc. no.: JQ010984). The red arrows point to the base positions where IncA/C family plasmids most frequently carry divergent bases. (**b**) Localization of the recombination hot spot (RecHS) in the Mob 2 region. The 11-bp IR1 and 14-bp IR2 inverse repeat motifs are shown as light green arrows. Vertical arrows indicate the positions of ARI$_{R16a}$ and Tn*6333* insertions in the related R16a and IP40a plasmids.

These regions showed no similarity in their sequence or their secondary structure (stem-loops, direct repeats) neither to each other nor to SGI1 *oriT*. To further analyse the function of the potential *oriT* sequences, KO mutants were generated on the close relatives of R55, the IncA/C plasmids R16a and IP40a, which have fewer resistance markers (KmR, ApR) that made easier to carry out site-directed gene KO. The conjugation tests of these mutants indicated that the *oriT* is located near to the *repA* gene, but far from the relaxase gene where $oriT_{A/C}$ was previously predicted. We identified the $oriT_{A/C}$ as a ca. 150 bp AT-rich sequence containing multiple short inverted and a 14-bp partially overlapping direct repeat motifs (Fig. 2a). The core domain of $oriT_{A/C}$ was confined to a 70-bp segment containing only two inverted repeats and one copy of the 14-bp direct repeat [17]. We have shown that *oriT* sequence covers the promoter region of *mobI* gene encoding an indispensable transfer factor of the plasmid [18]. The promoter of *mobI* has been identified by primer extension and shown to be constitutive promoter that drives the synthesis of *mobI* mRNA through the *oriT* region. Consequently, $oriT_{A/C}$ KO mutation also eliminated the *mobI* gene expression. This was proved by complementation tests where *mobI* gene was expressed in trans. We also showed that IncA/C plasmids have a single *oriT* locus and excluded the existence of a second one predicted previously [18].

The other mobilizable R55 region that appeared to be a second *oriT* in our experimental setup was found to be a recombination site, instead to be a real *oriT* (Fig. 2b). KO of this region or the divergent orf positioned similarly to *mobI* and *oriT* had no significant effect on the conjugation. Extensive complementation and PCR tests suggested that this region contains a recombination hot spot which could be responsible for high frequency transfer of the non-conjugative plasmid containing it. The transfer occurred by recombinational integration into the intact helper plasmid. The recombinogenic segment was identified as an 56-bp region including a 14-bp imperfect inverted repeat motif (RecHS) near the *rhs* gene in the conserved IncA/C backbone. We have shown that the recombination activity of this segment is independent of the host-encoded RecA and requires a plasmid-encoded recombinase [17]. Although the cognate recombinase of RecHS has not yet been identified, we suggest that it acts as a recombination site of a site-specific dimer resolution system. The comparative analyses of both $oriT_{A/C}$ and the RecHS sequences among IncA/C family plasmids proved that they are highly conserved parts of the IncA/C backbone. To date, more than 150 sequenced IncA/C family members can be found in public databases. Five of them lack $oriT_{A/C}$ due to large deletions spanning the whole region, including the *mobI* gene and 14 plasmids lack the RecHS region, however the vast majority of the family members preserved both with very low sequence divergences indicating the selective pressure towards retaining these functions.

Verification that traI of IncA/C plasmids encodes for the relaxase.

The conjugative transfer system of IncA/C plasmids has been classified into the MOB_H family [19] based on the protein sequence of the putative relaxase gene *traI*. However, the transfer genes including *traI* were mostly identified by their homology to characterized *tra* genes of other systems without experimental analyses. To prove that *traI* encodes the relaxase, the key enzyme in the transfer of the plasmid and also SGI1, the gene has been deleted from R16a helper plasmid and cloned into an expression vector. The Δ*traI* mutant plasmid lost its ability both for self-transfer and for mobilization of SGI1, however the complementation tests showed that the cloned relaxase gene can rescue both functions. More sensitive tests indicated that although the Δ*traI* mutation completely

erased the conjugation of the helper plasmid, a residual SGI1 transfer could be observed. This suggested that SGI1 transfer is not fully relaxase-dependent, however the relaxase increases its efficiency with orders of magnitude.

The Δ*traI* mutant was also applied in seeking $oriT_{A/C}$ and excluding that *oriT* is located near *traI* [17] as it was found in several unrelated conjugation systems.

Genomics of IncA/C family plasmids – establishing the third generation sequencing platform MinION.

For better understanding the SGI1-helper system, we investigated the characteristics of IncA/C plasmids by genetics and also genomics approaches. We applied in the experiments three IncA/C plasmids that were proved to be equivalent in mobilization of SGI1, but differed in their antibiotic resistance spectrum. Due to its multiple resistance (Km, Gm, Ap, Cm, Flo, Sul, Hg), R55 was, for example, poorly suitable for genetic manipulations (KO experiments), thus in these cases two related KmR, ApR plasmids, R16a and IP40a, were applied. During these experiments we observed the loss of one or both resistance markers of the plasmids. All the three plasmids were isolated in the early antibiotic era (early 1960s), but the sequence was available only for R55. To understand this phenomenon, we decided to determine the sequence of R16a and IP40a plasmids. First, both were sequenced on MySeq platform and the reads were de novo assembled by Mira software. The resulting contigs proved that the plasmids are closely related to each other and to R55, however both contain very complex regions composed of several different types of transposons (Tn) inserted into each other. This made almost impossible to assemble the intact sequence. Our group have participated in the test program of a third generation sequencing device MinION developed by Oxford Nanopore Ltd. MinION is a portable, single molecule sequencing device. As this instrument directly senses individual DNA fragments passing through a protein nanopore, it is able to read extremely long fragments (50-75 kb). The very long reads are important when assembling genomes as they span repetitive elements. We sequenced the two plasmids on MinION as well and the resulting long reads facilitated the exact assembly of MySeq contigs and establishing the ungapped sequence of two new IncA/C plasmids (both are >170 kb) (Fig. 3).

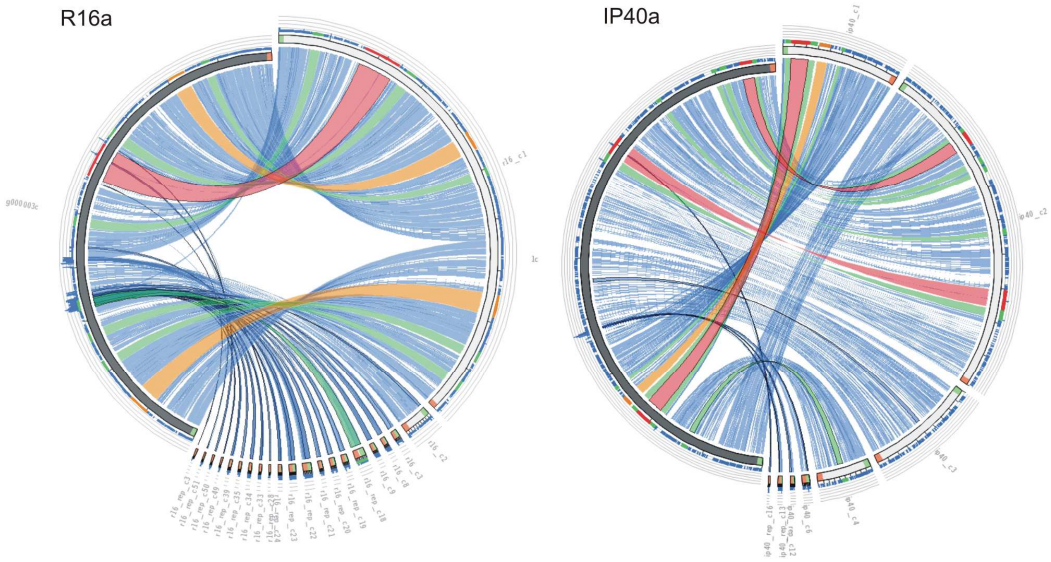

Fig. 3. Alignments of MIRA-contigs (assembled from Illumina reads) to the single Miniasm-contigs assembled from MinION reads obtained from R16a and IP40a. The Miniasm-contigs (g00003c and g00001c) are shown as dark grey, Mira-contigs are blank (available at http://emboss.abc.hu/minionarticle/). Ribbons represent the local alignments produced by BLAST. Colors correspond to the alignment bit scores in the four quartiles: red, 75-100%; orange, 50-75%; green, 25-50%; blue, <25% of the maximum bit score. Green-to-red direction of ideograms shows the 5'-3' orientation of contigs. Ribbons are inverted if reverse complementary sequence are aligned.

Comparative analysis based on our new plasmid sequences indicated that R16a and IP40a, unlike R55, which is a Type 2 plasmid, belong to the A/C2 Type 1 group and both contain unique antibiotic resistance islands (ARI) and a complete GI*sul2* island, which was not previously found in the family (Fig. 4ab). ARI of R16a consists of Tn*1*, Tn*6020*, and the newly identified Tn*6333* (Fig. 4c) harbouring the resistance genes *blaTEM-1D*, *aphA1b* and a mercury resistance module, respectively, a truncated Tn*5393* copy and a gene cluster with unknown function, while ARI of IP40a carrying *blaTEM-1D* and *aphA1b* genes is composed of Tn*1* with Tn*6023* insertion in it. Additionally, IP40a harbours single IS*2*, IS*186* and Tn*1000* insertions in the backbone, an IS*150* copy in GI*sul2* and a complete Tn*6333* at the position of ARI of R16a [12]. Interesting fact is that these plasmids carry the resistance genes as parts of composite or Tn*3*-family transposons instead of integrons. The exceptional presence of a complete GI*sul2* island in R16a and IP40a, which was suggested to be the ancient form of ARI-B islands [20] found in numerous IncA/C family plasmids suggest that they represent ancestral forms in Type 1 lineage and the early stages of IncA/C evolution. On the other hand, the analysis proved that the key transfer genes of the three plasmids are very conservative, which supported their interchangeability in our experimental setups.

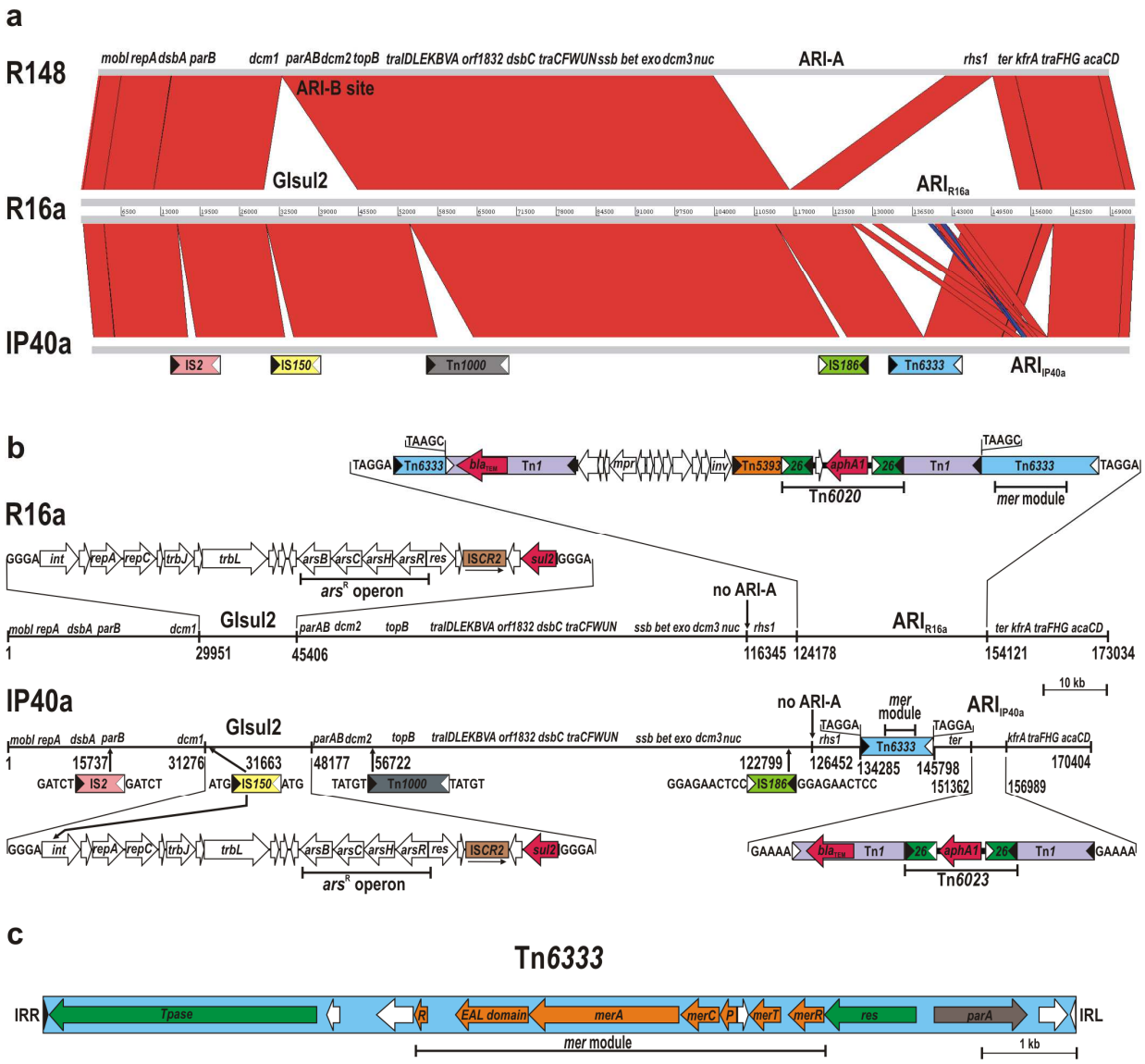

Fig. 4. (**a**) Comparison of R16a and IP40a to reference IncA/C_2 Type 1 plasmid R148. Major backbone genes and the site of ARIs are indicated. The intact transposons and IS elements located outside of ARI_{IP40a} are shown below the graph (**b**) Detailed structure of ARIs in R16a and IP40a. The horizontal lines with the major backbone genes represent the whole plasmid sequences. Coordinates below the maps show the ends of ARIs and the standard position of ARI-A. Orfs of ARIs encoding >50 amino acids are shown by arrows. Transposons and IS elements are represented by colour-coded rectangles with white (IRL) and black (IRR) arrowheads indicating the inverted repeats of the element. Direct repeats bracketing the mobile elements are shown in capitals. IS*26* variants are marked with '26'. Antibiotic resistance genes are marked as red, the IS*CR2* element is light brown, metal-resistance operons, compound transposons and

prophages are indicated. Insertion site of *E. coli*-related IS elements and Tn*1000* are indicated below the graph of IP40a. Figures are drawn not to scale. (**c**) Schematic representation of Tn*6333*. Orfs are colour coded, depending on functional annotations: green, transposition/recombination; orange, mer resistance, grey, partitioning; white, unknown function. EAL (glutamate-alanine-leucine) domain: gene encoding a putative signalling protein with the conserved diguanilate phosphodiesterase EAL domain. Linear maps are drown to scale except the enlarged regions showing ARIs in detail.

MinION sequencing made also possible to determine different plasmid rearrangements leading to the loss of marker genes of R55, R16a and IP40a observed before. Every phenotypic change proved to be the result of recombination events involving mobile elements. Comparative analysis of deletion derivatives and the original plasmids indicated that several transposon families (Tn*3*, IS*26*, ISCR*2*) are the key players in reshaping new plasmid derivatives. The genomics analysis of the whole IncA/C family based on >150 sequenced members and the case-studies of our deletion derivatives led us to better understanding the evolution of these plasmids, helped to define the conserved plasmid backbone, the resistance islands (ARIs) and revealed the role of transposons in the fast evolution of the family [12].

Investigations of the genetic background of the relationship between SGI1 and the conjugative helper plasmids.

SGI1 is unable for self-transfer and its conjugation depends on the IncA/C helper plasmids. SGI1 normally exists as a chromosomally integrated element, which is very stable and its excised form is hardly detectable [6]. However, the mobilization of the island obviously requires an excised circular form. Since SGI1 can not replicate as plasmid, the excised form would rapidly be lost from the cell population. This controversy between the stable vertical transmission and the requirements of horizontal transfer in the presence of a helper plasmid suggested that there must be some kind of communication between the island and the incoming helper. At the beginning of the project, such communication was unknown, thus we intended to study the possible interactions of the partners.

Discovery of the helper induced SGI1 excision.

The first indications for such communication came from segregation tests that demonstrated the high stability of SGI1 in *S.* T. DT104 in absence of IncA/C plasmids and the high rate of SGI1 loss in the presence of helper. PCR tests indicated that in the latter case, the circular (excised) form of SGI1 is much more abundant. These findings led us to support that the helper plasmid somehow induces SGI1 excision. To identify the plasmid-encoded inductor, we have developed a two libraries representing ca. 10× coverage of the IncA/C-family plasmid R55. The libraries were introduced into a *E. coli* host strain harbouring chromosomal SGI1-C variant (TG1Nal::SGI1-C). The libraries then were screened by PCR specific for attP of SGI1 representing the joined ends of the island in the excised circular form. Nine attP+ R55 subclones were found, which promoted SGI1 excision. Restriction and Southern analyses showed that the 9 clones represent two regions of the R55 plasmid. The clones containing the shortest inserts were selected for end-sequencing to determine the plasmid region. The genes responsible for SGI1 excision were identified by deletion and KO-mutagenesis performed on the isolated R55 segments. By this approach we could identify the flhDC-family regulator genes (designated as *acaCD*), which have been shown to trigger SGI1 excision [21]. We also showed that this function is indispensable for both the plasmid and SGI1 transfer and also responsible for the elevated instability of the island observed in the presence of helper plasmids. Deletion of these genes from the helper plasmid led to complete loss of conjugal transfer of both the plasmid and SGI1, suggesting that this regulator has a central role in the spread of SGI1 and its helper plasmid, as well. The *acaCD* genes have

been cloned into expression vectors, which could fully complement the *acaCD* deletion mutant. Since the excision of SGI1 depends on the functioning if the λ integrase-family recombinase Int and the recombination directionality factor Xis, we analised their expression in the presence of AcaCD. Promoters of both genes were analysed by microbial (β-galactosidase) and biochemical (primer extension) methods and the transcription start sites were identified. We showed that *int* gene is expressed constitutively from a relatively strong promoter, while *xis* expression is AcaCD-dependent (Fig. 5ab).

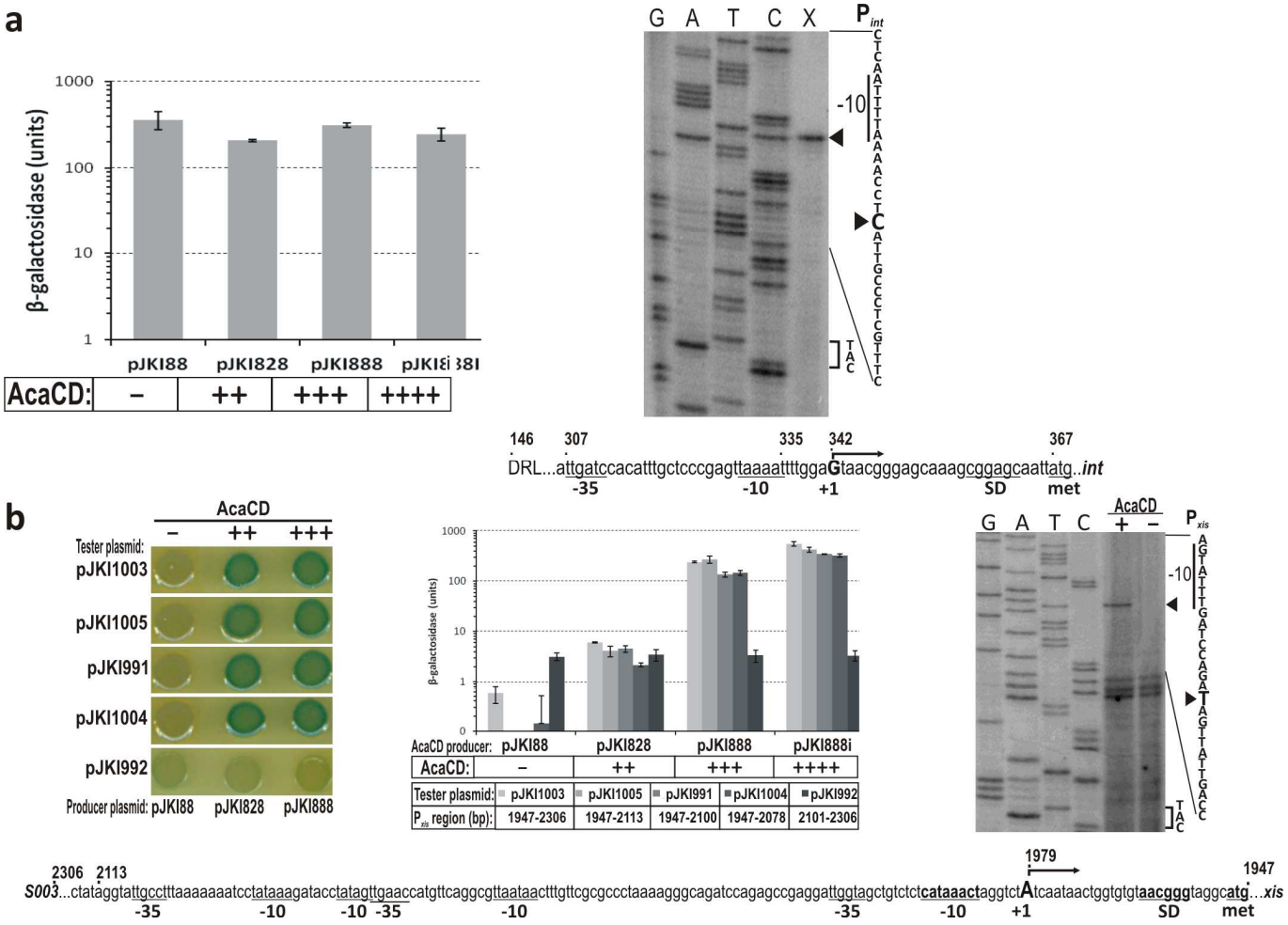

Fig. 5. (**a**) Analysis of promoter region P_{int}. The effect of AcaCD on the expression from P_{int} region was assayed by β-galactosidase assay carried out with TG1 strain containing the tester plasmid pJKI995 (P_{int}) and AcaCD producer plasmids pJKI828 or pJKI888, respectively. Vector pJKI88 lacking *acaCD* genes was used as negative control. Producer plasmid containing P_{tac} (pJKI888) was measured under non-inducing and inducing (pJKI888i) conditions. The relative expression levels of *acaCD* are indicated. Determination of *int* TSS was carried out by primer extension reaction (lane X) that was performed using total RNA purified from *E. coli* TG90 carrying pJKI995 and primer pUCfor21 annealing near to the start codon of *lac*Z gene. Lanes G, A, T, C: Sanger sequencing reactions obtained using pUCfor21 and pJKI995 template DNA. Arrowheads point to the C base on the non-transcribed strand corresponding to the G located 25 bp upstream from the ATG codon on the sense strand. The -10 box and the start codon are indicated. The promoter region of *int* is shown below the sequencing gel. Coordinates above the sequence refer to published SGI1 sequence (AF261825). The start codon, deduced Shine-Dalgarno, -10, -35 boxes and TSS of *int* are indicated. (**b**) Analysis of promoter region P_{xis}. Drop test and β-galactosidase assay of different regions of P_{xis} in the absence or presence of AcaCD shows AcaCD-dependent expression from P_{xis} promoter. The assay was carried out with TG1 strain containing the tester plasmids pJKI1003, pJKI1005, pJKI991, pJKI1004 and pJKI992. Right side panel shows the determination of *xis* TSS. Extension reactions were performed using primer pUCfor21 and total RNA purified from *E. coli* TG90 carrying tester plasmid pJKI1003 +/- AcaCD producer plasmid pJKI888 (lanes + and -). Lanes G, A, T, C: Sanger sequencing reactions obtained using pUCfor21 and pJKI1003 template DNA. Arrowheads point to the T base on the non-transcribed strand corresponding to the A located 28 bp upstream to the ATG codon on the sense strand. The putative -10 box and the start codon are indicated. The presence (+) or absence (-) of AcaCD is shown. Sequence of P_{xis} promoter region is shown below. Coordinates above the sequence refer to published SGI1 sequence. The startpoint of *xis* transcript (uppercase A) and the deduced Shine-Dalgarno, -10 boxes are in bold, other potential -10 and -35-like elements are also indicated.

Next, a protein purification method was established for the heteromeric AcaCD regulator complex, and the purified protein was used in EMSA and fingerprint experiments by which we could locate the binding site of the regulator in the promoter region of *xis* (Fig. 6).

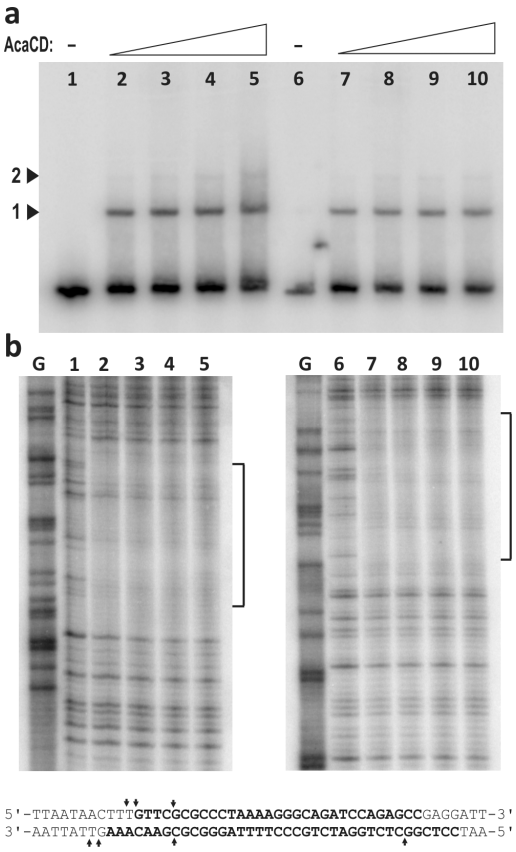

Fig. 6. Investigation of DNA-protein interactions between P_{xis} and the purified AcaCD activator. (**a**) Detection of P_{xis} DNA-AcaCD complexes by EMSA. ^{32}P single-end-labelled proximal fragment of P_{xis} (1947-2100 bp of SGI1) was subjected to AcaCD binding in 50 µl volume. Five µl reaction mix was applied for EMSA and the rest was used in the footprinting assay (panel **b**). DNA of P_{xis} was labelled at EcoRI end in lanes 1-5 (corresponding to the upper strand in the footprinting experiments and the sequence in panel **b**), and at HindIII end in lanes 6-10 (lower strand), respectively. AcaCD content of the 50 µl binding reactions was: lanes 1 and 6, 0 µg; lanes 2 and 7, 5 µg; lane 3 and 8, 10 µg; lanes 4 and 9, 20 µg; lanes 5 and 10, 30 µg. Arrowheads point to the primary (1) and higher order (2) complexes. (**b**) Determination of AcaCD binding site by DNaseI protection footprinting experiment. G, G-specific Maxam-Gilbert sequencing lane. Brackets indicate the regions protected by bound AcaCD, corresponding bases are in bold in the sequence below. Arrows indicate positions with enhanced DNaseI cleavage.

Further mutation analysis of the binding site revealed that the highly conserved 13 bp IR motif located asymmetrically in the AcaCD-protected region has a crucial role in AcaCD-dependent excision activation. This study demonstrated that the FlhDC-family activator AcaCD, which regulates the conjugation machinery of the IncA/C plasmids, serves as a signal and P_{xis} acts as a sensor of helper plasmid entry. Thus SGI1 not only exploits the conjugal apparatus of IncA/C plasmids but also hijacks their regulatory mechanisms controlling the conjugation system for the exact timing and activation of excision to ensure efficient horizontal transfer, while the constitutive expression of Int and the lack of Xis in absence of helper plasmids ensures the integration in the recipient and the stable vertical transmission of the island in the growing cell population [21].

Additionally, we found an *flhDC*-related gene on SGI1, which turned to be the closest known homologue of acaCD. This regulatory gene, named as $FlhDC_{SGI1}$, shows the same activities as its plasmid-borne counterpart. Although it can also trigger SGI1 excision when it is overexpressed from a strong promoter, $FlhDC_{SGI1}$ has much weaker destabilization effect on SGI1. Robust activation of SGI1 excision by $FlhDC_{SGI1}$ seems unlikely as spontaneous loss of the island (in absence of the helper plasmid) can not be observed [6] and excision is hardly detectable even by PCR. Thus, determination of the exact function of $FlhDC_{SGI1}$ needs further studies.

Analysis of the AcaCD-dependent SGI1 promoters

Based on homology searches several further acaCD binding sites were predicted in SGI1 [11]. Since we have shown that $FlhDC_{SGI1}$ identified from SGI1 as a close relative of the helper-plasmid-encoded AcaCD activator acts similarly on P_{xis} promoter, we decided to investigate all the SGI1 promoters carrying predicted AcaCD-binding site. Beside P_{xis} promoter, upstream regions of four additional SGI1 genes of unknown function, *S004*, *S005*, *S012* and *S018*, contain putative AcaCD-binding sites. We have analysed their functionality and shown that all are AcaCD-responsive promoters (Fig 7a.). The promoters were characterized and the transcription start sites were also determined by primer extension assays. It was also shown that they are activated by either AcaCD or $FlhDC_{SGI1}$ (Fig. 7b). Moreover, we provide evidence that both activators act on the same binding site in P_{xis} and that $FlhDC_{SGI1}$ is able to complement the *acaCD* deletion mutant IncA/C plasmid R16a. Complementation of R16aΔ*acaCD* plasmid

by an $FlhDC_{SGI1}$-expressing plasmid restored not only the conjugation of the helper but the mobilization of SGI1, indicating the compatibility of the two regulators [22].

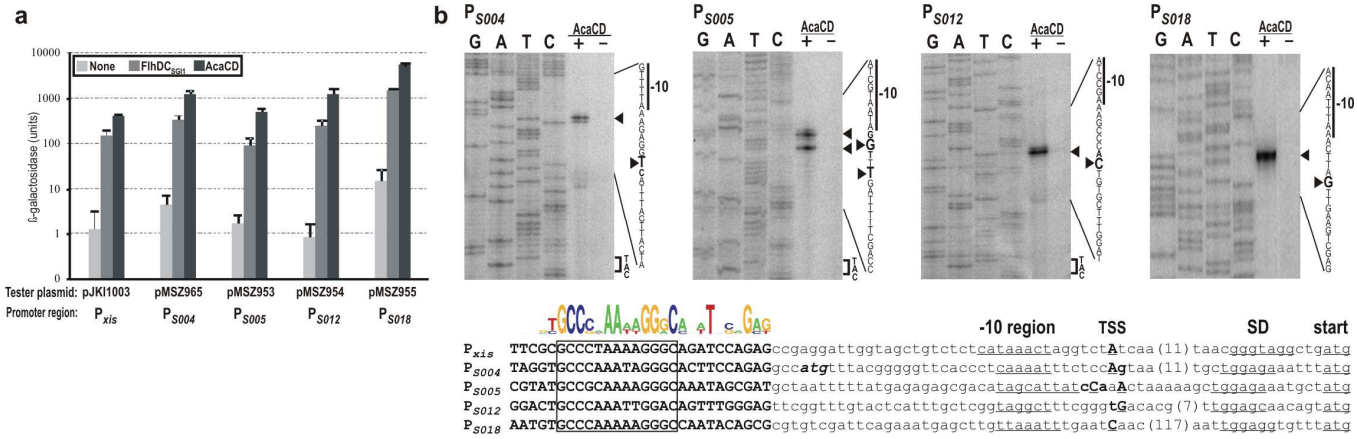

Fig 7. Analysis of the predicted AcaCD-dependent promoters of SGI1. (**a**) Activation of five SGI1 promoters containing predicted AcaCD-binding site by $FlhDC_{SGI1}$ and AcaCD. The activators were expressed from p15A-based producer plasmids pGMY6 and pJKI888, while the empty vector pJKI88 was applied as negative control. (**b**) Determination of TSSs of AcaCD/$FlhDC_{SGI1}$-responsive SGI1 genes. Primer extension reactions were performed using primer pUCfor21 and total RNA purified from *E. coli* TG1 carrying tester plasmids pMSZ965 (P_{S004}), pMSZ953 (P_{S005}), pMSZ954 (P_{S012}) and pMSZ955 (P_{S018}) +/- the AcaCD producer plasmid pJKI888 (lanes + and -). Lanes G, A, T, C: Sanger sequencing reactions obtained using pUCfor21 and the appropriate tester plasmid as template DNA. Arrowheads point to the base on the non-transcribed strand corresponding to the TSS on the sense strand. The putative -10 box and the start codon (if applicable) are indicated. The presence (+) or absence (-) of AcaCD is shown. The sequence alignment of the five AcaCD-dependent promoter regions are shown below the images. The putative -10 box, SD-box and the start codon are underlined. TSSs detected are shown as bold, underlined uppercases indicate the most frequent start sites. The predicted AcaCD-binding sites are shown as bold uppercase, the conserved core regions are boxed. Conserved positions are represented by sequence logo. The start codon of the annotated *S004* orf is indicated as bold italics.

The observed interchangeability of the plasmid- and SGI1-encoded activators rises that complicated crosstalk may exist between the helper and SGI1, similarly to that observed with other SGI1 genes[23]. During the analysis of the AcaCD-responsive promoters it was proved that orf *S004* is expressed from a different start codon than predicted earlier and the showed that, unlike the originally predicted S004 protein, the shorter protein expressed from the second start codon is biologically active, however its function is not yet clear.

Development of a minimal SGI1 model system.

At the beginning of the research most genes of SGI1 backbone were of unknown function. Since our approaches were mostly analytical, we decided to develop a minimal-SGI1 model system, which enables synthetic approaches and verifying the results obtained from the analytical experiments. We also expected from this modelling that we will learn more about this system and it will result in a model for further studies, which is easier to handle and manipulate. The main goal was to integrate all functions into the model that are required for horizontal spread of the wt island without unnecessary regions. The first model system containing only the site-specific recombination system of SGI1 has been developed by site-directed internal deletion in SGI1 located on a plasmid and the chromosome. This minimal SGI1 model retained only the excision/integration functions (DRL-*int*-*xis*-DRR). When it was tested for the activity of the two genes, high rate of excision was detected in the presence of the IncA/C helper plasmid, R55, indicating that the helper-encoded function, triggering the excision, worked similarly in the minimal system and the wt SGI1. However the extremely high rate of mini-SGI1 loss suggested that some kind of stabilizing system might be present in the wt SGI1. This hypothesis was later confirmed by our French partners as they found a new toxin-antitoxin system in SGI1 backbone, which were shown to stabilize the island in the presence of the helper [24].

The chromosomal minimal system has been further improved to a platform which is able to accept other SGI1-derived fragments by site-specific recombination. Exploiting the capacity of this system for gradual enlargement by

integration of specific segments of SGI1, the whole mob_{SGI1} region required for conjugative transfer of SGI1 has been inserted into the minimal system, resulting in the "minimal conjugative SGI1 model". This model was tested in mating assays and it was proved that its conjugal transfer occurred in the presence of the helper plasmid R55, however, the transfer frequency was less than that of wt SGI1. This suggested that some unknown functions may be missing from this system. To reveal these function(s), a set of 15 consecutive deletions in SGI1 was generated. Detailed phenotypical analysis of these deletion mutants that were characterized by conjugation, excision and stability assays led us to identify a region that influences the stability of the island. The functional analysis of this region will be continued.

WP 2.

The $SGI1^+$ clone of *S.* T. DT104 became globally prevalent in the 1990s. Its rapid spread among livestocks and the seemingly more serious symptoms caused by this clone compared to other $SGI1^-$ *Salmonella* strains raised the assumption that SGI1 may carry some factors influencing the pathogenicity of its host. It has also been proposed that SGI1 may be indirectly involved in the pathogenicity of its bacterial host [25] [26]. However, the question remained unanswered in the lack of isogenic pairs of $SGI1^+$ and $SGI1^-$ strains, which could have elucidated this issue. Therefore, we decided to generate and systematically study such parental and mutant strains in order to understand the possible pathogenetic significance of SGI1. The aim of WP 2 was to study the possible role of SGI1 in fitness and virulence of *S.* T. DT104 strains and in pathogenesis of *Salmonella* infections.

Studies on the interactions between SGI1 and its bacterial host.

Influence of SGI1 on the host's fitness.

The first step of this topic was the generation of isogenic $SGI1^+$ and $SGI1^-$ strains of *S.* T. DT104. Since the excision of SGI1 was reported to be detectable by PCR [9], the first approach to gain $SGI1^-$ *S.* T. DT104 derivatives was the attempt for isolation of clones that spontaneously lost SGI1. However, such derivatives could not be isolated even after 46 passages conducted without selection for SGI1 [6], which indicated the high stability of the island. Thus, SGI1 has been removed from three independent $SGI1^+$ *S.* T. DT104 isolates by the λ Red recombinase mediated site-directed deletion method [15]. The resulting strains harboured the CmR gene cassette at the position of the deleted SGI1. As a next step the resistance gene was also removed using a second, site-specific recombination event mediated by the Flp/FRT system of yeast. In the SGI1-free strains, a 83-bp insertion replaced the whole SGI1, however this insertion caused a change in the last 5 codons of *thdF* gene of which the 3' end serves as integration site for SGI1. *In vitro* competition experiments, however, indicated that the *thdF* mutation has a negative effect on the fitness, therefore these $SGI1^-$ strains could not be used as control strains in comparative analyses. This was supported by *in vivo* caecal colonization and organ invasion tests carried out by oral infection of chickens, which clearly indicated a reduced caecal colonization and organ invasion of the three *thdF*-mutant SGI^- strains.

After this, new approaches have been worked out to create $SGI1^{+/-}$ pairs of strains with strictly isogenic genetic background. i) SGI1 has been transferred into the originally SGI1-free *S.* T. DT104 strain ST903. First, the island was mobilized from the $SGI1^+$ ST919 *Salmonella* strain into *E. coli*, then, SGI1 was transferred from this *E. coli*

strain into the *S.* T. strain ST903 in a second conjugation step. The integration of SGI1 at the chromosomal *thdF* sequence of ST903 was verified by PCRs. ii) In the meantime, the identification of *acaCD* activator gene on the IncA/C helper plasmids solved the problem of SGI1 curing. When we recognized that SGI1 shows high segregation rates in the presence of *acaCD*, we could develop a well-curable *acaCD*-expressing plasmid with a thermo sensitive (*ts*) replication system. Using this plasmid, SGI1 has been cured from three SGI1$^+$ *S.* T. DT104 strains. The strains harbouring the curing plasmid were passaged without selection for SGI1, then the plasmid was also cured from the SGI1$^-$ segregants by growing at 42 °C. These SGI1$^-$ isolates and their SGI1$^+$ parental strains regarded as strictly isogenic were used in the comparative examinations of fitness and virulence

First, the growth rate of the three isogenic pairs of strains was tested and no significant difference was observed between the SGI1$^+$ and SGI1$^-$ counterparts. Then *in vitro* competition tests were carried out where equal amounts of the SGI1$^+$ and SGI1$^-$ counterparts were mixed and passaged 10 times without antibiotic selection for SGI1 (each passage represented ca. 8 generations). The proportion of SGI1$^+$ cells dropped below the 3×10^{-4}/ml at the 6th passage in each population, which clearly showed that the presence of SGI1 had negative effect on the fitness of its host strain in this experimental setup.

Th effect of SGI1 on the host's pathogenicity and virulence.

After preliminary experiments that were carried out to adjust the *in vivo* experimental setup, the pathogenicity of the SGI1$^{+/-}$ pairs of *S.* T. strains was compared in *in vivo* caecal colonization and organ (liver and spleen) invasion tests carried out by oral infection of chickens. Two sets of experiments were carried out with the three isogenic pairs of SGI1$^+$ and SGI1$^-$ strains (two derived from the SGI1-curing method and one from the SGI1-transfer into a new naïve host) using day-old SPF chickens. The *Salmonella* counts (CFU/g) were determined at 5 dpi from caecal, liver and spleen samples. The three pairs of strains showed no significant difference either in caecal colonization or organ invasion Fig. 2.

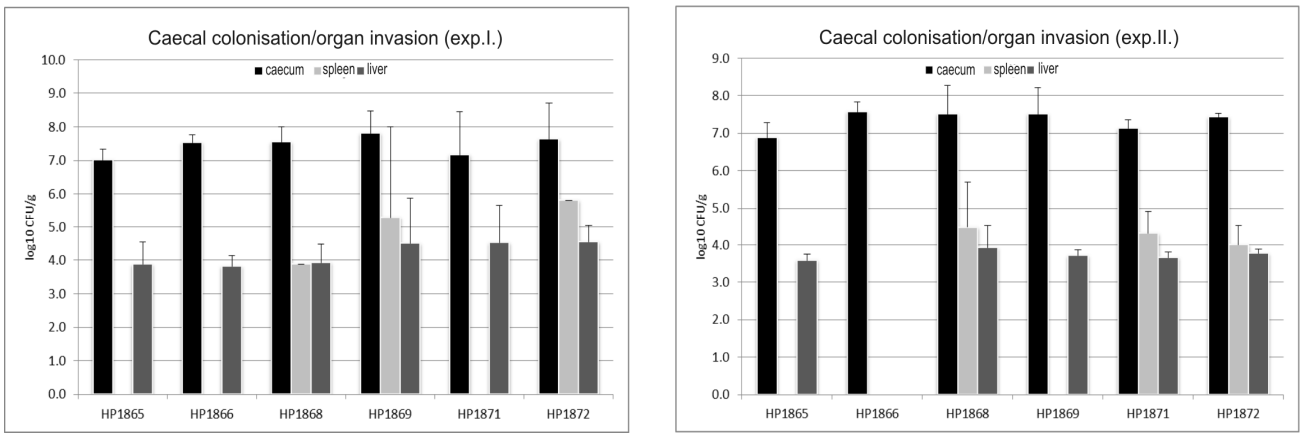

Fig. 8. Comparison of caecal colonization and organ invasion of SGI1$^{+/-}$ isogenic pairs of three *S.* T. DT104 strains in day-old chicks. The strains: ST903 (SGI1$^-$) derivatives - HP1865 (NalR, SGI1$^+$ transconjugant) / HP1866 (NalR, SGI$^-$ [SGI1-cured HP1865]); ST916 (SGI1$^+$) derivatives - HP1868 (NalR, SGI1$^+$ [parental]) / HP1869 (NalR, SGI1-cured HP1868); ST919 (SGI1$^+$) derivatives - HP1871 (NalR, SGI1$^+$ [parental]) / HP1872 (NalR, SGI1-cured HP1871).

Based on these and our previous data we concluded that SGI1 does not increase the pathogenicity or virulence of the tested *S.* Typhimurium strains. Moreover, SGI1 may represent a genetic load under nonselective conditions. Our results have recently been supported by a clinical monitoring work [27], where the severity of symptoms of *Salmonella* infections were statistically evaluated on 100 patients and SGI1-dependence could not be observed.

WP 3.

***In vivo* horizontal transfer of SGI1 in the intestinal microflora**

At the beginning of the project, SGI1-variants appeared to be restricted mainly to several *Salmonella* serovars and few *Proteus* strains. Interestingly, the presence of SGI1 has not been reported so far in commensal or pathogenic *E. coli* constituting integral part of the intestinal flora of chicken, despite of the fact that the laboratory K12 strains have been shown to be good recipients of SGI1 [9] [6]. This raised the question whether some natural barrier prevents the horizontal transfer of SGI1 to other bacteria occurring together with *Salmonellae* (such as *E. coli*) in the same host environment or alternatively, whether selective pressure by antimicrobials may facilitate the *in vivo* spread of SGI1. Although several further species (<10) have recently been shown to carry SGI1-related elements, interspecific transfer of SGI1 appears to be inefficient. Therefore the aim of WP3 was to monitor the capacity of chicken gut flora to accept and distribute SGI1 *in vivo* (in living animals) and the influence of pre-treatment with antimicrobials to the efficiency of this process, applying both metagenomics and classical microbiological approaches.

Construction of the SGI1 donor S. T DT104 strains

For investigation of *in vivo* spread of SGI1 in chicken gut flora, development of the appropriate donor strain is crucial. For donors the two basic requirements are that they have to possess both the mobilizable SGI1 and the helper plasmid. Further important factor is that the donor strain should disappear in a relatively short time after infection, but this reduced persistence in gut flora should be enough to achieve detectable level of SGI1 transfer to the potential recipients. Gradual decrease of the donor titer in gut flora is important because high titers of donor strains in the samples collected for the metagenomic analysis would interfere with detection of rare transconjugants. Further requirement was to introduce genetic markers into the donor strains that would allow the identification and separation of this strain from other bacteria of gut flora by microbial methods and PCR.

To fulfill these requirements we decided to develop a conditionally complemented lethal mutation in our donor strains, which ensures their maintenance under laboratory conditions, but prevents significant growth under natural *in vivo* conditions such as chicken intestine. The *asd* (aspartate-semialdehyde dehydrogenase) gene encoding for a key enzyme in cell wall synthesis and biosynthesis of several amino acids has been reported to be a good target gene [28] as its deletion significantly decreased the fitness of the mutant strains [29] [30]. Thus it was chosen for generating donor strains with reduced fitness.

To apply the one-step gene inactivation method for deleting *asd* gene of a wt *S.* T. DT104 isolate, we first had to cure SGI1 as the KO allele could not have been selected due to its multiresistance. First, SGI1 was cured from *S.* T. DT104 strains ST919 and its NalR derivative using the thermo sensitive curable plasmid expressing the AcaCD activator (see WP 2.). This step made the genetic manipulations of the strains possible as 7 resistance markers had been lost together with SGI1. In the next step *asd* gene was deleted from the chromosome resulting in the auxotroph mutant strains, which require diamino-pimelic acid (DAP) for their growth. For construction of complementing plasmids, *asd* gene was PCR-amplified and cloned in plasmids. Two alternative ways for conditional complementation were chosen. In one case, *asd* was placed under the control of P_{araBAD} promoter, as L-arabinose, the inductor of this promoter can easily be supplemented under lab conditions ensuring normal growth of the complemented strain, but occurs at a low concentration in animal intestine, which leads to repression of *asd*-

expression by AraC (encoded by the same plasmid) suspending the appropriate complementation of the chromosomal *asd* KO mutation. In the second case, the complementing plasmid contained *asd* gene with its own promoter, but the plasmid had a thermo sensitive pSC101 replication system, which can stably maintained at 30 ^{o}C but is unstable at 40-42 ^{o}C, the normal body temperature of chicks, preventing the maintenance of the complementing plasmid. After introduction of the two plasmids into the *Salmonella* strains they do not require antibiotic selection for stable maintenance as the conditional complementation of the lethal mutation [in the presence of arabinose or low (30 ^{o}C) growth temperature, respectively] stabilizes the plasmids. As a next step, SGI1 was reintroduced to the *asd*-complemented and the *asd*$^{-}$ auxotroph *Salmonella* donor strains from an appropriately designed *E. coli* strain, into which SGI1 was previously transferred from ST919. Finally, the helper plasmid had to be introduced into the donor strains. Our previous observations indicated that, regarding their resistance markers, the most stable IncA/C plasmid available for us is R55 [12], which has several common resistance markers with SGI1. To obtain the appropriate helper plasmid, the majority of resistance determinants of R55 were removed by multiple gene KO methods. The streamlined helper plasmid retaining only the *aadB* (KmR/GmR) gene, without the transposons causing its instability, was then conjugated into the *asd*$^{-}$, the conditionally *asd*-complemented and the original ST919 and ST919Nal donor strains.

Optimization of the experimental setup.

First the growth and conjugative abilities of the donor strains were examined under laboratory conditions. The cell counts of the non-complemented *asd*$^{-}$ strain in absence of DAP decreased after the 3rd hour and the CFU/ml dropped 3 orders of magnitude in 8^{h} growth time, which proved that *asd* KO has a serious fitness effect. In similar tests the two complemented strains showed similar growth rates with that of the wt strains under permissive (+ara or 30 ^{o}C) and restrictive (no ara or 42 ^{o}C) conditions. The latter observation suggested that termination of the complementation and the manifestation of *asd*$^{-}$ phenotype is not complete and/or needs longer time period.

The SGI1 donor activity was tested in mating assays using laboratory (TG1NalR or TG1RifR) and a natural, chicken-adapted *E. coli* PC2 recipient strains. Under permissive conditions all donor strains showed similar SGI1 transfer rates to that of wt donor strains (4.2-10.0×10^{-2}/recipient CFU), while under restrictive conditions the complemented strains showed no (no ara) or 3-fold (42 ^{o}C) reduction of transfer rate, while the non-complemented *asd*$^{-}$ strains showed 2 orders of magnitude lower frequencies. Using *E. coli* PC2 as recipient, similar results were obtained proving that SGI1 transfer can occur into this natural strain as well, however the transfer frequencies were 3 orders of magnitude lower than observed wit TG1 recipients probably due to the active restriction systems of the recipient. The ratios of donor and recipient titers under restrictive conditions indicated the reduced fitness of the non-complemented and the temperature-dependent strains, while the ara-dependent donor showed nearly wt phenotype during the mating.

To determine the optimal sampling period for testing the *in vivo* transfer of SGI1, the shedding of *S.* T. DT104 was monitored in chicks between 7-13th days post infection in comparison with the less invasive S. Infantis. Based on the data obtained it seemed that the ST919 donor strain colonize day-old chicks (and was shed from these chicks) at a lower level compared to the less invasive, but well colonizing S. Infantis, and the most feasible period for testing *in vivo* transfer of SGI1 from ST919 will be after the first 10 days post infection.

Assaying the in vivo horizontal spread of SGI1 in the chicken intestinal microflora.

The wt (HP2152) and two conditionally *asd*-complemented ST919 (NalS) donor strains (HP2154 – ara-dependent, HP2159 – *ts*) were applied in two separate *in vivo* infection experiments using 6-day old, and 1-day-old chicks, respectively. Chicks were infected with the PC2 recipient *E. coli* strain in 10^7 CFU/chick p.o. and challenged with the donor strains in 5×10^8 CFU/chick after 2 hours. Faecal shedding of the donor and recipient strains were monitored for 13 days of age. The data showed that the *ts* donor strain has a reduced viability compared to the wt strain in 6-day-old chicks (having a developed gut flora), as it disappeared from the faeces by 13th day post infection. In day-old chicks, where the infection occurred before development of the gut flora, titers of all donor strains decreased similarly (ca. 2 logs decrease at the 10th d.p.i.). This experimental setup allowed the isolation of PC2::SGI1 transconjugants with all three donor strains from faecal and caecal samples of day-old chicks (Table 1).

Table 1. PC2::SGI1 transconjugant strains isolated from faecal or caecal samples

Donor	Recip.	Exp	Sample	Sampling d. p. i.	Total isolated	No.	SGI1 PCR	E. coli PCR	Coliform agar	Km/GmR (R55+)	Designation
HP 2152	PC2	II.	faeces	3	2	1	+	+	*E. coli*	-	tkj HP2152/1
						2	+	+	*E. coli*	-	tkj HP2152/2
HP 2152	PC2	II.	caecum	6	4	1	+	+	*E. coli*	-	tkj HP2152/3
						2	+	+	*n.e.*	-	tkj HP2152/4
						3	+	+	*n.e.*	-	tkj HP2152/5
						4	+	+	*E. coli*	-	tkj HP2152/6
HP 2152	PC2	II.	faeces	8	3	1	+	+	*E. coli*	-	tkj HP2152/7
						2	+	+	*n.e.*	-	tkj HP2152/8
						3	+	+	*E. coli*	-	tkj HP2152/9
HP 2154	PC2	II.	caecum	6	154	1	+	+	*E. coli*	-	tkj HP2154/1
						2	+	+	*n.e.*	-	tkj HP2154/2
						3	+	+	*n.e.*	-	tkj HP2154/3
						4	+	+	*E. coli*	-	tkj HP2154/4
						5	+	+	*n.e.*	-	tkj HP2154/5
						6	+	+	*n.e.*	-	tkj HP2154/6
						7	+	+	*E. coli*	-	tkj HP2154/7
						8	+	+	*n.e.*	-	tkj HP2154/8
						9	+	+	*E. coli*	-	tkj HP2154/9
						10	+	+	*n.e.*	-	tkj HP2154/10
						11	+	+	*n.e.*	-	tkj HP2154/11
						12	+	+	*E. coli*	-	tkj HP2154/12
						13	+	+	*n.e.*	-	tkj HP2154/13
						14	+	+	*n.e.*	-	tkj HP2154/14
						15	+	+	*E. coli*	+	tkj HP2154/15
						16	+	+	*n.e.*	-	tkj HP2154/16
HP 2159	PC2	I.	faeces	1	1	1	+	+	*E. coli*	-	tkj HP2159/1

The presence of *E. coli* transconjugants could also be detected by nested PCR from metagenomic DNAs isolated from caecal samples. (Fig. 9).

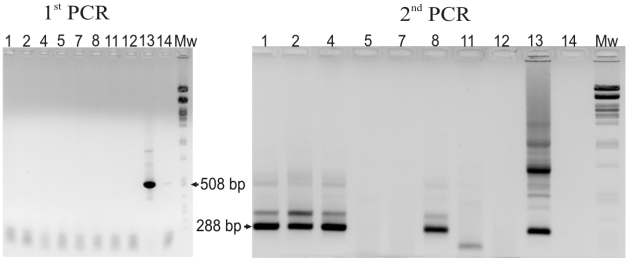

Fig. 9. Detection of PC2::SGI1 transconjugants by semi-nested PCR in metagenomic DNA isolated from chicken caecal samples. Lanes 1-12: metagenomic template DNA samples, 13: PC2::SGI1 genomic DNA template (+ control), 14: nontemplate control.

As expected, the wt and ara-dependent strains resulted in more transconjugants, but the *ts* strain proved also to be an active SGI1 donor. PC2 transconjugants were validated by colony PCRs specific for *E. coli* and SGI1 markers. Among 26 tested clones only one carried also the helper plasmid. In this way we could provide the first evidence on the *in vivo* horizontally transfer of SGI1 in gut flora of a living animal.

After further fine tuning, large scale infection experiments were carried out in two setups. In the first setup, 6-day-old chicks were infected with the wt and the *ts* donor strains (5×10^8 CFU/chick) w/o 100 mg/bwt kg tetracycline/day. In the second setup, the wt and the arabinose-dependent donor was applied (5×10^7 CFU/chick) to day-old chicks w/o tetracycline addition. Tetracycline is a broad-spectrum polyketide antimicrobial drug that has been used extensively in poultry-farming for several decades. Probably this practice has accelerated the spread of resistant pathogens, such as SGI1$^+$ *Salmonellae*. Therefore in this experiment, we intended to model the impact of preventive antibiotic treatment on the horizontal transfer of SGI1 in the intestinal flora. The absence of *Salmonellae* in the chicks was confirmed before infections by the standard detection method (ISO 6579-2002). Twelve chicks were infected in each group and faeces and caecal samples as well as agar cultures in 20% glycerol-PBS from caeca were collected at 3rd, 6th and 9th d.p.i. from 3 chicks/group/sampling. The donor titers were monitored and one representative chick/strain was controlled by PCR. Ca. 200 faecal and caecal samples were collected and stored at -70 °C until metagenomic DNA extraction.

For detection of SGI1 transconjugant bacteria in gut flora a PCR-based NGS metagenome-sequencing approach was applied. To avoid sequencing huge amounts of chicken and indigenous microbial DNA, a PCR selective for DNA fragments containing SGI1 – target junctions was designed. During library preparation, after the adapter ligation, PCR amplification was carried out using primers facing outward of SGI1 ends with primers annealing to the adapter. For optimization of the procedure, the first metagenome sequencing was carried out on total DNA samples obtained from chickens infected at 6-day-old age and where PC2::SGI1 transconjugants could be isolated. These transconjugants can control the sensitivity of the detection method. The average length of the libraries for the left (DRL) and right (DRR) end of SGI1 was 650 and 700 bp, respectively. The sequencing resulted in 4.0 and 3.2 Mbase assembled sequence. Unfortunately, vast majority of the reads represented offtarget (chicken and microbial DNA) sequences and only ca 1000 reads came from SGI1-junctions (all from the donor strain). These results indicated that the specificity of the amplification is unsatisfactory. Since careful fine-tuning the PCR conditions by gradient PCR could not significantly reduce offtarget amplification, a semi-nested PCR will be applied, where two SGI1-specific primers are used in consecutive amplification steps to increase the specificity. The optimization of library preparation by this semi-nested PCR method is ongoing. We hope that sequencing of an appropriately specific library will provide information on the recipient bacteria in gut flora that can accept SGI1 and the quantitative comparison of the results will inform us on the effects of antibiotic treatment regarding SGI1 spread.

Although this part of the project has not yet finished, it led to the development of appropriate donor strains for further in vivo studies and to the first demonstration of interspecific horizontal transfer of SGI1 under quasi-natural conditions in the intestinal flora of chickens.

NEW RESULTS

- Identification and characterization of $oriT_{SGI1}$
- Identification of two self-encoded conjugation factors (S020, S019) of SGI1.

- Localization of the promoter regions of the S020-S019 operon and S019.
- Establishment of expression and purification protocols for S020 and S019 proteins.
- Construction of a helper *E. coli* strain containing the mobSGI1 region integrated in the chromosome.
- Identification and characterization of $oriT_{A/C}$ of the IncA/C plasmids.
- Functional analysis of relaxase and *mobI* genes of IncA/C plasmids
- Identification of a recombination site (RecHS) of a sit-specific recombination system.
- Complete ungapped sequence of two ancient IncA/C plasmids.
- Dataset of IncA/C-family plasmids for comparative analyses
- Establishment of the third generation sequencing platform MinION from Oxford Nanopore.
- Identification of rearrangements, description of mechanisms involved in the evolution of IncA/C-family.
- Identification of new antibiotic resistance islands (ARIs) and transposons in R16a and IP40a.
- Verification and genetic analysis of the helper induced SGI1 excision.
- Identification and characterization of P_{int} promoter of SGI1.
- Demonstration of the sequence-specific binding of the plasmid-encoded AcaCD activator to the promoter region of xis_{SGI1}.
- Identification and characterization of four AcaCD-inducible promoters of SGI1.
- Identification and functional analysis of the *acaCD*-related *flhDC*-family regulator of SGI1.
- Development of a mini-SGI1 model system in plasmid and chromosome
- Establishment of an SGI1-curing method.
- Generation of isogenic $SGI1^+$ and $SGI1^-$ *S.* T. DT104 strains.
- Analysis of SGI1-host bacterium interaction: demonstration of the lack of pathogenicity/virulence factors on SGI1 and the genetic load of SGI1 harbouring host under nonselective conditions.
- Construction of conditional lethal *S.* T. DT104 donor strains for *in vivo* studies on the horizontal transfer of SGI1 and development of the appropriate *in vivo* model.
- First demonstration of *in vivo* interspecific transfer of SGI1 (*S.* T. DT104→*E. coli*) in the gut flora of chicken.

CONCLUDING REMARKS AND ACKNOWLEDGEMENTS

The project has established the studies of a system consisting of a mobilizable MDR genomic island and its helper plasmid, which both are important vehicle of lots of resistance determinants among numerous bacteria in *Enterobacteriaceae*. We have reported detailed analyses of basic elements of their conjugative system (*oriT*s, conjugation genes) and uncovered the genetic background of the mode of hijacking the regulatory system of IncA/C plasmids by SGI1. We have developed appropriate experimental systems for investigations of the SGI1 – host interactions and the *in vivo* horizontal spread of SGI1. Although the research has raised several new and exciting questions, the project has achieved its original aims. Furthermore, the research contributes to the PhD thesis of four young researchers, which is undoubtedly an important outcome of this project. We thank to OTKA/NKFIH for supporting our research proposal.

References

(Publications from this project are shown as bold.)

1. Levings RS, Lightfoot D, Partridge SR, Hall RM, Djordjevic SP. The genomic island SGI1, containing the multiple antibiotic resistance region of Salmonella enterica serovar Typhimurium DT104 or variants of it, is widely distributed in other S. enterica serovars. J Bacteriol. 2005;187: 4401–4409. doi:10.1128/JB.187.13.4401-4409.2005
2. Levings RS, Partridge SR, Djordjevic SP, Hall RM. SGI1-K, a variant of the SGI1 genomic island carrying a mercury resistance region, in Salmonella enterica serovar Kentucky. Antimicrob Agents Chemother. 2007;51: 317–323. doi:10.1128/AAC.01229-06
3. Levings RS, Djordjevic SP, Hall RM. SGI2, a relative of Salmonella genomic island SGI1 with an independent origin. Antimicrob Agents Chemother. 2008;52: 2529–37. doi:10.1128/AAC.00189-08
4. Doublet B, Praud K, Bertrand S, Collard J-M, Weill F-X, Cloeckaert A. Novel insertion sequence- and transposon-mediated genetic rearrangements in genomic island SGI1 of Salmonella enterica serovar Kentucky. Antimicrob Agents Chemother. 2008;52: 3745–3754. doi:10.1128/AAC.00525-08
5. Siebor E, Neuwirth C. The new variant of Salmonella genomic island 1 (SGI1-V) from a Proteus mirabilis French clinical isolate harbours blaVEB-6 and qnrA1 in the multiple antibiotic resistance region. J Antimicrob Chemother. 2011;66: 2513–20. doi:10.1093/jac/dkr335
6. Kiss J, Nagy B, Olasz F. Stability, Entrapment and Variant Formation of Salmonella Genomic Island 1. Ahmed N, editor. PLoS One. 2012;7: e32497. doi:10.1371/journal.pone.0032497
7. Siebor E, Neuwirth C. Emergence of Salmonella genomic island 1 (SGI1) among Proteus mirabilis clinical isolates in Dijon, France. J Antimicrob Chemother. 2013;68: 1750–1756. doi:10.1093/jac/dkt100
8. Siebor E, Neuwirth C. Proteus genomic island 1 (PGI1), a new resistance genomic island from two Proteus mirabilis French clinical isolates. J Antimicrob Chemother. 2014;69: 3216–3220. doi:10.1093/jac/dku314
9. Doublet B, Boyd D, Mulvey MR, Cloeckaert A. The Salmonella genomic island 1 is an integrative mobilizable element. Mol Microbiol. 2005;55: 1911–1924. doi:10.1111/j.1365-2958.2005.04520.x
10. Douard G, Praud K, Cloeckaert A, Doublet B. The Salmonella genomic island 1 is specifically mobilized in trans by the IncA/C multidrug resistance plasmid family. PLoS One. 2010;5: e15302. doi:10.1371/journal.pone.0015302
11. Carraro N, Matteau D, Luo P, Rodrigue S, Burrus V. The Master Activator of IncA/C Conjugative Plasmids Stimulates Genomic Islands and Multidrug Resistance Dissemination. PLoS Genet. 2014;10: e1004714. doi:10.1371/journal.pgen.1004714
12. **Szabó M, Nagy T, Wilk T, Farkas T, Hegyi A, Olasz F, et al. Characterization of Two Multidrug-Resistant IncA/C Plasmids from the 1960s by Using the MinION Sequencer Device. Antimicrob Agents Chemother. 2016;60: 6780–6786. doi:10.1128/AAC.01121-16**
13. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31: 3406–3415. doi:10.1093/nar/gkg595
14. Chen J, Pappas DL, Galli DM. Mapping of the nick site on conjugative plasmid pVT745 by interrupted mating. Plasmid. Elsevier Inc.; 2010;63: 136–142. doi:10.1016/j.plasmid.2010.02.001
15. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97: 6640–5. doi:10.1073/pnas.120163297
16. Miller JH. Experiments in molecular genetics. Cold Spring Harb Lab Press Cold Spring Harb NY. 1972; Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17416361
17. **Hegyi A, Szabó M, Olasz F, Kiss J. Identification of oriT and a recombination hot spot in the IncA/C plasmid backbone. Sci Rep. 2017;7: 10595. doi:10.1038/s41598-017-11097-0**
18. Carraro N, Sauvé M, Matteau D, Lauzon G, Rodrigue S, Burrus V. Development of pVCR94ΔX from Vibrio cholerae, a prototype for studying multidrug resistant IncA/C conjugative plasmids. Front Microbiol. 2014;5: e44. doi:10.3389/fmicb.2014.00044
19. Garcillán-Barcia MP, Francia MV, De La Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev. 2009;33: 657–687. doi:10.1111/j.1574-6976.2009.00168.x
20. Harmer CJ, Hall RM. The A to Z of A/C plasmids. Plasmid. 2015;80: 63–82. doi:10.1016/j.plasmid.2015.04.003
21. **Kiss J, Papp PP, Szabó M, Farkas T, Murányi G, Szakállas E, et al. The master regulator of IncA/C plasmids is recognized by the Salmonella Genomic island SGI1 as a signal for excision and conjugal transfer. Nucleic Acids Res. 2015;43. doi:10.1093/nar/gkv758**
22. **Murányi G, Szabó M, Olasz F, Kiss J. Determination and Analysis of the Putative AcaCD-Responsive Promoters of Salmonella Genomic Island 1. Hayes F, editor. PLoS One. 2016;11: e0164561. doi:10.1371/journal.pone.0164561**
23. Carraro N, Durand R, Rivard N, Anquetil C, Barrette C, Humbert M, et al. Salmonella genomic island 1 (SGI1) reshapes the mating apparatus of IncC conjugative plasmids to promote self-propagation. PLOS Genet. 2017;13: e1006705. doi:10.1371/journal.pgen.1006705
24. Huguet KT, Gonnet M, Doublet B, Cloeckaert A. A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable Salmonella Genomic Island 1. Sci Rep. 2016;6: 32285. doi:10.1038/srep32285
25. Golding GR, Olson AB, Doublet B, Cloeckaert A, Christianson S, Graham MR, et al. The effect of the Salmonella genomic island 1 on in vitro global gene expression in Salmonella enterica serovar Typhimurium LT2. Microbes Infect. 2007;9: 21–27. doi:10.1016/j.micinf.2006.10.004
26. Carlson S a, Sharma VK, McCuddin ZP, Rasmussen M a, Franklin SK. Involvement of a Salmonella genomic island 1 gene in the rumen protozoan-mediated enhancement of invasion for multiple-antibiotic-resistant Salmonella enterica serovar Typhimurium. Infect Immun. 2007;75: 792–800. doi:10.1128/IAI.00679-06

27. de Curraize C, Amoureux L, Bador J, Chapuis A, Siebor E, Clément C, et al. “Does the Salmonella Genomic Island 1 (SGI1) confer invasiveness properties to human isolates?” BMC Infect Dis. BMC Infectious Diseases; 2017;17: 741. doi:10.1186/s12879-017-2847-1
28. Galán JE, Nakayama K, Curtiss R. Cloning and characterization of the asd gene of Salmonella typhimurium: use in stable maintenance of recombinant plasmids in Salmonella vaccine strains. Gene. 1990;94: 29–35. doi:10.1016/0378-1119(90)90464-3
29. Santander J, Xin W, Yang Z, Curtiss R. The aspartate-semialdehyde dehydrogenase of Edwardsiella ictaluri and its use as balanced-lethal system in fish vaccinology. PLoS One. 2010;5. doi:10.1371/journal.pone.0015944
30. Kim SW, Kang HY, Hur J, Gal SW, Bang WY, Cho KK, et al. Construction of a conditional lethal Salmonella mutant via genetic recombination using the ara system and asd gene. J Microbiol Methods. Elsevier B.V.; 2011;87: 202–207. doi:10.1016/j.mimet.2011.08.004
31. Szabó M, Nagy T, Wilk T, Farkas T, Hegyi A, Olasz F, et al. Characterization of two multidrug-resistant IncA/C plasmids from the 1960s by using the MinION sequencer device. Antimicrob Agents Chemother. 2016;60. doi:10.1128/AAC.01121-16